



NEON Data Standards Technical Working Group

2020 Annual Report



1685 38th St., Suite 100 | Boulder, CO 80301 | 720.746.4844 | www.neonscience.org

National Ecological Observatory Network (NEON) is a project sponsored by the National Science Foundation and proudly operated by Battelle.

Introduction

Since its inception, NEON has relied on expertise within the science, education, and engineering communities to advise on key areas impacting the design, construction, and maintenance of the observatory with the goal to optimize its operation. Currently, two types of external advisory bodies support staff and leadership in making key decisions that guide all of NEON's activities: the Science, Technology & Education Advisory Committee (STEAC) and Technical Working Groups (TWGs). Both bodies are comprised of experts nominated to serve in these roles who are selected by NEON staff following a rigorous selection process.

NEON currently relies upon input from 22 TWGs. These groups play an important role by providing input to NEON's data collection and processing methods and ensuring that NEON infrastructure, data, and programs are a valuable community resource. Working groups are participatory and advisory; they are often tasked with providing input on issues that have scientific, educational, engineering, or operational implications.

This document includes a summary of activities, recommendations, and NEON's response to those recommendations for the Data Standards TWG during the 2020 funding year (November 2019-October 2020).

The Data Standards Technical Working Group is tasked with making recommendations about effective ways to provide NEON's data products to the broader scientific, educational, and policy communities. Topics may include 1) principles, standards, and policies for open data and software, 2) data discovery, exploration, and delivery mechanisms, 3) improvement of data products to increase utility, and 4) monitoring impact of NEON data use on research.

Q1 – November 2019-January 2020

Summary of Activities

The Data Standards TWG kicked off this year's TWG on December 19, 2019. We reviewed the charter and logistics of meetings and collaboration for coming year. We discussed needs for easier ways to cite data and code packages, as well as need for making the AOP data viewer usable in a dedicated webpage, in addition to the current model. In our second meeting, on January 6, 2020, the TWG reviewed NEON's acknowledgement and citation guidance/policy documents. We discussed CC0 and CC-BY licensing and the opportunities for NEON to provide more coherent guidelines to the ecological research community. The TWG also discussed upcoming changes to NEON's web portals.

TWG Recommendations

The Data Standards TWG made the following recommendations in Q1:

1. Include a bib/ris text file for data product citations with data downloads in addition to the existing copy button.
2. Create a "make citation" function in the neonUtilities package.
3. Add a new functionality to the neonUtilities package - when a user runs *stackByTable()*, the function will print out guidance on how to properly cite the data.
4. Create a separate interface for AOP data viewer, in addition to the current modals.

5. Use CCO licensing where possible for data.
6. Improve documentation around citing NEON outputs.

NEON Response

- 1-3: These are feasible, and our team has done some preliminary research.
- 4: The AOP viewer iframe can be popped out into a separate webpage/interface, and this has been recommended by other users. This can put on data portal team's to-do list, likely in late Spring.
5. Need to compile rationale for Science Data Quality Integrated Product Team (SDQ IPT) and leadership.
6. Our team has been working on a draft citation document, which has been preliminarily reviewed by the SDQ IPT. The need for better citations as previously been noted by Annual Operations Review (AOR) panel and others.

Q2 – February 2020-April 2020

Summary of Activities

The Data Standards TWG held three meetings in Q2. The Science Data Quality Integrated Product Team (SDQ IPT) had some questions about the costs and benefits of CCO licensing over CC-BY and requested that we continue the conversation and draft a memo. We held thorough discussions over the February and March meetings which led to a re-affirmation of the CCO model. A memo was produced for the SDQ IPT; this memo was discussed and approved by the SDQ IPT and sent forward to leadership for further review. In April, the TWG discussed the COVID-19 impacts on NEON, particularly the impacts of disrupted sampling and shipping to the bio-archive, as well as a potentially heightened interest in NEON data as academic sampling has also been disrupted. Members pointed out that this could support justification for more automated sampling (e.g., acoustic). We also discussed NEON's discussions with the Environmental Data Initiative (EDI) about hosting data derived from NEON data; it was noted that we could consider 1) pulling information about NEON derived data hosted on EDI via API hooks, 2) talk to Figshare about a NEON landing page such as what NIH has (<https://nih.figshare.com/>), and 3) think about other aggregators or data repositories.

TWG Recommendations

The Data Standards TWG made the following recommendations in Q2:

1. Use CCO licensing where possible for data.
2. Consider metrics and use cases for our upcoming API tokens.
3. Consider how to pull information about datasets derived from NEON data that may be hosted in repositories such as EDI, DataDryad, Figshare, Zenodo, and Mendeley.
4. Create better citation exposure/generation within neonUtilities and during download on the Data Portal.

NEON Response

1. We will continue working on approval of CCO licensing for NEON data.
- 2-3. We will research both ideas.

4. We are prototyping a function that creates a bibtex citation file during data downloading in a future release of neonUtilities.6. Our team has been working on a draft citation document, which has been preliminarily reviewed by the SDQ IPT. The need for better citations as previously been noted by Annual Operations Review (AOR) panel and others.

Q3 – May 2020-July 2020

Summary of Activities

The TWG met once, in June, and discussed progress toward data portal and neonscience.org rebuild goals. We also discussed the new data formats and conventions webpage.

TWG Recommendations

The primary recommendation from the TWG is that we be especially careful to include information about missing values and how users will and should interact with them.

NEON Response

We will include this recommendation into the data formats and conventions webpage.

Q4 – August 2020-October 2020

Summary of Activities

At the August 2020 meeting, the TWG discussed the upcoming plans for a data release and reviewed progress made toward previous recommendations. There was a brief discussion about a possibility of generating annual releases for the Darwin core records from the Biorepository, but more discussion would be warranted.

TWG Recommendations

Generally, the TWG agreed with the overall plans for data releases. The primary recommendation was to be very clear about the provisional data, in particular, to help users understand that these data may improve over time (as is typical with many datasets that are thoroughly QA/QC'd).

NEON Response

The internal data release working group is proceeding with developing the workflow and outputs of the first data release, and we are being very careful to consider the user perspective by implementing documentation about provisional vs. released data from the data portal webpages, within the data packages (readme and manifest files), and in the API responses