12-1-2021

# Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community

R. Chelsea Nagy
*University of Colorado Boulder*

Jennifer K. Balch
*University of Colorado Boulder*

Erin K. Bissell
*Metropolitan State University of Denver*

Megan E. Cattau
*Boise State University*

Nancy F. Glenn
*Boise State University*

*See next page for additional authors*

## Authors
R. Chelsea Nagy, Jennifer K. Balch, Erin K. Bissell, Megan E. Cattau, Nancy F. Glenn, Benjamin S. Halpern, Nayani Ilangakoon, Brian Johnson, Maxwell B. Joseph, Sergio Marconi, Catherine O'Riordan, James Sanovia, Tyson L. Swetnam, William R. Travis, Leah A. Wasser, Elizabeth Woolner, Phoebe Zarnetske, Mujahid Abdulrahim, John Adler, and Grenville Barnes

esa ECOSPHERE

SPECIAL FEATURE:
HARNESSING THE NEON DATA REVOLUTION

# Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community

R. Chelsea Nagy,[1,†] Jennifer K. Balch,[1,2] Erin K. Bissell,[3] Megan E. Cattau,[4] Nancy F. Glenn,[4,5] Benjamin S. Halpern,[6,7] Nayani Ilangakoon,[1] Brian Johnson,[1] Maxwell B. Joseph,[1] Sergio Marconi,[8] Catherine O'Riordan,[9] James Sanovia,[10] Tyson L. Swetnam,[11] William R. Travis,[1,2] Leah A. Wasser,[1,2] Elizabeth Woolner,[1] Phoebe Zarnetske,[12] Mujahid Abdulrahim,[13] John Adler,[2,14] Grenville Barnes,[15] Kristina J. Bartowitz,[16] Rachael E. Blake,[17] Sara P. Bombaci,[18] Julien Brun,[6,7] Jacob D. Buchanan,[19] K. Dana Chadwick,[20,21] Melissa S. Chapman,[22] Steven S. Chong,[6,7,23] Y. Anny Chung,[24] Jessica R. Corman,[25] Jannelle Couret,[26] Erika Crispo,[27] Thomas G. Doak,[28] Alison Donnelly,[29] Katharyn A. Duffy,[30] Kelly H. Dunning,[31] Sandra M. Duran,[32] Jennifer W. Edmonds,[33] Dawson E. Fairbanks,[34] Andrew J. Felton,[35] Christopher R. Florian,[36] Daniel Gann,[37] Martha Gebhardt,[38] Nathan S. Gill,[39] Wendy K. Gram,[40] Jessica S. Guo,[41] Brian J. Harvey,[42] Katherine R. Hayes,[43] Matthew R. Helmus,[44] Robert T. Hensley,[36] Kelly L. Hondula,[17] Tao Huang,[4,45] Wiley J. Hundertmark,[46] Virginia Iglesias,[1] Pierre-Andre Jacinthe,[47] Lara S. Jansen,[48] Marta A. Jarzyna,[49,50] Tiona M. Johnson,[51] Katherine D. Jones,[36] Megan A. Jones,[52] Michael G. Just,[53] Youssef O. Kaddoura,[15] Aurora K. Kagawa-Vivani,[54] Aleya Kaushik,[55] Adrienne B. Keller,[56] Katelyn B. S. King,[57] Justin Kitzes,[58] Michael J. Koontz,[1] Paige V. Kouba,[59] Wai-Yin Kwan,[60] Jalene M. LaMontagne,[61] Elizabeth A. LaRue,[62] Daijiang Li,[63,64] Bonan Li,[65] Yang Lin,[66] Daniel Liptzin,[67] William Alex Long,[68] Adam L. Mahood,[2] Samuel S. Malloy,[69] Sparkle L. Malone,[37] Joseph M. McGlinchy,[1] Courtney L. Meier,[36] Brett A. Melbourne,[70] Nathan Mietkiewicz,[71] Jeffery T. Morisette,[72] Moussa Moustapha,[73] Chance Muscarella,[34] John Musinsky,[36] Ranjan Muthukrishnan,[74] Kusum Naithani,[75] Merrie Neely,[76,77] Kari Norman,[22] Stephanie M. Parker,[36] Mariana Perez Rocha,[78] Laís Petri,[79] Colette A. Ramey,[3] Sydne Record,[80] Matthew W. Rossi,[1] Michael SanClements,[36] Victoria M. Scholl,[1,2] Anna K. Schweiger,[81] Bijan Seyednasrollah,[30] Debjani Sihi,[82] Kathleen R. Smith,[3] Eric R. Sokol,[36,83] Sarah A. Spaulding,[83] Anna I. Spiers,[1,70] Lise A. St. Denis,[1] Anika P. Staccone,[84] Kaitlin Stack Whitney,[85] Diane M. Stanitski,[55] Eva Stricker,[86] Thilina D. Surasinghe,[87] Sarah K. Thomsen,[88] Patrisse M. Vasek,[10] Li Xiaolu,[89] Di Yang,[90] Rong Yu,[29] Kelsey M. Yule,[91] and Kai Zhu[92]

[1]*Earth Lab, CIRES, University of Colorado Boulder, Boulder, Colorado, USA*
[2]*Department of Geography, University of Colorado Boulder, Boulder, Colorado, USA*
[3]*Biology Department, Metropolitan State University of Denver, Denver, Colorado, USA*
[4]*Human-Environment Systems, Boise State University, Boise, Idaho, USA*
[5]*University of New South Wales Sydney, Sydney, New South Wales, Australia*
[6]*National Center for Ecological Analysis and Synthesis (NCEAS), Santa Barbara, California, USA*
[7]*University of California Santa Barbara, Santa Barbara, California, USA*
[8]*School of Natural Resources & Environment, University of Florida, Gainesville, Florida, USA*
[9]*Ecological Society of America, Washington, D.C., USA*
[10]*Department of Math, Science, and Technology, Oglala Lakota College, Kyle, South Dakota, USA*
[11]*BIO5 Institute, University of Arizona, Tucson, Arizona, USA*
[12]*Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA*
[13]*Department of Civil and Mechanical Engineering, University of Missouri Kansas City, Kansas City, Missouri, USA*
[14]*CIRES, University of Colorado Boulder, Boulder, Colorado, USA*
[15]*Department of Forest, Fisheries and Geomatics Sciences, University of Florida, Gainesville, Florida, USA*
[16]*Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow, Idaho, USA*
[17]*National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, Maryland, USA*
[18]*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA*

[19]*Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio, USA*
[20]*Department of Geological Sciences, University of Texas Austin, Austin, Texas, USA*
[21]*Department of Integrative Biology, University of Texas Austin, Austin, Texas, USA*
[22]*Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, California, USA*
[23]*University of California Berkeley Library, University of California Berkeley, Berkeley, California, USA*
[24]*Departments of Plant Biology and Plant Pathology, University of Georgia, Athens, Georgia, USA*
[25]*School of Natural Resources, University of Nebraska Lincoln, Lincoln, Nebraska, USA*
[26]*Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island, USA*
[27]*Department of Biology, Pace University, New York City, New York, USA*
[28]*Department of Biology, Indiana University, Bloomington, Indiana, USA*
[29]*Department of Geography, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA*
[30]*School of Informatics, Computing & Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA*
[31]*School of Forestry and Wildlife, Auburn University, Auburn, Alabama, USA*
[32]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA*
[33]*Department of Physical and Life Sciences, Nevada State College, Henderson, Nevada, USA*
[34]*Department of Environmental Science, University of Arizona, Tucson, Arizona, USA*
[35]*Department of Wildland Resources, Utah State University, Logan, Utah, USA*
[36]*Battelle, National Ecological Observatory Network, Boulder, Colorado, USA*
[37]*Department of Biological Sciences, Florida International University, Miami, Florida, USA*
[38]*School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona, USA*
[39]*Department of Natural Resources Management, Texas Tech University, Lubbock, Texas, USA*
[40]*University Corporation for Atmospheric Research, Boulder, Colorado, USA*
[41]*College of Agriculture and Life Sciences, University of Arizona, Tucson, Arizona, USA*
[42]*School of Environmental and Forest Sciences, University of Washington, Seattle, Washington, USA*
[43]*Department of Integrative and Systems Biology, University of Colorado Denver, Denver, Colorado, USA*
[44]*Department of Biology, Temple University, Philadelphia, Pennsylvania, USA*
[45]*Cary Institute of Ecosystem Services, Millbrook, New York, USA*
[46]*Department of Earth and Environment, Boston University, Boston, Massachusetts, USA*
[47]*Department of Earth Sciences, Indiana University Purdue University, Indianapolis, Indiana, USA*
[48]*Department of Environmental Science & Management, Portland State University, Portland, Oregon, USA*
[49]*Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, Ohio, USA*
[50]*Translational Data Analytics Institute, The Ohio State University, Columbus, Ohio, USA*
[51]*Atlanta, Georgia, USA*
[52]*Boulder, Colorado, USA*
[53]*US Army, ERDC CERL, Champaign, Illinois, USA*
[54]*Department of Geography and Environment, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA*
[55]*National Oceanic and Atmospheric Administration, Boulder, Colorado, USA*
[56]*Department of Ecology, Evolution, and Behavior, University of Minnesota Twin Cities, St. Paul, Minnesota, USA*
[57]*Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, USA*
[58]*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[59]*Department of Plant Sciences, University of California Davis, Davis, California, USA*
[60]*CALeDNA, University of California Los Angeles, Los Angeles, California, USA*
[61]*Department of Biological Sciences, DePaul University, Chicago, Illinois, USA*
[62]*Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, USA*
[63]*Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA*
[64]*Center for Computation & Technology, Louisiana State University, Baton Rouge, Louisiana, USA*
[65]*Department of Biological & Ecological Engineering, Oregon State University, Corvallis, Oregon, USA*
[66]*Soil and Water Sciences Department, University of Florida, Gainesville, Florida, USA*
[67]*Soil Health Institute, Morrisville, North Carolina, USA*
[68]*Science and Technology Innovation Program, Woodrow Wilson International Center for Scholars, Washington, D.C., USA*
[69]*Battelle Center for Science, Engineering and Public Policy in the John Glenn College of Public Affairs, Ohio State University, Columbus, Ohio, USA*
[70]*Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, Colorado, USA*
[71]*CoreLogic, Irvine, CA, USA*
[72]*U.S. Department of Agriculture Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA*
[73]*Department of Biological Science, University of Ngaoundere, Ngaoundere, Adamawa, Cameroon*
[74]*Environmental Resilience Institute, Indiana University, Bloomington, Illinois, USA*
[75]*Department of Biological Sciences, University of Arkansas-Fayetteville, Fayetteville, Arkansas, USA*
[76]*GEO AquaWatch, Clearwater, Florida, USA*
[77]*Global Science and Technology, Inc, Greenbelt, Maryland, USA*
[78]*Department of Biology, University of Oklahoma, Norman, Oklahoma, USA*

[79]School for Environment and Sustainability, University of Michigan, East Lansing, Michigan, USA
[80]Department of Biology, Bryn Mawr College, Bryn Mawr, Pennsylvania, USA
[81]Remote Sensing Laboratories, Department of Geography, University of Zurich, Zurich, Switzerland
[82]Department of Environmental Sciences, Emory University, Atlanta, Georgia, USA
[83]INSTAAR, University of Colorado Boulder, Boulder, Colorado, USA
[84]Department of Ecology, Evolution, & Environmental Biology, Columbia University, New York, New York, USA
[85]Department of Science, Technology, and Society, Rochester Institute of Technology, Henrietta, New York, USA
[86]Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA
[87]Department of Biological Sciences, Bridgewater State University, Bridgewater, Massachusetts, USA
[88]Department of Integrative Biology, Oregon State University, Corvallis, Oregon, USA
[89]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York, USA
[90]Wyoming GIS Center, University of Wyoming, Laramie, Wyoming, USA
[91]Biodiversity Knowledge Integration Center, Arizona State University, Tempe, Arizona, USA
[92]Department of Environmental Studies, University of California, Santa Cruz, Santa Cruz, California, USA

**Citation:** Nagy, R. C., J. K. Balch, E. K. Bissell, M. E. Cattau, N. F. Glenn, B. S. Halpern, N. Ilangakoon, B. Johnson, M. B. Joseph, S. Marconi, C. O'Riordan, J. Sanovia, T. L. Swetnam, W. R. Travis, L. A. Wasser, E. Woolner, P. Zarnetske, M. Abdulrahim, J. Adler, G. Barnes, K. J. Bartowitz, R. E. Blake, S. P. Bombaci, J. Brun, J. D. Buchanan, K. D. Chadwick, M. S. Chapman, S. S. Chong, Y. A. Chung, J. R. Corman, J. Couret, E. Crispo, T. G. Doak, A. Donnelly, K. A. Duffy, K. H. Dunning, S. M. Duran, J. W. Edmonds, D. E. Fairbanks, A. J. Felton, C. R. Florian, D. Gann, M. Gebhardt, N. S. Gill, W. K. Gram, J. S. Guo, B. J. Harvey, K. R. Hayes, M. R. Helmus, R. T. Hensley, K. L. Hondula, T. Huang, W. J. Hundertmark, V. Iglesias, P.-A. Jacinthe, L. S. Jansen, M. A. Jarzyna, T. M. Johnson, K. D. Jones, M. A. Jones, M. G. Just, Y. O. Kaddoura, A. K. Kagawa-Vivani, A. Kaushik, A. B. Keller, K. B. S. King, J. Kitzes, M. J. Koontz, P. V. Kouba, W.-Y. Kwan, J. M. LaMontagne, E. A. LaRue, D. Li, B. Li, Y. Lin, D. Liptzin, W. A. Long, A. L. Mahood, S. S. Malloy, S. L. Malone, J. M. McGlinchy, C. L. Meier, B. A. Melbourne, N. Mietkiewicz, J. T. Morisette, M. Moustapha, C. Muscarella, J. Musinsky, R. Muthukrishnan, K. Naithani, M. Neely, K. Norman, S. M. Parker, M. Perez Rocha, L. Petri, C. A. Ramey, S. Record, M. W. Rossi, M. SanClements, V. M. Scholl, A. K. Schweiger, B. Seyednasrollah, D. Sihi, K. R. Smith, E. R. Sokol, S. A. Spaulding, A. I. Spiers, L. A. St. Denis, A. P. Staccone, K. Stack Whitney, D. M. Stanitski, E. Stricker, T. D. Surasinghe, S. K. Thomsen, P. M. Vasek, L. Xiaolu, D. Yang, R. Yu, K. M. Yule, and K. Zhu. 2021. Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community. Ecosphere 12(12):e03833. 10.1002/ecs2.3833

**Abstract.** It is a critical time to reflect on the National Ecological Observatory Network (NEON) science to date as well as envision what research can be done right now with NEON (and other) data and what training is needed to enable a diverse user community. NEON became fully operational in May 2019 and has pivoted from planning and construction to operation and maintenance. In this overview, the history of and foundational thinking around NEON are discussed. A framework of open science is described with a discussion of how NEON can be situated as part of a larger data constellation—across existing networks and different suites of ecological measurements and sensors. Next, a synthesis of early NEON science, based on >100 existing publications, funded proposal efforts, and emergent science at the very first NEON Science Summit (hosted by Earth Lab at the University of Colorado Boulder in October 2019) is provided. Key questions that the ecology community will address with NEON data in the next 10 yr are outlined, from understanding drivers of biodiversity across spatial and temporal scales to defining complex feedback mechanisms in human–environmental systems. Last, the essential elements needed to engage and support a diverse and inclusive NEON user community are highlighted: training resources and tools that are openly available, funding for broad community engagement initiatives, and a mechanism to share and advertise those opportunities. NEON users require both the skills to work with NEON data and the ecological or environmental science domain knowledge to understand and interpret them. This paper synthesizes early directions in the community's use of NEON data, and opportunities for the next 10 yr of NEON operations in emergent science themes, open science best practices, education and training, and community building.

**Key words:** community; continental-scale ecology; diversity; inclusion; National Ecological Observatory Network; open data; open science; Special Feature: Harnessing the Neon Data Revolution.

**Received** 31 March 2021; revised 4 June 2021; accepted 24 June 2021. Corresponding Editor: Debra P. C. Peters.

† **E-mail:** nagyrc@gmail.com

## Introduction

### *Summary of foundational thinking around NEON*

Environmental challenges facing today's society require thinking across scales from local to continental or global, data collected across different ecoregions and over decades, multidisciplinary expertise, team science approaches, and training in skills like computer and data science (Keller et al. 2008, Schimel 2011). A fundamental goal of the National Ecological Observatory Network (NEON) is to improve our understanding of and ability to predict the effects of environmental change (e.g., climate change, land-use change, biological invasions, altered nutrient cycling) at continental scales representing both terrestrial and aquatic ecosystems (Field et al. 2006, Schimel et al. 2011, Kao et al. 2012). The scientific infrastructure of NEON was designed to meet this goal by using a standardized, multiscale sampling strategy consisting of systematically deployed aquatic, ground, and tower-based sensors, field sampling, and high-resolution airborne remote sensing (Schimel et al. 2007, Kampe et al. 2010, Taylor et al. 2011). NEON delivers a coordinated and standardized set of calibrated and documented data on key plant and animal taxa as well as microbes and algae, environmental and atmospheric variables, and remote sensing data across the United States (Field et al. 2006, Kampe et al. 2010). These data can help reveal linkages between ecological patterns and processes across scales and identify drivers of change and the resultant ecological consequences (Atkins et al. 2018, Read et al. 2018, Hall et al. 2020, Marconi et al. 2021). The Observatory design (including 81 sites across 20 ecoclimatic domains; NEON 2021*a*) and sampling protocols (Keller et al. 2008, Kao et al. 2012) will capture temporal scales across local, regional, and continental spatial extents to enable meaningful connections to satellite remote sensing, geospatial, and other network data. The transformational potential of NEON as a highly integrated scientific observatory was recognized early on in its deployment (Balch et al. 2020*b*), as were the challenges of working with disparate types of ecological and environmental data. Realizing the full potential of NEON will require that data are easy to access and use by scientific and educational communities. However, NEON data will not be able to answer all questions; it will not replace the need for field ecology or skills to conduct hypothesis-driven, experimental research (e.g., study design, data collection), the value of an intimate understanding of a particular organism or ecosystem, or the utility of other individual sites, networks, and data sources, but rather can be used in conjunction with all of these other existing sources of information.

### *History of NEON*

The necessity for a long-term, geographically widespread ecological observatory network with consistent data collection had long been recognized by the scientific community. Workshops in 2000–2005 led to an initial plan for NEON with an ambitious 30-yr timeline (NEON 2021*b*). The following five years were dedicated to planning and designing, and in 2011, the National Science Foundation (NSF) approved funds to build NEON (NEON 2021*b*). The construction of NEON sites was completed in 2019 (NEON 2021*b*). While some data products at some sites were available as early as 2010, it was not until 2019 that all sites had each data product available. Thus, after almost 20 yr of envisioning, planning, and construction, NEON is now fully operational. The challenges of running an operation at this scale are significant and have been noted elsewhere (Mervis 2015, Cesare 2016, Collins and Knapp 2019, Rogers 2019). While some of the community has been hesitant or resistant to embrace NEON due to the steep learning curve and other challenges (Sagoff 2019), some of the community is eager to use NEON data because of the potential that this large NSF investment offers.

Some individual researchers and lab groups began analyzing components of NEON data even before the full scope of the Observatory was complete (Anderegg and Diffenbaugh 2015, Ghabbour et al. 2015, Read et al. 2018, Scholl et al. 2020), yet further efforts were needed to build a cohesive NEON user community. Furthermore, while there is an understanding in much of the community about the power of the data being produced by the Observatory, many potential users may face barriers to utilizing these data sets (Balch et al. 2020*b*). Therefore, Earth Lab at the University of Colorado Boulder hosted the first NEON Science Summit in 2019 to continue to build a robust and sustainable NEON user community, which is essential for the Observatory to realize its full potential.

### *2019 NEON Science Summit*

The NEON Science Summit, held at the University of Colorado Boulder in October 2019, was the first "unconference" (a meeting with participant-driven agenda and working group topics) focused on building new science from NEON data products. In total, there were 170 participants through a mix of in-person and remote participation. Throughout two-and-a-half days, ~15 breakout working groups used NEON data to explore questions such as: What are the environmental drivers of microbial community composition across sites (see Qin et al., *in this special issue*)? How can ground, uncrewed aerial systems (UAS), airborne, and satellite data at NEON sites be linked for applications such as detecting and segmenting individual trees (see Koontz et al. and Gann et al., *in this special issue*)? Does the relationship between native and non-native species richness change with spatial scale (see Gill et al., *in this special issue*)? This paper synthesizes the work from the 2019 NEON Science Summit and the Grand Challenges that the NEON user community identified as priority areas to address.

## NEON Open Science

Open science principles and methods (e.g., making samples, data, workflows, software, publications freely available) are changing the field of ecology (Hampton et al. 2015). As a key tenet of the NEON mission at the outset, this commitment to open science has the potential to accelerate ecological research and increase the diversity of scientists involved by removing barriers to access. The diversity and number of data products (NEON 2021*c*), tutorials (NEON 2021*d*), and analytic tools (neonUtilities (Lunch et al. 2021) and geoNEON (NEON 2020) packages in R) that NEON provides are a key resource for open ecological research. In addition, NSF requires all scientists funded by Macrosystems Biology and NEON Enabled Science grants to archive their data with the Environmental Data Initiative (EDI; EDI 2021) to promote data discovery and use. An extended commitment of the scientists using these resources to make their data, code, and workflows open will increase efficiency and facilitate greater coordination across a larger collaborative community. Key opportunities that are expected to bring added value to open NEON data include the following: harmonization with other observation networks (such as the Long-Term Ecological Research (LTER), Long-Term Agroecosystem Research (LTAR), Critical Zone Observatory (CZO), AmeriFlux, USA-NPN National Phenology Network, and others) and data sources, open science contributions from the NEON user community, and facilitation, training, and curation that lead to a robust and popular NEON software toolbox.

### *Data harmonization to answer continental-scale ecology questions*

*The why of linking data.*—Participants of the 2019 NEON Science Summit identified that network-to-network data collaborations are critical for continental-scale ecology (Balch et al. 2020*b*; SanClements et al., *in this special issue*). NEON has existing collaborations with AmeriFlux, the PhenoCam network, the LTER network, the LTAR network, the National Phenology Network, and others. These collaborations leverage multiple sources of spatial and temporal resolution and thus improve our ability to understand complex processes, phenomena, and change over space and time compared to individual networks. Because of its spatial and disciplinary breadth, NEON is well poised to act as a central hub in a network of networks (SanClements et al., *in this special issue*). Expanding these partnerships would be beneficial to NEON and the whole community.

*Broadening our networks and multi-scale analysis.*—NEON data can be used in conjunction with data from other networks at co-located sites, to expand sites in spatial extent, to standardize data collection protocols, or to synthesize complementary data products across networks. Existing long-term (i.e., decadal) ecological networks (e.g., CZO, LTER) at co-located sites provide inferential power and historical context for contemporary patterns observed by NEON (Hinckley et al. 2016*b*). For example, long-term experiments at LTER sites manipulate some of the drivers observed by NEON, contextualizing patterns observed at NEON sites (Jones et al. 2021). Other continental-scale monitoring efforts (e.g., community or citizen science data, North American Breeding Bird Survey, eBird) can fill in gaps between NEON sites and expand the spatial reach of NEON data. NEON also carries the potential to contribute to global-scale networks such as GLEON (Global Lakes Ecological Observatory Network) and GEO BON (Group on Earth Observations Biodiversity Observation Network). NEON collaborated with the USA National Phenology Network to ensure that data collection protocols between the two networks were standardized from NEON's inception. A recent collaboration between the Environmental Data Initiative (EDI), LTER, and NEON resulted in a harmonized data model (ecocomDP) for community ecology observations that provides an analysis ready data product for synthesis of community ecology data sets across the LTER and NEON (O'Brien et al., 2021; Record et al. 2021; Li et al., *in this special issue*).

NEON site data and Airborne Observation Platform (AOP) observations can be used synergistically for calibration and validation and inform existing and future missions/networks. The ability to task NEON resources in support of these studies will continue to expand the utility of the NEON program and synergies. For example, the AOP was tasked in 2018 in support of the Department of Energy's ongoing Watershed Function Science Focus Area (Chadwick et al. 2020). This project produced publicly available data sets and functional trait models that are now being used for a wide range of studies, including assessment of sensitivity requirements for NASA's Surface Biology and Geology (SBG) Designated Observable (Cawse-Nicholson et al. 2021; Thompson et al., *unpublished manuscript*). A priority of the SBG program is to understand the global distribution of vegetation functional types and traits. Precursor studies like this one will allow NEON AOP studies to inform SBG architecture. Over the long term, together, imaging spectroscopy satellite missions and NEON data will provide repeat observations of chemical properties of vegetation, aquatic biomass, and soils at variable, and complimentary, resolutions and repeat times.

NEON data can be extended vertically and horizontally with UAS, airborne, and satellite-based observations to answer questions requiring multi-scale observations and analysis. For example, following on Schimel et al. (2019), plot data and flux tower data can be integrated with UAS, airborne, and satellite observations to obtain productivity and carbon measurements from the eddy covariance footprint to continental scales. The multi-temporal observations from NEON data are also primed for trend analyses by integrating with data from Landsat, Landsat/Sentinel harmonization, Moderate Resolution Imaging Spectroradiometer (MODIS), Global Ecosystem Dynamics Investigation (GEDI), and long-term monitoring plots from other networks to understand vegetation dynamics from disturbances (e.g., fire, beetle kill) and soil dynamics from experimental manipulations (Wieder et al. 2020). Lastly, there is an opportunity to utilize UAS data for field validation and scaling to observations at NEON sites, as well as non-NEON sites through NEON's assignable assets program (NEON 2021*e*).

### *Best practices for sharing data, code, software, and entire workflows*

All components of NEON are documented and intended to be reused as community standards for data collection and processing. The science carried out using NEON data must be open and reproducible, encouraging the creation of online space to store, review, and share tools and software to build upon each other's efforts. Yet successfully building a community of scientists who share code, software, and data products built upon NEON data is one of the key challenges identified during the Summit and requires community adoption of best practices (Hey 2009, Bechhofer et al. 2013).

*Open software, tools, and code to support efficient open science.*—In modern science, software underlies a majority of science outputs; it is critical that researchers strive for openness in their work to the degree possible. This includes making code accessible with permissive licensing (Dabbish et al. 2012, Loeliger and McCullough 2012), citable using DOIs, documented and maintained over time to support reproducibility and re-use. Although these approaches are historically not part of traditional science training, organizations such as rOpenSci (rOpenSci 2021) and pyOpenSci (pyOpenSci 2021) provide community support and a peer-review process to ensure both citation credit for software developed and high-quality scientific tools. The review process is further supported by the Journal of Open Source Software and Methods in Ecology and Evolution to simplify publication avenues. Tools and workflows for working with NEON data can make use of these open science principles to advance scientific discovery by building on collaborative knowledge. An open science approach will also ensure continued improvements of NEON-specific analytical tools, expansion of the macroecological knowledge base, and systematically address knowledge gaps.

*Open data and processing pipelines.*—The volume and variety of data produced by the NEON user community are significant and require the adoption of best practices for sharing, curating, and archiving data workflows and derived products following FAIR data principles (Wilkinson et al. 2016). To make the greatest use of NEON and extended data sources, data are given structure in the form of a schema or index and must have community established metadata templates, for example, DublinCore (Weibel et al. 1998), Ecological Metadata Language (EML; Fegraus et al. 2005), or DarwinCore (DwC; Wieczorek et al. 2012). Derived data can be hosted on public repositories such as Dryad (Isard et al. 2007), Pangaea (Pangea 2021), the EDI Data Portal (EDI 2021), Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE; Varadharajan et al. 2019), or CyVerse Data Commons (CyVerse 2021).

When extended to the regional and continental scale over many years and decades, the computational and management specifications of NEON data require analyses on distributed, scalable cyberinfrastructure. While some researchers pay for computing and storage services via platforms like Amazon Web Services, limited access to computational resources by students or researchers at small and underserved institutions can be remedied by using free and open scientific cyberinfrastructure resources. High-performance computing (HPC), high-throughput computing (HTC), and cloud services are freely available to U.S.-based researchers via the eXtreme Scientific and Engineering Discovery Environment (XSEDE; Towns et al. 2014) and CyVerse (Merchant et al. 2016, Swetnam et al. 2016, Bucksch et al. 2017). Additionally, privately operated CyberGIS platforms, such as Google Earth Engine (GEE; Gorelick et al. 2017), re-host some public Earth observation system (EOS) satellite and aerial remote sensing data and provide a resource for exploratory data analysis. Workflow Management Systems (WMS) enable researchers to analyze vast quantities of data on distributed cyberinfrastructure. Exemplar WMS, such as Pangeo (Eynard-Bontemps et al. 2019), SnakeMake, Makeflow, and WorkQueue (Albrecht et al. 2012, Köster and Rahmann 2012, Zheng and Thain 2015), are used by the Life and Earth science communities for analyzing massive corpuses of scientific data and could be adopted by NEON users. Processing NEON's environmental, soil metagenomic data, or AOP data across the entire observatory requires scalable computing which are most easily accomplished via a WMS (Thessen et al. 2020).

### *The NEON toolbox will empower an open, collaborative NEON community*

The scientific community has already built many different tools and products to make workflows for processing and analyzing NEON data more efficient (NEON 2021*f*); however, these efforts are widespread, often disconnected, and the tools developed are not easily discoverable. Further, the tools are often developed for single project use and thus not generalized to support the broader community of NEON data users (but see Li et al., *in this special issue*). There are several potential advantages to sharing resources among NEON users including (1) reduced redundancy in efforts as groups independently develop tools with similar functionality; (2) adoption of methods or algorithms with novel approaches better suited for tackling questions with the available volume of data rather than those traditionally used by a specific community; and (3) lower investment of skills, resources, and time for individual groups that do not have pre-existing tools and workflows. Sharing derived data, tools, and software that users generate to draw greater meaning from NEON public data products will significantly improve the use and scope of NEON in addressing emerging scientific challenges across the continent.

Existing efforts could be leveraged to make NEON tools more discoverable. For example, rOpenSci (rOpenSci 2021) and pyOpenSci (pyOpenSci 2021) support software peer review and community development of citable, tested, well-documented, and discoverable software. Along with software and tools, the user community will also be adding packages to more easily visualize, download, and redistribute portions of data products from pipelines that would require computational architecture too expensive to run on local computers. Derived data sets can be stored in open-access repositories (e.g., Zenodo, Dryad, Figshare, or EDI) with options for version control and DOIs. Although NEON will not directly maintain these software, tools, or data, it can play a central role in making them discoverable by the community.

## The Intellectual Merit of NEON Science: Status and Future

The ecological community has already taken advantage of existing NEON data products and NEON Biorepository samples and specimens. As of October 2020, 267 publications have described, referenced, and used NEON data and network resources. While the range of topics varies greatly, drawing on the 181 open-access data products and 63 collections of physical samples, certain key themes have emerged (Fig. 1). The large emphasis on data suggests how valuable these products have been for the ecological community. Prominent topics include tracking phenology changes, forest structural dynamics and tree classification, soil organic matter (or carbon) dynamics, ecological forecasting, and small mammal biodiversity patterns. For example, early work has used tree classification and mapping techniques, and convolutional neural networks with combinations of hyperspectral imagery, lidar, and RGB, to identify tree species and individual tree crowns (Dalponte et al. 2019, Fricker et al. 2019*b*, Weinstein et al. 2019, Scholl et al. 2020). Others have examined changes in plant phenology in deciduous forests (Seyednasrollah et al. 2020) and alpine systems (Dorji et al. 2020) in a changing climate.


Fig. 1. Word cloud created from article titles from 180 NEON-related publications from 2017 to 2020.

Researchers also evaluated continental to global-scale dynamics of soil carbon by leveraging NEON resources (Kramer and Chadwick 2018, Hall et al. 2020).

The NEON Science Summit Steering Committee surveyed the work funded through NSF Macrosystems (and other agencies/programs) and papers that have used NEON data to inform our virtual breakout groups preceding the Science Summit. These preliminary calls were then used to distill areas of interest among the 170 participants. While much thought has been given to what will be possible with NEON data over the course of decades (e.g., tracking phenological changes resulting from climate change), the focus of the 2019 NEON Science Summit was to determine what can be done with NEON data and NEON-compatible products right now and in the next decade. This section synthesizes the main areas that Summit participants are addressing through individual efforts and group work that stemmed from the Summit. Topics range across fundamental ecology and contemporary problems such as response to environmental change among species, communities, and ecosystems. There were also some emergent themes that may not have been anticipated in the foundational thinking around NEON.

### *Fundamental ecology*

Ecology as a discipline aims to understand and predict how biotic and abiotic features of the environment interact. Fundamental questions span multiple scales of biological organization, ecosystem processes, space, and time. A survey of ecologists identified 100 fundamental yet unanswered ecological questions (Sutherland et al. 2013), covering the basic understanding of eco-evolutionary feedbacks, processes driving population, community, and biodiversity patterns, species interactions and invasion dynamics, unique qualities of disease and microorganisms, ecosystem functioning, and human interactions. NEON is well poised to address many of these questions with its spatially nested hierarchical design and systematic sampling of Earth's biotic and abiotic components across North America (Schimel et al. 2007, Keller et al. 2008, Schimel and Keller 2015).

The NEON Science Summit identified several key fundamental areas in ecology that NEON data can be used to address now: (1) testing of our understanding of ecological patterns and processes across spatial and temporal scales; (2) determining the drivers of biodiversity patterns across the United States; and (3) documenting ecosystem processes and the services that nature provides. Some examples of basic questions that arose from the Summit include the following: Which ecological patterns and dynamics scale up and/or down in time and space? What controls decomposition? How generalizable are ecological observations from single sites to the continental scale? What drives trait variation? Which ecological patterns and processes are more context-dependent or species-specific, and which are more generalizable?

### *Biodiversity across scales*

Understanding temporal and spatial scales at which drivers affect biodiversity is necessary to inform robust modeling of historical, current, and future patterns (Delsol et al. 2018, Gonzalez et al. 2020). Spatial synchrony of populations and communities is an emerging area of research that can help determine the degree of resilience and resistance of biota to environmental change (Zelnik et al. 2018). Central to quantifying synchrony is identifying the appropriate scales at which ecological systems function. NEON's hierarchical design—from plots to sites to domains—enables investigations of the drivers of biodiversity within and among spatial scales. For example, at continental scales, NEON data on small mammals and intraspecific body size variation helped reveal the role of biotic interactions and climate in mediating patterns of community composition and trait plasticity across NEON domains (Read et al. 2018). This research has contributed to greater understanding of the biotic mechanisms that drive the Latitudinal Diversity Gradient (Read et al. 2018). When combined with remotely sensed data (e.g., airborne and satellite imagery) and organismal data across the United States (e.g., North American Breeding Bird Survey, USDA Forest Inventory & Analysis), NEON's organismal and Airborne Observation Platform (AOP) data will further advance continental-scale assessment of biodiversity patterns. The variety of taxa sampled by NEON allows investigators to seek generalities in terms of the spatial and temporal scales of drivers, while also accounting for taxon-specific life-history characteristics and traits as well as other biological constraints (e.g., regional species pools and biogeographic filters) (Kao et al. 2012).

Researchers are already tackling the scales of biodiversity drivers and responses using NEON data and asking such questions as: What can NEON data tell us about the spatial synchrony of populations and communities at the continental scale? What are the drivers responsible? How do temporal dynamics (e.g., shifts in temperature, precipitation, discharge) affect biodiversity patterns in aquatic communities across the United States? How do different metrics of biodiversity (e.g., species richness) relate to structural (LaRue et al. 2019), functional, and spectral diversity? Future directions could include mapping changes in biodiversity in a changing climate. Where will the greatest losses of biodiversity occur and how can this inform conservation and management?

### *Evaluating disturbance dynamics with NEON*

Disturbance plays an important role in the structure and function of terrestrial and aquatic ecosystems (Thom and Seidl 2016, Daam et al. 2019). The impacts of disturbance on an ecosystem are controlled by its intensity, frequency, size, spatial pattern, and spatial extent. Some disturbances like wildfires can occur multiple times during an ecosystem's response period, affecting the resilience and recovery of the system (Bartowitz et al. 2019; Mahood & Balch 2019). NEON's sampling design, including repeat sampling covering the full range of U.S. ecological and geo-climatic diversity, helps evaluate the impacts of these frequent disturbance events. NEON's high spatial and temporal resolution airborne remote sensing data (e.g., lidar, hyperspectral, and photogrammetry) allow mapping patterns of disturbance within and across biomes (Fig. 2), though the extent to which NEON captures disturbance history for various disturbances has yet to be examined. For example, the extent of bark beetle attack, forest fire, or flood/hurricane damage can be estimated using remotely sensed vegetation structure and function over the disturbed area. Further, the validation of remotely sensed data with coincident ground-based data can facilitate more accurate estimation of disturbance scale. Finally, NEON recently released a "Site management and event reporting" data product (DP1.10111.001) which includes on-the-ground documentation of disturbance events as well.

The wealth of NEON data includes detailed field measurements about plants, animals, soil, microbes, nutrients, freshwater, and the atmosphere that can be used to advance the understanding of fine-scale variation of ecosystems under variable conditions. The combination of sites for long-term observation and assignable assets that move through time further enables the capture of ecosystem dynamics in response to disturbance. The multi-scalar observations from NEON can provide a macrosystem view of disturbance dynamics and ecological patterns. For the disturbances that NEON sites do not capture well, metrics of impact and/or environmental correlates (e.g., ecohydrological variables that indicate fluvial disturbances) can be determined by combining NEON and associated data sets. Furthermore, NEON data provide a unique opportunity to improve forecasts of future disturbances (e.g., extreme weather events such as droughts, spread of invasive species) and inform management recommendations at the continental scale.

### *Carbon and climate dynamics*

A major opportunity for NEON data use is exploring biogeochemical feedbacks in the climate system (e.g., carbon–climate feedbacks). Baseline C storage and fluxes (e.g., productivity and decomposition) across ecoregions can now be assessed, and processes directly tied to climatic gradients and vegetation cover across the United States can be evaluated. Further, it is possible to examine how other nutrient pools and dynamics in soils and plants affect C fluxes across NEON sites. Patterns in soil chemistry (e.g., C and nitrogen (N) concentrations, net N mineralization and nitrification rates, C and N isotopes), as well as changes in soil C stability at individual sites and across the United States, can be examined in a systematic and repeated design not previously available (Hinckley et al. 2016*a*, Weintraub et al. 2017).

In the long term, the decades of data collected by NEON on vegetation cover, aboveground biomass, soil physical and chemical properties, litterfall and fine woody debris production, litter chemical properties, root biomass and chemistry, soil carbon dioxide concentrations, and climate

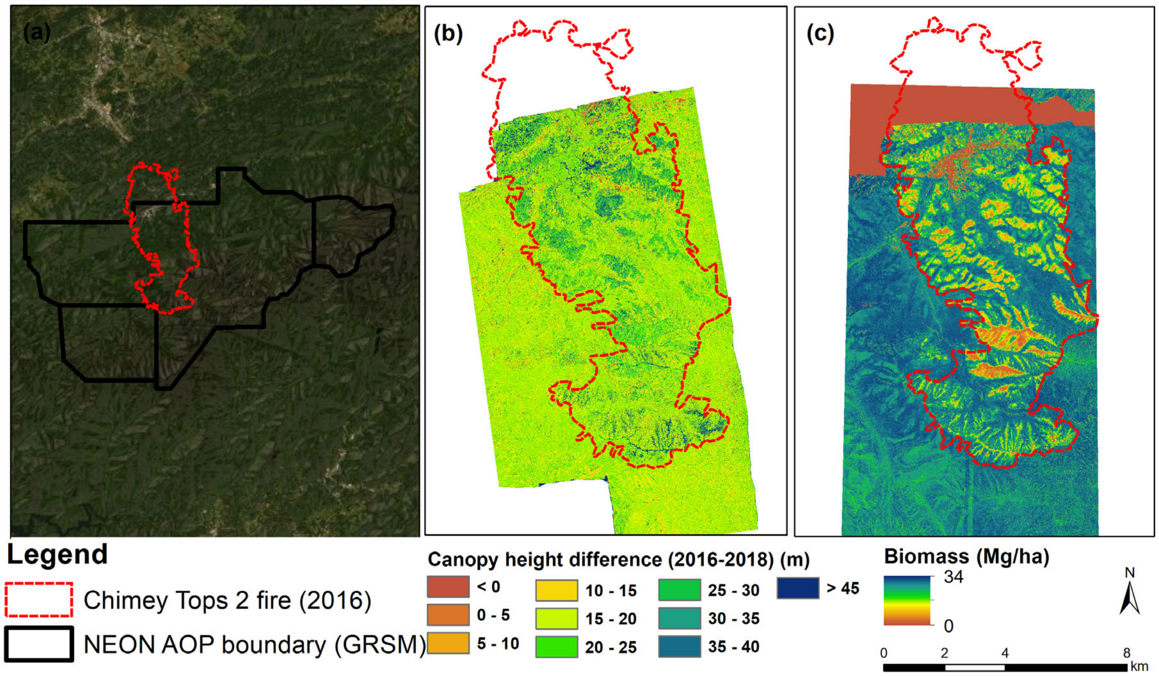


Fig. 2. NEON AOP captures impact of disturbances at high spatial resolution. (a) Chimney Tops 2 fire boundary (2016) at Great Smoky Mountain, Twin Creek (GRSM) NEON site. (b) Canopy height difference calculated using NEON lidar canopy height models from the years 2016 (pre-fire) and 2018 (post-fire). (c) Biomass (2017) estimated from correlation with NDVI and LAI parameters derived from NEON AOP data.

will offer a site-by-site detailed history to track climate change and carbon–climate feedbacks. These data will be extremely valuable for supporting climate action and related decision-making. Furthermore, the capacity of ecosystems to sequester C can be better defined to improve future climate modeling scenarios (Kramer and Chadwick 2018). The timing of vegetation phenology and the duration of the growing season play a key role in determining the amount of C sequestered by different ecosystems. Therefore, the phenological data being recorded at NEON sites will be critical in the development of accurate C budgets across ecosystems.

### *Situating NEON in a coupled human–environment systems framework*

The growing human population and associated activities (e.g., agricultural expansion, urbanization, transportation) across the globe influence nearly all patterns and functions of natural systems, while changing natural systems directly impact the socioeconomic and physical well-being of humans (Fedele et al. 2017). Thus, human and natural systems must be considered as a coupled social–environmental system (SES; Liu et al. 2007, Grimm et al. 2017, Balch et al. 2020*a*). The NEON network was designed to capture environmental rather than social gradients. However, the integration of local environmental observations and multi-temporal remote sensing data from the NEON observatory, combined with socioeconomic data, could enable understanding of human–environment interactions (Pricope et al. 2019). Research Coordination Networks, such as the newly funded project, "Landscape Exchange Network for Socio-environmental systems research (LENS)," which will leverage detailed observations from the NEON AOP to study SES across the United States, can improve our capacity to use NEON toward SES research more broadly. NEON sites subjected to active land management can also inform future land-use planning and adaptive natural resource management in the Anthropocene. Additionally, NEON's assignable assets, or rapidly deployable mobile suites of NEON instruments, could be used to target processes of interest in human–environmental systems. Ultimately, this knowledge can be integrated into management solutions to help society to adapt to environmental change. The ecological forecasting, land-use change, and climate change science that NEON enables in the long term involve tight coupling between humans and the environment and can bridge natural and social sciences.

### *Ecological forecasting*

Ecology is increasingly turning to ecological forecasting to aid in preparing for the future drivers of change that affect ecosystems, species, and communities. An important role for NEON data is to establish baselines of biodiversity, biogeochemical pools and fluxes, ecological structure, and other indicators to initialize forecasts and compare against future scenarios. NEON is taking the pulse of changing U.S. ecosystems and helping us predict their future health. As the field of ecological forecasting, or eco-forecasting, moves forward, NEON-enabled science can contribute iterative, probabilistic projections (Dietze and Lynch 2019). A step toward this goal of community members building iterative, near-term forecasts with NEON data is the NEON Ecological Forecast Challenge, the first round of which occurred in 2020, put on in collaboration with the Ecological Forecasting Initiative (EFI 2021). Ultimately, the aim is to develop the capability to anticipate or forecast ecological change to better prepare for, adapt to, or prevent change.

### *Emergent themes*

NEON-enabled science has tackled some classic questions in ecology, and the Science Summit suggested some unique, and perhaps unanticipated, possibilities for NEON observations data (in combination with other data) that have not been possible until now (Box 1). A basic opportunity for new insights built into the NEON design stems from the integrated, co-located measurements of many processes with NEON tower data collection. In this way, independent research can capitalize on the NEON infrastructure to add additional measurements and experiments at NEON sites. Other new efforts include researchers mapping individual trees and tree species using high-resolution airborne and satellite data (Weinstein et al. 2019, Scholl et al. 2020). And several novel questions examine scale: a multi-scale understanding of the carbon cycle by integrating data sources from field plot data to satellite imagery, better understanding the changing predictors of canopy height across scales (Fricker et al. 2019*a*), and how biodiversity patterns change across temporal scales.

The potential to address cutting-edge, multidisciplinary science questions with NEON (and NEON-compatible) data is unprecedented. Although NEON was envisioned as a stand-alone

**Box 1.**

The following are the emergent topics that are possible to address now using NEON (and other) data products, by category of questions.

*Foundational ecology*

1. What controls metabolic rates?
2. What drives trait variation?
3. What are the patterns in biodiversity across taxa —from beetles to trees?
4. What controls decomposition or transpiration rates?

*Species from space*

1. How can tree species of individual tree crowns be identified?
2. How do spectral signatures correspond to leaf traits?

*Change detection and forecasting*

1. What's the phenological response to interannual variability in climate?
2. How is the hydrology of streams changing?
3. What would it take to forecast ecological processes?

*Invasive species*

1. How does plasticity of the genome level predict plant success, across native and invasive species?

*Data harmonization and scaling*

1. How can different types and sources of data be integrated?
2. How does ecological pattern and process scale?

observatory, it now sits within a data ecosystem with many points of leverage. Linking across data sets and networks of observatories has begun, including in some ways not anticipated in early thinking around NEON. While one goal of NEON is to detect change over time, which will become possible with decades of data, early NEON and NEON-linked science is already used for change detection. For example, using NEON data to understand vegetation recovery after a fire requires integration of additional data sets (e.g., Landsat, MODIS, UAS) since only a couple of the NEON sites have experienced a fire. Other interdisciplinary opportunities abound including use of NEON data for ecohydrological studies like hierarchical clustering of surface water chemistry across NEON sites (see Edmonds et al., *in this special issue*) and soil data for pedological research on biogeochemical fluxes. Finally, opportunities exist to advance NEON technology as innovations in instrumentation, observing platforms, and data science surpass what the observatory initially offered (e.g., unmanned aerial vehicles (UAVs)). Engagement by a diverse community of investigators enhances this opportunity for continued scientific advancement.

## Building the NEON Community

### *Diversity, inclusion, and accessibility in STEM*

*How NEON can serve a diverse and inclusive community.*—Building a diverse and inclusive community expands professional opportunities for NEON users by encouraging the creation of a more accessible user experience. There is the potential to mobilize NEON's vast resources to grow such a community while also addressing persistent disparities in STEM participation. While open NEON infrastructure and data are central draws (and by themselves remove a central barrier of the resources needed to collect empirical data that many researchers lack), training opportunities (e.g., workshops that provide hands-on experience with the data) and networks of people are critical for building a diverse and inclusive community. These resources can be incentives for the larger community of environmental scientists who have yet to join the NEON community. For example, a training exercise might provide early-career or underrepresented scientists the opportunity to build their skill sets while also becoming a resident NEON data expert at their home institution. Similarly, these scientists may find new professional linkages and interests within the NEON user network.

*Diversity.*—Diverse backgrounds are critical to developing innovative and creative scientific perspectives and will expand the visibility of NEON beyond the current ecological community. Diversity can be measured in many different ways, but attracting participation from groups that have smaller representation in science and engineering fields than their representation in the U.S. population will increase diversity among NEON users. These groups include women, people with disabilities, and underrepresented racial/ethnic groups including Black, Indigenous, and People of Color (BIPOC). Post-event anonymous surveys following the 2019 NEON Science Summit indicated that 31.3% identified as non-white and 55% identified as female. Furthermore, developing mechanisms to attract users at different career stages, including students (K-12, undergraduate, and graduate); early-, mid-, and late-career scholars; educators and public outreach professionals; and professionals outside of academia, will also diversify and enrich the user community. During the NEON Science Summit, participants were asked to consider the question: "Who isn't here and how can they be brought into the fold?" Many of the responses went beyond identification of racial and ethnic minorities and emphasized the importance of building a coalition of NEON data users that includes a wide range of perspectives and members beyond the academic ecological research community (Fig. 3).

*Inclusion.*—The NEON user community can build on previous efforts to increase diversity with targeted outreach, recruitment, and training to groups at underrepresented institutions, such as Historically Black Colleges and Universities (HBCUs), Hispanic Serving Institutions (HSIs), and Tribal Colleges and Universities (TCUs). The Environmental Data Science Inclusion Network (EDSIN) was created to foster an online community to develop training opportunities, shared resources, and leadership experience for scientists who are traditionally underrepresented in data science and environmental science. Inclusive practices not only increase opportunities for underrepresented populations, they also foster diversity in

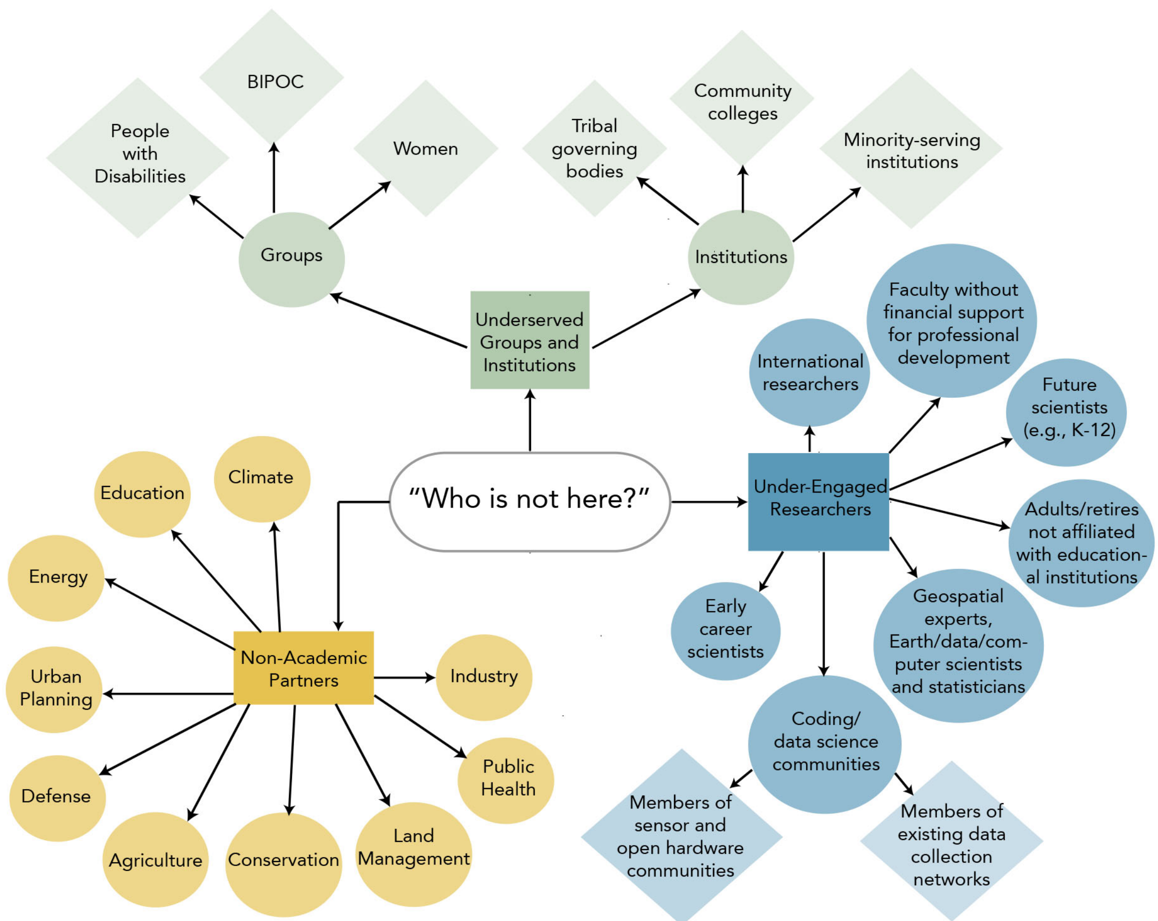


Fig. 3. Responses to the question asked of 2019 NEON Science Summit participants about who was not at the Summit that should have been (i.e., who or what groups were not well represented).

the types of research questions addressed using NEON open-source data and expand the visibility of the network to more potential users.

The knowledge generated from NEON products could be deepened through the co-production of science by members of the NEON community and stakeholders (i.e., leaders of underrepresented groups) to determine what types of ecological research questions are of interest to a range of communities. Building trust and meaningful relationships to jointly tackle research questions requires time and space to both listen and respond to the issues identified by all groups. Collaborations stemming from these efforts will help identify potential research sites for assignable assets that enable targeted groups to connect with NEON data.

However, it is not enough to identify and recruit members of underrepresented groups to join the NEON community; an inclusive environment encourages the development of essential skills and is open to new opportunities. Such an environment will recognize that the inclusion of diverse voices in science strengthens the creation of knowledge. Summit participants identified three main elements critical to these efforts: (1) education and training, (2) financial support, and (3) accessibility (Fig. 4).

*Accessibility in STEM.*—Continued funding for infrastructure, both physical and virtual, is essential for training ecologists and environmental scientists at all career stages, from undergraduate students to senior scientists, and from different professional settings, including

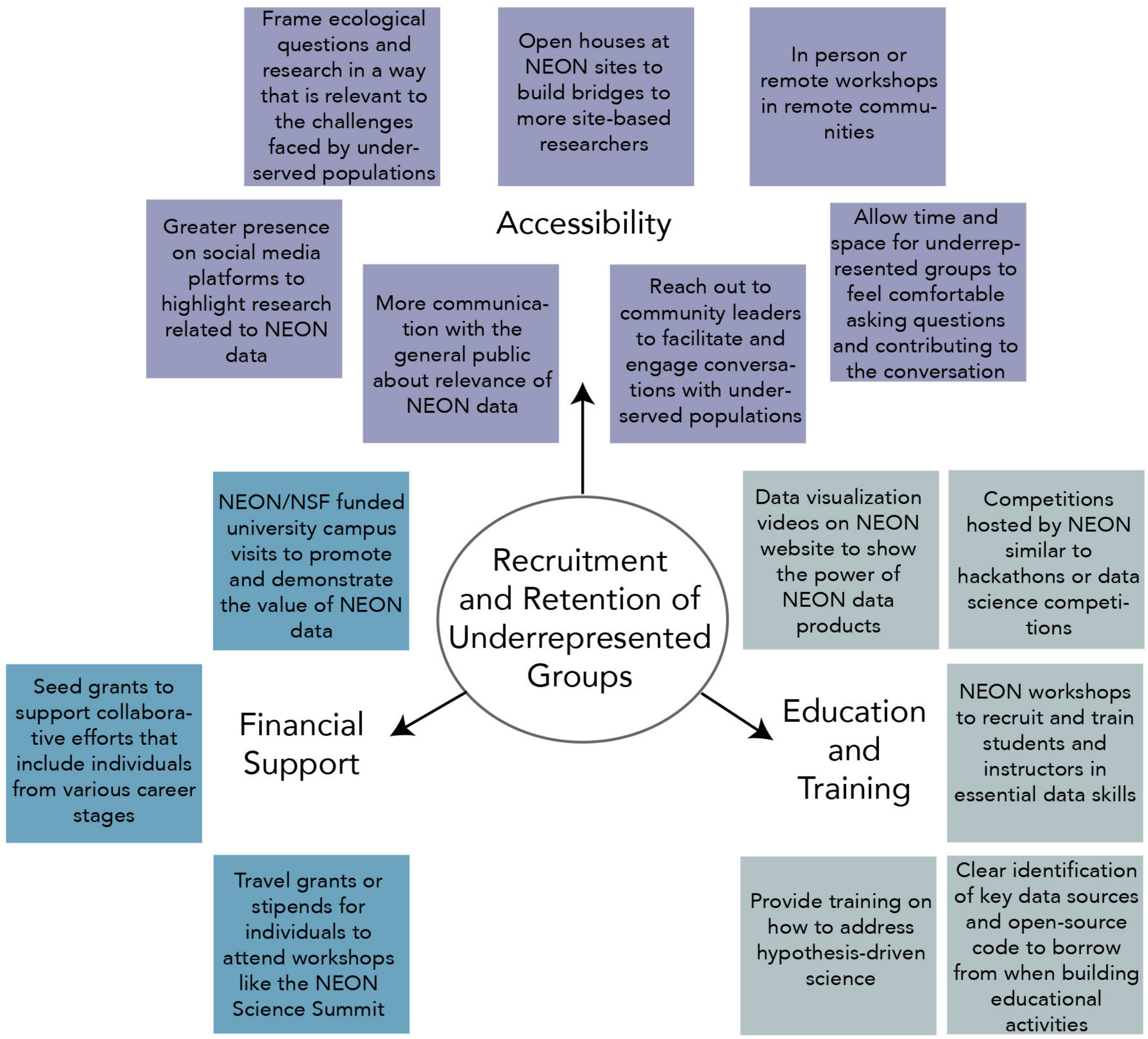


Fig. 4. Ideas from 2019 NEON Science Summit participants about how to recruit and retain underrepresented groups.

non-academic institutions. Creating a broad network of users will begin to address the challenge of accessibility to NEON data. The NSF-funded Quantitative Undergraduate Biology Education and Synthesis (QUBES) is a consortium of academic institutions, NSF projects, and professional societies that connects researchers across disciplines and supports efforts to improve quantitative literacy and data skills at the undergraduate level by offering training and course development opportunities to faculty (QUBES 2021*a*). Efforts like these can be expanded to meet the needs of diverse populations using NEON data. Other key infrastructural challenges include access to computers and reliable internet. Each institution may have unique barriers and the larger community needs to be mindful of the current states of physical infrastructure and levels of training to make NEON data accessible for all.

An online hub could help support individuals and institutions that serve underrepresented populations to encourage the use of NEON data and provide opportunities for learning how to use them. To kick things off, one opportunity is to host a series of short webinars to give college instructors and their students an introduction on how to access NEON data. These webinars could be followed by a virtual conference that includes lightning talks by members of target groups to present on their use of NEON data. Continued mentorship for faculty that serve underrepresented populations is key and will require time and commitment to the people at those schools/institutions. In the long term, for a sustainable diverse NEON community, members should explore the mandates and pursue the resources of NEON, NSF, and others to engage with these target groups (see *Community Engagement Initiatives*). Lastly, the community needs to spread the word to national, regional, and local agencies (e.g., Tribal offices, land management agencies, scientific societies, conservation organizations, and community groups) that opportunities exist for those from all professional and personal backgrounds. These outreach efforts can complement and build on the curriculum development and teaching resources described below.

### *NEON presents a need for new data-intensive curriculum in ecology*

NEON data offer an opportunity to teach scientific inquiry and ecological principles relevant to STEM careers but also transferable data science and critical thinking skills. A major challenge identified at the Summit is the limited number of people with skills to use NEON data among both the faculty and the students of the ecological community.

*A different model of teaching and doing ecological science.—*The availability of NEON data, while a tremendous opportunity for understanding environmental change, presents some unique challenges. The data are collected by an external entity rather than individual scientists and that collection is driven by NEON-focused mission science requirements. The scientist then develops questions that can be answered using NEON and other data. This model of science requirements driving data collection is not new: Remote sensing, astronomical telescope, particle collider, and other agency-driven data collection missions have implemented this approach to support the science community for decades (National Academies of Sciences, Engineering, and Medicine et al. 2018). However, data like those being collected by NEON at the U.S. continental scale, with standard measurements and more than 180 data products, are new to the natural resource sciences.

NEON thus presents a suite of new technical challenges associated with using data that are collected externally in formats and structures that support long-term and large data sets; their use will be new to many ecologists. These formats include HDF5 (the structure of the hyperspectral remote sensing data), and text files with relational database-like structures, to name just two. Additional challenges are associated with reading and interpreting metadata, understanding documentation (e.g., Algorithm Theoretical Basis Documents; ATBD), describing sampling designs, data collection methods, calibration procedures, nested structures in data storage (which vary by protocol), and uncertainty calculations. Further, many data sets require specific training to understand and evaluate quality (e.g., data from imaging spectroscopy, flux towers); it is not always appropriate to use products from the data portal without close consideration and filtering, which requires familiarity with domain literature and expertise. The cross-domain nature of the data collected by NEON (e.g., remote sensing, flux towers, organismal diversity) further encourages questions that are best addressed by large interdisciplinary teams, thus demanding a new suite of skills in the realm of communication and collaboration that also are not traditionally taught in natural resource science curricula.

*Addressing the community gap in data science skills.—*The skills needed to effectively work with NEON and other large data sets are not currently taught in most curricula, which struggle to keep pace with changes in technology and data processing. Education and training opportunities for students, educators, researchers, and community partners are key to building a vibrant and diverse community (Fig. 4). An example of a workshop that provides such opportunities is the "Critical Skills to Scale Up Ecology: An ESA SEEDS and NEON Workshop,"—an intensive week-long training designed to introduce ecological data skills to graduate students with underrepresented backgrounds (originally scheduled for June 2020 and postponed due to the COVID-19 pandemic). Another example is the Earth Lab Earth Data Science Corps (EDSC), funded by the NSF. The EDSC provides students at Tribal and other schools serving historically underrepresented groups in STEM with data skills training, mentorship, and a paid summer internship where they work on a real-world project: One Tribal student in the 2020 internship used NEON lidar data to evaluate forest structural diversity in the western United States. Improved access to these types of opportunities will also require financial support for those such as K-12 teachers, contingent university faculty, and non-academic professionals. Providing funding for early-career scientists and underrepresented populations can encourage participation at in-person training opportunities. Furthermore, the development of hybrid or remote conference and workshop opportunities make these events accessible to a broader population, especially for people who face obstacles to travel, such as families, resource constraints, and heavy teaching loads. Finally, the identification and development of tools that improve accessibility of NEON data, such as GUIs (graphical user interfaces), and IDEs (integrated development environments), and creation of teaching modules that help build the skills necessary to engage with NEON data can be promoted across the NEON user community.

*Instructor challenges for teaching with NEON data.*—A working group at the 2019 NEON Science Summit identified challenges facing instructors interested in using NEON data in undergraduate courses. While some guidance and educational modules and lessons do exist for instructors who wish to use NEON data, these are not housed in a centralized location or maintained by a single organization. Additionally, some instructors are lacking the data skills themselves to implement the lessons in the classroom or face other infrastructural and technological challenges. Some examples are given below of NEON Data Education faculty mentoring networks (see Naithani et al., *in this special issue*) and partnerships for instructors wishing to implement NEON data in the classroom.

Resources for instructors include teaching and learning modules, and full courses available on the NEON Learning Hub (NEON 2021*g*), the Earth Data Science learning portal (earthdatascience.org; Earth Lab 2021), the QUBES portal (QUBES 2021*b*), and the Environmental Data-Driven Inquiry and Exploration (EDDIE) website (EDDIE 2021). These modules can be added to existing courses or used for self-paced learning can further democratize access to these data skills. However, they lack a centralized platform and a clearly identified agency responsible for curating NEON instructional resources which creates barriers to their use in the classroom. The instructional level and types of data science skills required to complete the modules are not always easily discerned from the current descriptions. A framework is needed that would allow instructors to easily identify the intended audience (e.g., undergraduate course level), learning goals, and assessment tools included in these modules. This would also allow instructors interested in creating new modules to identify and fill gaps in the current resources. Sharing educational materials on a maintained and funded portal like QUBES or CyVerse would help with their adoption. Additionally, adding tags and providing links to these resources on NEON's website would further facilitate discoverability.

Instructors may face many barriers to implementing NEON data in their classrooms including the instructor's own lack of data science skills and familiarity with the data products among others (e.g., lack of funding for new curriculum, technology, and time). Sustained and improved outreach and instructor training is necessary to overcome this first barrier. The NEON community should identify and target populations and institutions that are not currently engaged with or utilizing NEON data, such as primarily undergraduate institutions (PUIs), minority-serving institutions, and Tribal colleges. One such example is developing a curriculum with PUI ecology and GIS faculty that is centered around testing the use of NEON terrestrial and airborne collections to introduce spatial ecology and macrosystems biology concepts in undergraduate courses (Styers et al. *unpublished manuscript*). A lesson learned from these activities is that it is helpful for NEON staff to participate and introduce some of the NEON sampling design and data portal, and for researchers familiar with working with NEON data to help PUI faculty quickly get up-to-speed on how to work with the data. It is also important to consider the resources available at these institutions and address technological challenges related to hardware and Internet connectivity that may exist.

Several faculty mentoring networks and partnerships exist including collaborations with NEON, Ecological Research Education Network (EREN), and the Biological and Environmental Data Education (BEDE) Network and are hosted by QUBES, which support faculty in both using and developing new data-intensive curricula for their courses (BEDE 2021, EREN 2021, NEON 2021*h*). These faculty mentoring networks work with faculty members from undergraduate teaching institutions to develop, implement, and publish data-driven teaching modules to empower undergraduate instructors to incorporate data science skills in their courses. Such partnerships could provide a fruitful avenue for incorporating real-world skills and experiences using NEON data in undergraduate curricula.

### *Community engagement initiatives*

In order to maximize the benefit of NEON, the community needs activities that support researchers within, and attract researchers from beyond, current NEON users. Achievement of this robust community will require enhanced funding to support broad community engagement initiatives, and a science-forward centralized mechanism to share and advertise those opportunities.

This support should take on at least three synergistic forms: (1) engage and train early-career researchers, (2) build an inclusive community of scientists, and (3) prioritize community engagement activities. All three support mechanisms require nuanced approaches to building trust in and transparency around data and shared goals that enable the transition from place-based science to research that spans across and links ecological systems. Building an inclusive and expansive research community involves dialogue with multiple interests and recognition of the emergence of a new culture of ecology based on team science and data sharing. Workshops, lecture series, and training are some activities that require funding. In particular, outreach and funding are needed to attract and support underrepresented groups and those who are not current NEON users. Additionally, outreach to other disciplines will help increase awareness of how NEON data can contribute to the conversation on public policy and applications to address emerging societal and ecological challenges. Such an effort can enhance the value of this revolutionary ecological infrastructure.

## Why NEON Matters

It is important to remember how NEON is similar to and different from other efforts that came before it. In particular, many of the questions remain the same; abiding environmental questions and problems still drive individual PIs as well as networks such as NEON, LTER, CZO, AmeriFlux, etc. However, NEON was established in the era of open science that aims to democratize the access to knowledge with freely available data to all students and in all classrooms.

The success of NEON matters both to the ecological community and society as a whole. NEON represents the largest single U.S. investment in ecology to date, and it will inform future scientific missions by other agencies and data networks. The observatory will provide critical science for key decisions in the management of ecosystems and habitats. It is up to the community of users to increase awareness of how NEON data can contribute to informed public policy and socially relevant applications. It offers new opportunities for tackling big challenges such as climate change, land-use change, and ecological transformations that affect all life on Earth. NEON will not stand alone as a solution to these complex challenges (see SanClements et al., *in this special issue*), but offers opportunities to build on other approaches, data, and knowledge to support individual scientists and the ecological community to address them.

## Conclusions

NEON is intended to help us observe, understand, and interpret the response of species, communities, and ecosystems to our changing environment. This revolutionary observatory network for ecology will enable us to ask and answer local- to continental-scale questions with a design that includes sites located across ecoregions. NEON aids understanding of ecological change across space and time and the forecasting of future conditions with drivers like land-use change and climate change. The standardization of collection methods across sites, and open access to the data can foster a new, open ecology for the next generation of scientists, including microbial ecologists, biogeochemists, community ecologists, atmospheric scientists, and more. NEON became fully operational during the data and analytics revolutions and now is poised to contribute to advancing the discipline in conjunction with other long-term observatories and data sets. As data science and analytics have become critical skills for today's scientists, training on how to access and analyze NEON data is also critical. There is a need to develop a curriculum to teach these skills to a diverse, inclusive NEON user community. Collaborative, hands-on workshops like the 2019 NEON Science Summit will build a broad community, and investing in other inclusive mechanisms will sustain this engaged community.

## Acknowledgments

R. C. Nagy and J. K. Balch contributed equally to the work reported here and were co-lead authors. Funding for the 2019 NEON Science Summit was provided by NSF Award #DBI 1906144. Additional funding was provided by Earth Lab through the University of Colorado, Boulder's Grand Challenge Initiative, the Cooperative Institute for Research in Environmental Sciences (CIRES), and the North Central Climate Adaptation Science Center (NC CASC). AKS was supported by the University of Zurich's University Research Priority Programme on Global Change and Biodiversity. We are grateful to all of those who helped make the 2019 NEON Science Summit a success: Dawn Umpleby, Linda Pendergrass, David Zakavec, Rebecca Stossmeister, Alycia Crall, Kate Thibault, Claire Lunch, Megan Jones, Katie Weeman, Katy Human, Ally Faller, Lauren Herwehe, Jenny Palomino, Tim Dunn, Nathan Campbell, Nan Regnier, Elizabeth Woolner, Katie Lamb, Troy Burke, Jim Lee, Jay Burghardt, and Michelle Meighan. Several co-authors on this paper including Christopher R. Florian, Robert T. Hensley, Katherine D. Jones, Courtney L. Meier, John Musinsky, Stephanie M. Parker, Michael SanClements, and Eric R. Sokol are currently employed by Battelle and the National Ecological Observatory Network. The findings and conclusions in this manuscript are those of the authors and should not be construed to represent any official U.S. Government determination or policy.

## Literature Cited

Albrecht, M., P. Donnelly, P. Bui, and D. Thain. 2012. Makeflow: a portable abstraction for data intensive computing on clusters, clouds, and grids. Pages 1–13 *in* Proceedings of the 1st ACM SIGMOD Workshop on Scalable Workflow Execution Engines and Technologies - SWEET '12. ACM Press, Scottsdale, Arizona, USA.

Anderegg, W. R., and N. S. Diffenbaugh. 2015. Observed and projected climate trends and hotspots across the National Ecological Observatory Network regions. Frontiers in Ecology and the Environment 13:547–552.

Atkins, J. W., R. T. Fahey, B. S. Hardiman, and C. M. Gough. 2018. Forest canopy structural complexity and light absorption relationships at the subcontinental scale. Journal of Geophysical Research: Biogeosciences 123:1387–1405.

Balch, J. K., et al. 2020*a*. Social-environmental extremes: rethinking extraordinary events as outcomes of interacting biophysical and social systems. Earth's Future 8:e2019EF001319.

Balch, J. K., R. C. Nagy, and B. S. Halpern. 2020*b*. NEON is seeding the next revolution in ecology. Frontiers in Ecology and the Environment 18:3.

Bartowitz, K. J., P. E. Higuera, B. N. Shuman, K. K. McLauchlan, and T. W. Hudiburg. 2019. Post-fire carbon dynamics in subalpine forests of the Rocky Mountains. Fire 2:58.

Bechhofer, S., et al. 2013. Why linked data is not enough for scientists. Future Generation Computer Systems 29:599–611.

BEDE. 2021. Biological and Environmental Data Education (BEDE) Network. https://qubeshub.org/community/groups/bede

Bucksch, A., A. Das, H. Schneider, N. Merchant, and J. S. Weitz. 2017. Overcoming the Law of the Hidden in Cyberinfrastructures. Trends in Plant Science 22:117–123.

Cawse-Nicholson, K., et al. 2021. NASA's surface biology and geology designated observable: a perspective on surface imaging algorithms. Remote Sensing of Environment 257:112349.

Cesare, C. 2016. January 19. Ecology's $434,000,000 test. Nature News.

Chadwick, K. D., et al. 2020. Integrating airborne remote sensing and field campaigns for ecology and Earth system science. Methods in Ecology and Evolution 11:1492–1508.

Collins, S. L., and A. K. Knapp. 2019. NEON should be run by ecologists for ecologists. BioScience 69:319.

CyVerse. 2021. Data Commons. https://datacommons.cyverse.org/

Daam, M. A., H. Teixeira, A. I. Lillebø, and A. J. A. Nogueira. 2019. Establishing causal links between aquatic biodiversity and ecosystem functioning: status and research needs. Science of the Total Environment 656:1145–1156.

Dabbish, L., C. Stuart, J. Tsay, and J. Herbsleb 2012. Social coding in GitHub: transparency and collaboration in an open software repository. Page 1277 *in* Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12. ACM Press, Seattle, Washington, USA.

Dalponte, M., L. Frizzera, and D. Gianelle. 2019. Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data. PeerJ 6:e6227.

Delsol, R., M. Loreau, and B. Haegeman. 2018. The relationship between the spatial scaling of biodiversity and ecosystem stability. Global Ecology and Biogeography 27:439–449.

Dietze, M., and H. Lynch. 2019. Forecasting a bright future for ecology. Frontiers in Ecology and the Environment 17:3.

Dorji, T., K. A. Hopping, F. Meng, S. Wang, L. Jiang, and J. A. Klein. 2020. Impacts of climate change on flowering phenology and production in alpine plants: the importance of end of flowering. Agriculture, Ecosystems & Environment 291:106795.

Earth Lab. 2021. Earth Data Science learning portal. https://www.earthdatascience.org/

EDI. 2021. Data Portal Home. https://portal.edirepository.org/nis/home.jsp

EDDIE. 2021. Environmental data-driven inquiry & exploration. https://serc.carleton.edu/eddie/enviro_data/index.html

EFI. 2021. EFI RCN NEON Ecological Forecast Challenge. Ecological Forecasting Initiative.

EREN. 2021. EREN-NEON Flexible Learning Projects FMN. https://qubeshub.org/community/groups/erenneonfmn

Eynard-Bontemps, G., R. Abernathey, J. Hamman, A. Ponte, and W. Rath. 2019. The PANGEO Big Data Ecosystem and its use at CNES. Pages 49–52 *in* Proc. of the 2019 conference on Big Data from Space (BiDS'2019), EUR 29660 EN. Publications Office of the European Union, Luxembourg.

Fedele, G., B. Locatelli, and H. Djoudi. 2017. Mechanisms mediating the contribution of ecosystem services to human well-being and resilience. Ecosystem Services 28:43–54.

Fegraus, E. H., S. Andelman, M. B. Jones, and M. Schildhauer. 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. Bulletin of the Ecological Society of America 86:158–168.

Field, C., R. DeFries, D. Foster, M. Grove, R. Jackson, B. Law, D. Lodge, D. Peters, and D. Schimel. 2006. Integrated Science and Education Plan for the National Ecological Observatory Network. NSF NEON.

Fricker, G. A., N. W. Synes, J. M. Serra-Diaz, M. P. North, F. W. Davis, and J. Franklin. 2019*a*. More than climate? Predictors of tree canopy height vary with scale in complex terrain, Sierra Nevada, CA (USA). Forest Ecology and Management 434:142–153.

Fricker, G. A., J. D. Ventura, J. A. Wolf, M. P. North, F. W. Davis, and J. Franklin. 2019*b*. A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. Remote Sensing 11:2326.

Ghabbour, E. A., G. Davies, A. A. Sayeed, M. T. Croman, B. A. Hoehing, and E. Ayres. 2015. Measuring the total and sequestered organic matter contents of grassland and forest soil profiles in the national ecological observatory network initiative. Soil Horizons 56:1–11.

Gonzalez, A., et al. 2020. Scaling-up biodiversity-ecosystem functioning research. Ecology Letters 23:757–776.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202:18–27.

Grimm, N. B., S. T. A. Pickett, R. L. Hale, and M. L. Cadenasso. 2017. Does the ecological concept of disturbance have utility in urban social–ecological–technological systems? Ecosystem Health and Sustainability 3:e01255.

Hall, S. J., C. Ye, S. R. Weintraub, and W. C. Hockaday. 2020. Molecular trade-offs in soil organic carbon composition at continental scale. Nature Geoscience 13:687–692.

Hampton, S. E., et al. 2015. The Tao of open science for ecology. Ecosphere 6:art120.

Hey, A. J. G. editor. 2009. The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond, Washington, USA.

Hinckley, E. S., et al. 2016*a*. The soil and plant biogeochemistry sampling design for The National Ecological Observatory Network. Ecosphere 7. https://doi.org/10.1002/ecs2.1234

Hinckley, E.-L.-S., et al. 2016*b*. Optimizing available network resources to address questions in environmental biogeochemistry. BioScience 66:317–326.

Isard, M., M. Budiu, Y. Yu, A. Birrell, and D. Fetterly. 2007. Dryad: distributed data-parallel programs from sequential building blocks. ACM SIGOPS Operating Systems Review 41:59–72.

Jones, J. A., et al. 2021. Synergies among environmental science research and monitoring networks: a research agenda. Earth Futures. 9:e2020EF001631.

Kampe, T. U., B. Johnson, M. Kuester, and M. Keller. 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. Journal of Applied Remote Sensing 4:043510.

Kao, R. H., et al. 2012. NEON terrestrial field observations: designing continental-scale, standardized sampling. Ecosphere 3:art115.

Keller, M., D. S. Schimel, W. W. Hargrove, and F. M. Hoffman. 2008. A continental strategy for the National Ecological Observatory Network. Frontiers in Ecology and the Environment 6:282–284.

Köster, J., and S. Rahmann. 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28:2520–2522.

Kramer, M. G., and O. A. Chadwick. 2018. Climate-driven thresholds in reactive mineral retention of soil carbon at the global scale. Nature Climate Change 8:1104–1108.

LaRue, E. A., B. S. Hardiman, J. M. Elliott, and S. Fei. 2019. Structural diversity as a predictor of ecosystem function. Environmental Research Letters 14:114011.

Liu, J., et al. 2007. Complexity of Coupled Human and Natural Systems. Science 317:1513–1516.

Loeliger, J., and M. McCullough. 2012. Version control with Git. Second edition. O'Reilly, Beijing, China.

Lunch, C., C. Laney, N. Mietkiewicz, E. Sokol, K. Cawley, and NEON (National Ecological Observatory Network). 2021. neonUtilities: Utilities for Working with NEON Data. R package version 2.1.0. https://CRAN.R-project.org/package=neonUtilities

Mahood, A. L., and J. K. Balch. 2019. Repeated fires reduce plant diversity in low-elevation Wyoming big sagebrush ecosystems (1984–2014). Ecosphere 10:e02591.

Marconi, S., S. J. Graves, B. G. Weinstein, S. Bohlman, and E. P. White. 2021. Estimating individual-level plant traits at scale. Ecological Applications 31: e02300.

Merchant, N., E. Lyons, S. Goff, M. Vaughn, D. Ware, D. Micklos, and P. Antin. 2016. The iPlant Collaborative: cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. PLOS Biology 14: e1002342.

Mervis, J. 2015, September 24. Ecology's megaproblem, Science News.

National Academies of Sciences, Engineering, and Medicine, Space Studies Board, Division on Engineering and Physical Sciences, and National Academies of Sciences, Engineering, and Medicine. 2018. Thriving on our changing planet: a decadal strategy for Earth observation from space. Page 24938. National Academies Press, Washington, D.C., USA.

National Ecological Observatory Network (NEON). 2020. geoNEON: geolocation data access for NEON data. R package version 1.0.0.9100.

NEON. 2021*a*. About Field Sites and Domains. https://www.neonscience.org/field-sites/about-field-sites

NEON. 2021*b*. History. https://www.neonscience.org/about/about/history

NEON. 2021*c*. Explore Data Products. https://data.neonscience.org/data-products/explore

NEON. 2021*d*. Data Tutorials. https://www.neonscience.org/resources/data-tutorials

NEON. 2021*e*. Research Support and Assignable Assets. https://www.neonscience.org/resources/research-support

NEON. 2021*f*. Code Resources. https://www.neonscience.org/resources/code-resources

NEON. 2021*g*. Learning Hub. https://www.neonscience.org/resources/learning-hub

NEON. 2021*h*. NEON Community Group. https://qubeshub.org/community/groups/neon/educational_resources

O'Brien, M., C. A. Smith, E. R. Sokol, C. Gries, N. Lany, S. Record, and M. C. N. Castorani. 2021. ecocomDP: a flexible data design pattern for ecological community survey data. Ecological Informatics 64:101374.

Pangea. 2021. Home. https://www.pangaea.de/

Pricope, N., K. Mapes, and K. Woodward. 2019. Remote sensing of human-environment interactions in global change research: a review of advances. Challenges and Future Directions. Remote Sensing 11:2783.

pyOpenSci. 2021. Welcome to pyOpenSci. https://www.pyopensci.org//

QUBES. 2021*a*. Home. https://qubeshub.org/

QUBES. 2021*b*. Resources. https://qubeshub.org/publications/browse

Read, Q. D., J. M. Grady, P. L. Zarnetske, S. Record, B. Baiser, J. Belmaker, M.-N. Tuanmu, A. Strecker, L. Beaudrot, and K. M. Thibault. 2018. Among-species overlap in rodent body size distributions predicts species richness along a temperature gradient. Ecography 41:1718–1727.

Record, S., et al. 2021. Novel insights to be gained from applying metacommunity theory to long-term, spatially replicated biodiversity data. Frontiers in Ecology and Evolution 8:479.

Rogers, N. 2019. The turbulent launch of NEON, ecology's megaproject. Inside Science.

rOpenSci, 2021. rOpenSci. https://ropensci.org/

Sagoff, M. 2019. Will NEON kill ecology? Issues in Science and Technology 35:54–62.

Schimel, D. 2011. The era of continental-scale ecology. Frontiers in Ecology and the Environment 9:311.

Schimel, D., et al. 2011. 2011 Science Strategy: enabling Continental-Scale Ecological Forecasting. Page 56. NSF NEON.

Schimel, D., W. Hargrove, F. Hoffman, and J. MacMahon. 2007. NEON: a hierarchically designed national ecological network. Frontiers in Ecology and the Environment 5:59.

Schimel, D., and M. Keller. 2015. Big questions, big science: meeting the challenges of global ecology. Oecologia 177:925–934.

Schimel, D., F. D. Schneider, J. P. L. Carbon, and E. Participants. 2019. Flux towers in the sky: global ecology from space. New Phytologist 224:570–584.

Scholl, V. M., M. Cattau, M. B. Joseph, and J. K. Balch. 2020. Integrating National Ecological Observatory Network (NEON) airborne remote sensing and in-situ data for optimal tree species classification. Remote Sensing 12:1414.

Seyednasrollah, B., A. M. Young, X. Li, T. Milliman, T. Ault, S. Frolking, M. Friedl, and A. D. Richardson. 2020. Sensitivity of deciduous forest phenology to environmental drivers: implications for climate change impacts across North America. Geophysical Research Letters 47:e2019GL086788.

Sutherland, W. J., et al. 2013. Identification of 100 fundamental ecological questions. Journal of Ecology 101:58–67.

Swetnam, T. L., C. D. O'Connor, and A. M. Lynch. 2016. Tree morphologic plasticity explains deviation from metabolic scaling theory in semi-arid conifer forests, Southwestern USA. PLoS One 11: e0157582.

Taylor, J. R., E. Ayres, H. Luo, and H. W. Loescher 2011. The National Ecological Observatory Network's Fundamental Instrument Unit: challenges to Managing Thousands of Environmental Sensors. Page JWE16. Renewable Energy and the Environment, OSA, Austin, Texas.

Thessen, A., et al. 2020. Predicting Phenotype from Multi-Scale Genomic and Environment Data using Neural Networks and Knowledge Graphs. Zenodo.

Thom, D., and R. Seidl. 2016. Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests: Disturbance impacts on biodiversity and services. Biological Reviews 91:760–781.

Towns, J., et al. 2014. XSEDE: accelerating Scientific Discovery. Computing in Science & Engineering 16:62–74.

Varadharajan, C., S. Cholia, C. Snavely, V. Hendrix, C. Procopiou, D. Swantek, W. J. Riley, and D. A. Agarwal. 2019. Launching an accessible archive of environmental data. Eos 100. http://dx.doi.org/10.1029/2019EO111263

Weibel, S., J. Kunze, C. Lagoze, and M. Wolf. 1998. Dublin Core Metadata for Resource Discovery. Page RFC2413. RFC Editor.

Weinstein, B. G., S. Marconi, S. Bohlman, A. Zare, and E. White. 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. Remote Sensing 11:1309.

Weintraub, S. R., L. Stanish, and E. Ayers. 2017. Using NEON Data to Test and Refine Conceptual and Numerical Models of Soil Biogeochemical and Microbial Dynamics. Pages B44B-04.

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: an evolving community-developed biodiversity data standard. PLoS One 7: e29715.

Wieder, W. R., et al. 2020. SoDaH: the SOils DAta Harmonization database, an open-source synthesis of soil data from research networks, version 1.0. Earth System Science Data Discussions 1–19.

Wilkinson, M. D., et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3:160018.

Zelnik, Y. R., J.-F. Arnoldi, and M. Loreau. 2018. The impact of spatial and temporal dimensions of disturbances on ecosystem stability. Frontiers in Ecology and Evolution 6:224.

Zheng, C., and D. Thain. 2015. Integrating Containers into Workflows: a Case Study Using Makeflow, Work Queue, and Docker. Pages 31–38 *in* Proceedings of the 8th International Workshop on Virtualization Technologies in Distributed Computing - VTDC '15. ACM Press, Portland, Oregon, USA.

## Data Availability Statement

Data are available from Dryad: https://doi.org/10.5061/dryad.866t1g1rj