Optimizing Available Network Resources to Address Questions in Environmental Biogeochemistry

EVE-LYN S. HINCKLEY, SUZANNE P. ANDERSON, JILL S. BARON, PETER D. BLANKEN, GORDON B. BONAN, WILLIAM D. BOWMAN, SARAH C. ELMENDORF, NOAH FIERER, ANDREW M. FOX, KELI J. GOODMAN, KATHERINE D. JONES, DANICA L. LOMBARDOZZI, CLAIRE K. LUNCH, JASON C. NEFF, MICHAEL D. SANCLEMENTS, KATHARINE N. SUDING, AND WILLIAM R. WIEDER

An increasing number of network observatories have been established globally to collect long-term biogeochemical data at multiple spatial and temporal scales. Although many outstanding questions in biogeochemistry would benefit from network science, the ability of the earth- and environmental-sciences community to conduct synthesis studies within and across networks is limited and seldom done satisfactorily. We identify the ideal characteristics of networks, common problems with using data, and key improvements to strengthen intra- and internetwork compatibility. We suggest that targeted improvements to existing networks should include promoting standardization in data collection, developing incentives to promote rapid data release to the public, and increasing the ability of investigators to conduct their own studies across sites. Internetwork efforts should include identifying a standard measurement suite—we propose profiles of plant canopy and soil properties—and an online, searchable data portal that connects network, investigator-led, and citizen-science projects.

Keywords: ecology, environmental science, modeling, monitoring/mapping, nutrient cycling

Networks differ in their origins, management structures, and science themes, but they are unified by their efforts to collect integrated data sets that span extensive spatial and/or temporal scales. Sometimes this promise is realized; sometimes it is not. Regardless, there is enormous potential in using network science to address some of the outstanding questions in biogeochemistry. Many questions in this field could benefit from data that extend beyond traditional single-investigator projects and instead come from networks that produce standardized, long-term, and spatially extensive data streams (Peters et al. 2014). The potential applications are diverse. For example, investigators have conducted syntheses of aboveground net primary production (ANPP) across a gradient of biomes represented by sites within the Long-Term Ecological Research (LTER) network (Knapp and Smith 2001). In other cases, modelers have sought to integrate physical and biological data from multiple sources, including networks. Using an Earth system model (ESM) framework to investigate the influence of climate on carbon (C) cycle feedbacks (Bonan et al. 2012) relies not only on single data streams that are standardized within particular networks (e.g., carbon dioxide, CO_2, fluxes from across AmeriFlux sites) but also integrates compatible data from other sources, such as data repositories (e.g., the TRY leaf database; see Kattge et al. 2011) and neighboring network sites (such as foliar nitrogen, N, collected at LTER sites).

Although there are examples of major networks in the United States and beyond that are generating data streams that could help address these and other complex research topics in biogeochemistry, and there have been pleas over the years to align existing and new networks (Bricker and Ruggiero 1998, Murdoch et al. 2014, O'Neel et al. 2015), there remain many barriers to integrating data streams across (or even within) networks. This is a crucial time for funding agencies, organizations that operate networks, and individual researchers to evaluate their investments in this diverse landscape of network science data and infrastructure. Networks that collect biogeochemical data as part of their science focus are at various stages of maturity: The oldest LTER sites have now been in operation for 35 years; the Critical Zone Observatory (CZO) network has expanded since its 2007 inception in the United States (Anderson et al. 2008), and the idea is spreading internationally (Banwart et al. 2013); the newer National Ecological Observatory Network (NEON) is beginning to generate data despite a contraction in scope (Mervis 2015a, 2015b, 2015c); and there are many grassroots, collaborative efforts by investigators,

BioScience 66: 317–326.
doi:10.1093/biosci/biw005 Advance Access publication 17 February 2016

such as DroughtNet (*http://wp.natsci.colostate.edu/droughtnet*). As network science moves forward, we must consider how a few targeted changes could help to maximize the different resources that networks bring, explore ways that their accumulated data can be integrated with other types of studies, and ensure that they are aligned with the scientific priorities of the broader community of researchers, educators, and policymakers.

In this article, we address the challenges associated with accessing and using network data resources for biogeochemical studies. We examine these challenges through examples of ecological synthesis projects integrating data from within a particular network, as well as recent efforts to use suites of observations from a variety of sources—including networks—to understand how climate affects above- and belowground C storage using ESMs (see figure 1). We provide suggestions for improving intra- and internetwork coordination that would maximize the investment in pre-existing networks by US funding agencies and improve the use of network resources by modelers and observational scientists alike.

The diverse landscape of current network resources

Networks vary in terms of management and funding structures, degree of standardization, and scalability of measurements. These characteristics influence the ability of data users to quantify ecological trends and answer questions at large spatial and long temporal scales. Several recent papers, such as Peters and colleagues (2014) and Collins and Childers (2014), explored these characteristics, which we highlight briefly here. For example, long-term research networks—including the LTER sites, the Long-Term Agricultural Research (LTAR) sites, and the CZO network—are internally coordinated and have individual investigator-driven management structures, as well as site-based instrumentation, methodologies, and measurement suites. This “bottom-up” network structure invites investigators to ask targeted, mechanistic, and site-based questions. However, the ability to synthesize data across space and time, although encouraged, is limited because of the lack of funds for coordinated cross-site research efforts, methodological standardization, and, in some cases, spatial and temporal coverage. In contrast, ecological observatory networks, such as NEON, provide a centralized, top-down management structure that facilitates standardized instrumentation, field and laboratory methodologies, and data curation but do not offer the same opportunities for site-specific exploration or experimentation.

A third model falls between these two: Coordinated distributed observations and experiments through networks, such as AmeriFlux and the National Atmospheric Deposition Program (NADP), provide standardized, spatially distributed data via a centralized repository. These entities rely on investments by interested investigators sometimes catalyzed by specific funding mechanisms. They have standardized methodologies and data processing across sites and typically ask a very specific set of research questions. For example, AmeriFlux focuses on understanding terrestrial CO_2 fluxes, whereas NADP provides data on atmospheric deposition. Because these networks rely on commitment from individual investigators, sites may be discontinued if the lead investigator’s funding ends or research focus shifts. The Department of Energy has tried to minimize this outcome by funding “core sites” within the AmeriFlux network.

There is increasing investment in other national and international networks outside of the United States. Some examples that are relevant to biogeochemistry include the Global Lake Ecological Observatory Network (GLEON), the International Soil Carbon Network (ISCN), the International Long-Term Ecological Research (ILTER) Network, and FLUXNET. These networks have different challenges associated with coordinating across political boundaries and funding sources, but there are many similarities with the types of US networks outlined above—and many of the same challenges in accessing their data resources. Although we focus here on a subset of US networks, we acknowledge that global-scale questions in biogeochemistry require global-scale participation and investment (Banwart et al. 2013); therefore, we make many of our points with attention to this broader scope.

Characteristics of the ideal network

We identify some of the primary characteristics of network science that make it ideally suited for use by the community (figure 2). These characteristics really call for a more specific definition or set of requirements for a collection of sites that is identified as a network. Specifically, we believe that a network is most successful when the majority of its data collection is standardized in its methodologies with repeated measurements taken across space and time. For existing networks, this would primarily be a revision for “bottom-up” structures, such as LTER and CZO. The data sets that fall into this category depend on the network’s overarching science theme; the measurements chosen for standardization across AmeriFlux’s network of eddy covariance towers differ from those that might be chosen for collection within observational plots and stream gaging stations at LTER or CZO sites. For investigators actively working at particular sites within a network, a common standardized set of measurements would be useful to aid studies that extend beyond the scope of a particular investigator’s research. Data users conducting cross-site syntheses desire similar standardized measurements. The problems with not having standardized measurements are described in the next section.

We believe that rapid reporting to make results publicly available is increasingly important in network science, not only for the broader community of researchers but also for justifying continued federal support for the network. Reporting of quality controlled data should be complete with associated metadata (location and time of collection, methodologies used) that are necessary for interpreting the results, as well as connecting them with other data streams from within or outside of a particular network. The need to incentivize rapid reporting of results from existing networks goes across the three types—bottom up, top down,

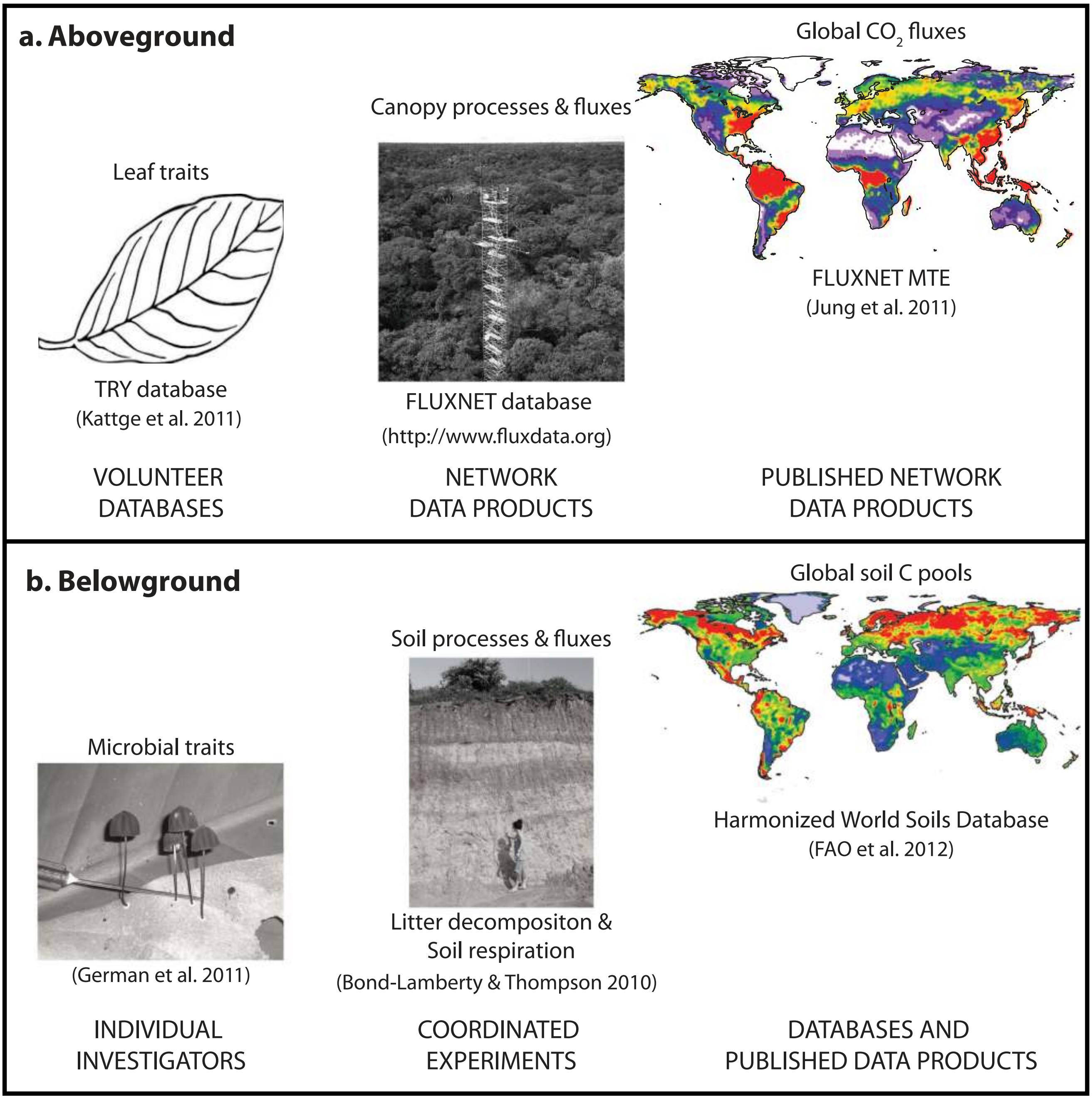

Figure 1. Representative data sources that are or could be used to evaluate biogeochemical processes in Earth system models a. aboveground and b. belowground. At increasing scales of interest, data may come from trait databases (leaf and microbial), cross-site measurements of fluxes and biogeochemical processes (canopy and soil pedons), and observationally derived estimates of global carbon (C) fluxes and pools. Examples of data sources are cited below each image. The flux tower image is from **www.bgc-jena.mpg.de/public/carboeur/archive/foto.html**. *The other images are from William R. Wieder.*

and hybrid—introduced above; largely, it calls for a cultural shift from individuals treating data as proprietary to willingly releasing them into the public domain and coordinated efforts to create data reporting systems that ease the process of publishing data online.

Networks should have the flexibility to incorporate new measurements and investigator-led research. Although this capability is more common in many of the bottom-up networks, it can be challenging in the top-down models. The Nutrient Network's (NutNet) protocol calls for leaving a subset of replicates available for site-specific studies, or additional experiments, whereas an investigator interested in working at a NEON site must work out logistics, permitting, and other arrangements with each of 47 site hosts.

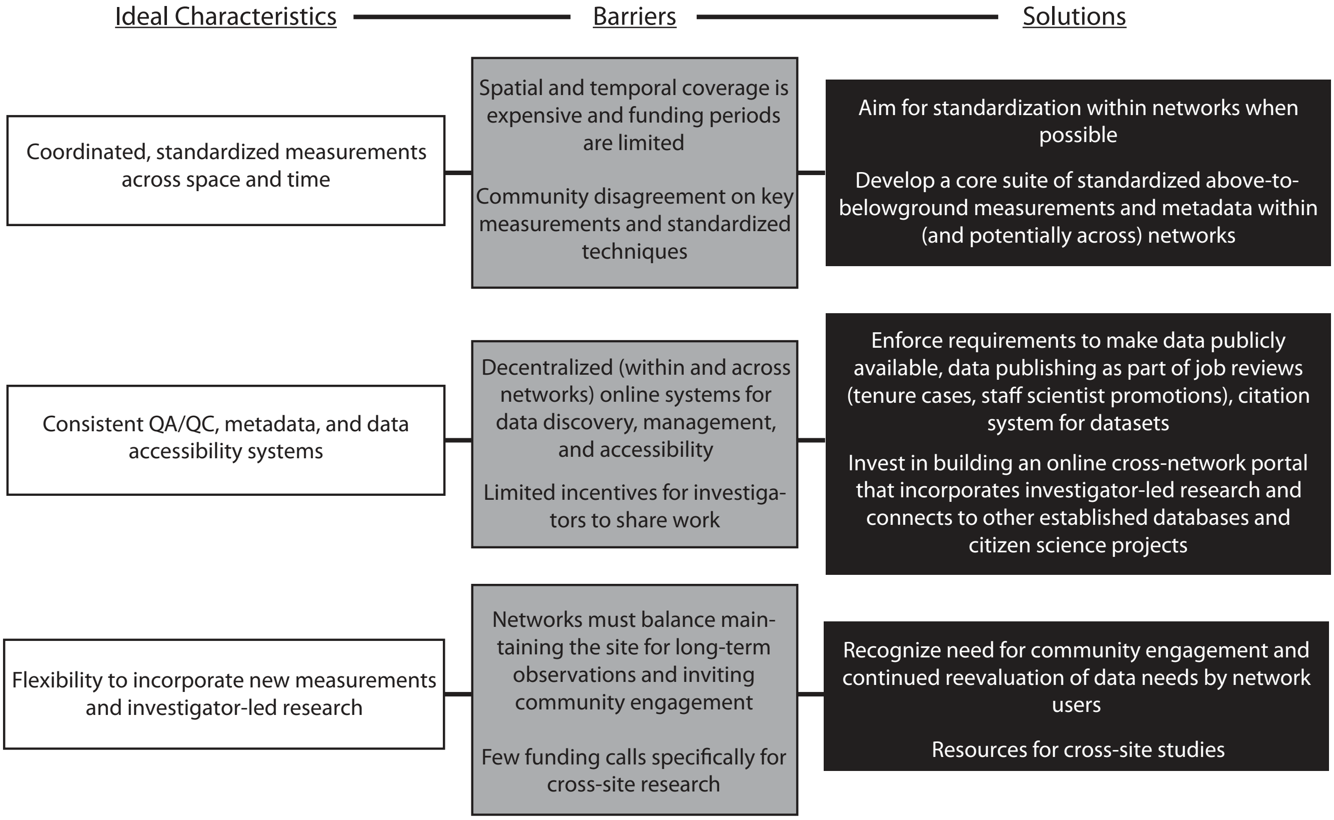

Figure 2. A summary of ideal characteristics of networks, current barriers to optimizing data use, and potential solutions to move networks toward greater usability.

However, information gleaned from data on long-term trends—often accomplished through monitoring—is greatest when it directly stimulates further investigations and informs process-based studies. For example, many monitoring efforts have detected increasing export of dissolved organic carbon (DOC) in streams across the Northern Hemisphere over the last two decades (see Driscoll et al. 2003, Evans et al. 2005). However, these observations alone were insufficient to pinpoint the factors responsible for this trend, which could include rising atmospheric CO_2 concentrations, climate warming, decreased sulfate deposition, and changing hydrology (Evans et al. 2005, Porcal et al. 2009). Determining the primary large-scale drivers and local mechanisms leading to increasing stream DOC export is an active area of research. Inspiring further targeted studies across gradients in climate or ecosystem types, for example, should be a key value in collecting long-term data at network sites.

Why do network resources fall short?

There are many examples from observational syntheses and modeling studies in biogeochemistry that demonstrate the current problems with using network data and other resources, summarized in figure 2. Synthesis studies that seek to draw on observations of ecosystem parameters or processes across sites often rely on meta-analysis techniques to deal with incomplete or nonstandardized data sets. Meta-analysis techniques are commonly employed in such instances because the statistical analyses can be tailored to handle the discrepancies. The meta-analysis approach typically involves combining the results of individual studies by converting raw measurements at each site into an effect size, which accounts for differences in variance and/or scale among sites. One disadvantage of meta-analyses is that large effect sizes can result from either large change in mean response or small variances (Rosenberg et al. 2013). As a result, meta-analyses that employ common effect size metrics, such as the standardized mean difference, allow inference about the direction of a response (e.g., an increase in C fluxes across a gradient of sites) and statistical significance but not the magnitude of change in biologically meaningful units (e.g., change in kilograms of C over time or space).

The synthesis study of Clark and colleagues (2007) illustrated many of the problems associated with integrating data sets from across network sites using meta-analysis. They sought to determine ecosystem predictors of plant community production and species richness in response to N additions. The study synthesized data from across 34 experiments conducted in nine herbaceous ecosystems in North America. The investigators reported problems with missing (i.e., not collected) and incomparable (i.e., inconsistent collection and analysis methodologies across sites) data, primarily from

LTER sites. In particular, they reported these issues with respect to the soil and biogeochemical parameters. Although they were able to include soil pH in their analysis, it was not available for all sites, let alone experimental plots within sites. They acknowledged that net nitrification rates, soil phosphorus, and soil-water content are important mechanistic predictors of plant community production and species richness, but the data were unavailable. Finally, although Clark and colleagues included net N mineralization data, they acknowledged that they had to integrate results generated from multiple sampling protocols (e.g., field and laboratory soil incubations), which may have affected the accuracy of their results. Because of these issues, they were left with few soil parameters on which to base their conclusions.

The problem of assembling comparable soil biogeochemistry data across sites is notorious, because often, there are a range of analytical techniques that could be employed to measure a particular chemical constituent, resulting in data that are operationally defined and often incomparable (Quevauviller 1998). This problem of diverse protocols persists with biological parameters as well. In their analysis of ANPP across North American biomes, Knapp and Smith (2001) reported that site-based studies used unique methods across all 11 sites included, making direct comparisons difficult. Estimating ANPP has the additional complication of requiring different methods based on vegetation type. However, there is the potential to standardize across biome types. Unfortunately, this potential does not exist with comparing measurements of soil processes or properties. When there is the potential for long-term data sets to be collected of above- or belowground ecosystem properties, it is worth considering that maintaining consistent approaches across space and through time may be better than using one's unique, preferred protocol or trying to incorporate the latest methodologies. With respect to adopting new methods for core (i.e., part of the long-term design) network measurements, it is often better to use outdated methodologies than to have incomplete data sets, because the value of network data is its continuity across space and time.

Carbon-cycle modeling efforts require many different data sets collected at multiple spatial and temporal scales, often from a range of sources (figure 1). The challenges in this research include a dearth of critical ecosystem parameters, as well as limited spatial and temporal coverage in network databases (figure 2). For example, models of aboveground C cycling would benefit from observations such as leaf area index (LAI) and foliar N, which are important for determining surface fluxes and photosynthetic capacity but are not necessarily available from the same ecosystems where eddy flux tower data are collected. Foliar N, in particular, is important for estimating V_{cmax} and J_{max} for the Farquhar photosynthesis model included in many ESMs (Kattge et al. 2009). Missing data on these ecosystem parameters from sites co-located with flux towers restricts the use of flux tower network databases for model testing and requires that modelers estimate values on the basis of limited data available from similar—although not co-located—ecosystems, as was discussed in Bonan and colleagues (2014).

Comparable efforts to model belowground C cycling have come about as recent analyses reveal substantial mismatches between modeled and observed soil C, highlighting the need to resolve major uncertainties in soil organic C stocks and changes over time (see Torn et al. 1997, Goidts et al. 2009, Todd-Brown et al. 2013, Jandl et al. 2014). However, data sets that describe belowground properties and processes across meaningful edaphic and climatic gradients are even more difficult to find than those describing aboveground canopy properties (figure 1b). For example, some data on microbial traits that allow for mechanistic representation of soil biogeochemical processes are beginning to emerge (e.g., German et al. 2012, Frey et al. 2013, Haggerty et al. 2014). Incorporating these data into modeling frameworks presents other challenges, because most commonly used models do not explicitly consider microbial activity or physiology (Schmidt et al. 2011, Wieder et al. 2013, 2015a). Overcoming the difficulties of aligning disparate data streams from a number of sources (e.g., networks, databases, and individual investigators) is a feature of such modeling studies, especially those geared toward representing biogeochemical processes at broad spatial scales. However, advancement in this and other research areas within biogeochemistry is slowed by not having physically co-located key ecosystem parameters above- and belowground, data resources that are not publicly available or reported with relevant metadata, and decentralized systems for data reporting and access, as we summarize in figure 2.

Finally, it can be very difficult for community researchers to conduct companion studies at or across multiple sites within a network. At NEON, there is continued debate about how to set up observational plots that will sustain at least 30 years of sampling while also fulfilling its requirement to incorporate and facilitate investigator-led projects. Concerns about long-term site management preclude widespread community engagement and the development of new research projects at these sites. Moreover, many of the measurements that NEON is making—especially with respect to terrestrial biogeochemistry—are best categorized as *site characterization* (e.g., broad soil surveys and periodic measurement of foliar canopy chemistry). NEON's terrestrial design does not include process-based measurements of biogeochemical cycles, with the exception of soil and ecosystem CO_2 fluxes at some locations.

Because measurements of soil biogeochemical transformations, trace gas fluxes, and soil-water transfers and chemistry are challenging to measure in a network context, outside investigators should be encouraged to conduct companion studies at network sites. Their detailed measurements would benefit from the characterization of soils and plant communities that NEON will provide. In turn, outside investigators often bring a deep knowledge of a particular site or ecological community, which is crucial to informing long-term network sampling designs. One opportunity is to initiate investigator-led research on changes to microbially mediated biogeochemical processes across broad gradients (e.g., climate, biomes, atmospheric deposition), which is important for advancing our ability to represent and predict soil C storage (figure 1b; Wieder et al. 2015b). Such an effort does not need to occur at NEON sites per se but could be conducted using a standardized approach across any of the networks whose sites span such gradients. Promoting community engagement is important for all network types—bottom-up, top-down, and hybrid—and requires both logistical assistance from networks and targeted funding calls.

What do we want from network resources?

There are opportunities to improve usability of network resources for the larger data user community by first making adjustments to the design, operation, and data curation within current network structures (i.e., intranetwork modifications). Summarized in figure 2, these improvements would alleviate many of the issues we have highlighted and bring networks closer to attaining the ideal characteristics. First, for basic, core measurements that are collected across sites within a network, making attempts to standardize the instrumentation or methods of collection and processing (e.g., from soil temperature and moisture sensors to analysis of leaf tissue N content) is crucial. This requires some awareness and flexibility on the part of individual investigators working at those sites, along with the enforcement of guidelines by site leads and funding agencies.

There have been many attempts within networks to identify a common suite of measurements that would be beneficial across sites beyond those related to isolated short- or long-term studies. Periodically, the LTER has visited this issue, resulting in a few methodological guides for the sites (e.g., Robertson et al. 1999, Fahey and Knapp 2007), but generally, there is not consensus on standardized measurements within the network. Recent efforts within the CZO have explored how sites should be designed for long-term measurement of the critical zone (e.g., Brantley et al. 2015, Chorover et al. 2015) and discussed suites of measurements across the network for understanding particular aspects of critical zone science, such as microbial communities or biogeochemistry. However, these efforts are only in their initial stages. We do not suggest that networks, especially those with bottom-up structures, eliminate all ability of investigators to innovate or bring new methodologies to sites but rather that intranetwork standardization be a planning topic regularly visited by site leads and considered a target goal for core measurements. In particular, we believe greater attention to standardizing soil analyses is paramount in the near term.

Second, it is important to provide better incentives for investigators to process and make their data publicly available, including digital object identifiers (DOIs) for data sets and other forms of citation/recognition. With respect to the latter, including the publication of data as an essential community service that is part of one's evaluation in tenure-track faculty or staff-scientist positions would be a means to help shift the culture from treating data as completely proprietary. Data reporting should include the publication of adequate metadata, including the methods of field collection and laboratory analysis, as well as basic sample location and timing information. Apart from the minimum metadata, individual networks may identify other useful fields to include, but they should be done so consistently across all sites within the network using commonly agreed-on terminology and bases for assessment.

Finally, we believe that resources should be allocated to support more studies by individual (or small groups of) investigators within and especially across network sites. These types of studies are encouraged across many of the networks discussed here, but there are limited calls to fund them. Specifically, research that focuses on illuminating particular soil-biogeochemical mechanisms, or changes in the distribution of microbial functional groups in response to climate or other drivers, would help network science (e.g., through community engagement), data users (e.g., C cycle modelers interested in representing microbial processes in ESM frameworks; figure 1b), and broader (i.e., spatial and temporal) understanding of changes to biogeochemical cycles. Realistically, these types of observational studies or manipulative experiments are better carried out by the research community rather than as part of core network sampling designs. NutNet is an example of a coordinated network effort to carry out a cross-site experiment measuring the effects of nutrient additions on plant productivity and diversity. However, the network is specifically designed to conduct this experiment and is focused on one ecosystem type (grasslands).

Coupled to improvements at the scale of individual networks, there are two primary cross-network initiatives that would benefit researchers asking questions in biogeochemistry, as well as other data users in the earth and environmental sciences. The first is a common, standardized suite of measurements across networks. This suggestion takes our assertion that networks should attempt to standardize methodologies when possible a step further. When it comes to defining this suite of measurements, the answer to the question "what do we want from network resources?" is difficult. Do we need consistent characterization of ecosystem properties across space and time? Do we need a new experiment? Do we just need more sites? Not only are ecosystems complex, but as question-driven scientists, the key set of measurements is often based on what is needed to address one's research questions. It is simply not possible to measure everything everywhere all the time. With the exception of NEON, which is designed to have the same suite of measurements across all of its sites, the community of investigators at particular network sites—say, within the LTER or CZO—have driven instrumentation and data-collection choices. Their measurement designs are predominately optimized for the research questions of individual investigators and selected on the basis of site-specific conditions. This organic development has been the norm.

We suggest that across networks, above-to-belowground profiles of properties in the plant canopy through, at minimum, the top 1 meter of soil (i.e., using the standards of the

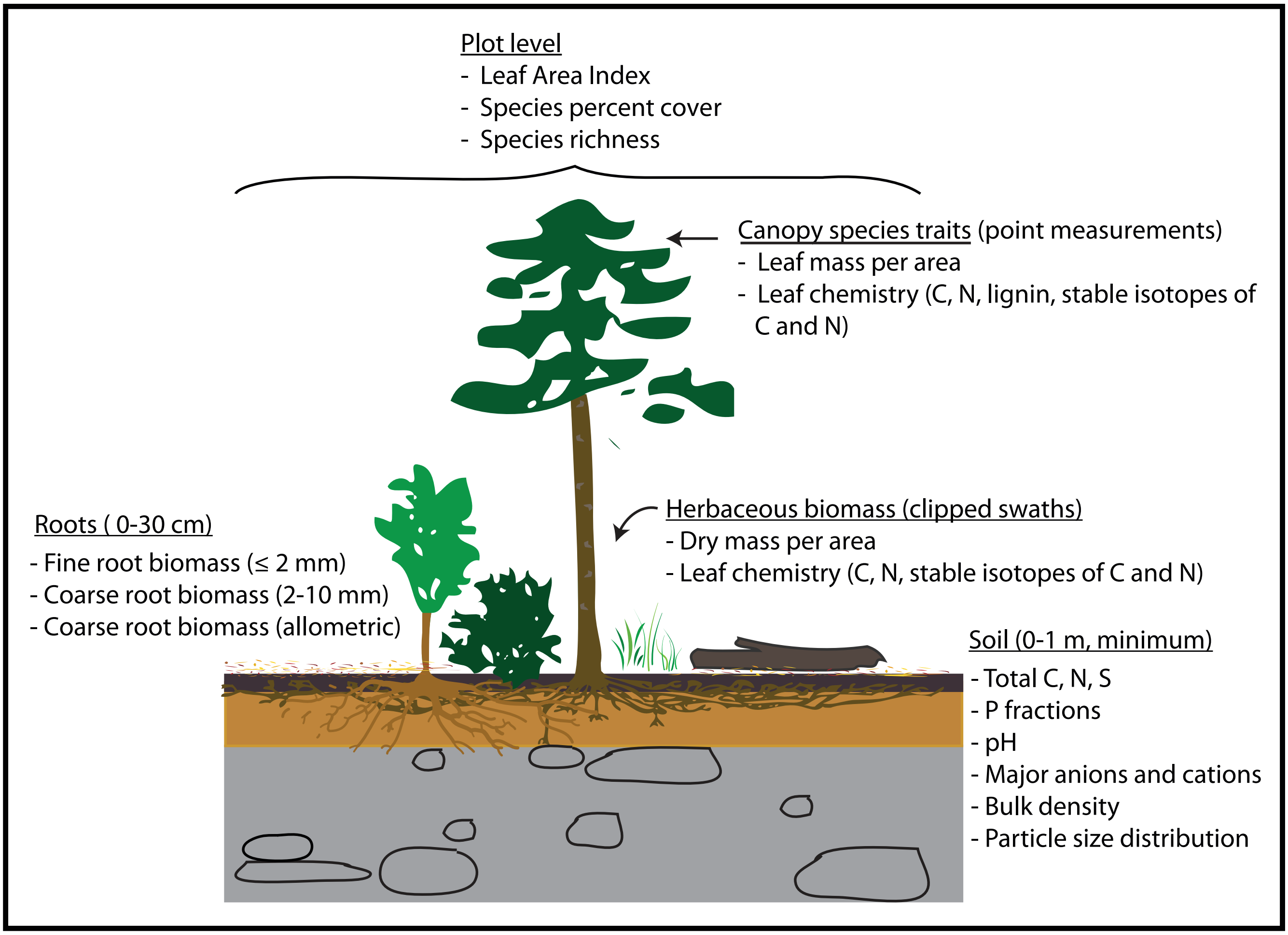

Figure 3. Above-to-belowground measurements of canopy and soil properties that could be collected across networks using standardized methods available online through the National Ecological Observatory Network. This data set would be of broad use to investigators at networks focused on collecting biogeochemical data and their data user communities. Source: Adapted with permission from original artwork by Courtney Meier.

National Resources Conservation Service soil survey; see Schoeneberger et al. 2012) be measured at least once every 10 years. These data would be very useful to future observational and modeling studies of terrestrial biogeochemistry (figure 3). Moreover, these above-to-belowground profiles are aligned with the science themes and interests of many networks that seek to explain some aspect of ecological patterns or subsurface architecture and evolution. Parameters measured could include the common canopy chemistry and structure, as well as soil chemical and physical properties summarized in figure 3. These measurements are already underway at operational NEON sites, and we feel that coordinated profile measurements at multiple points across other networks would be useful not only to researchers working within those sites but also to the broader community of data users. Coordinated, distributed networks such as NutNet and AmeriFlux might choose to contribute some fraction of the total profile of measurements but use the same methodologies so that the resulting data could be integrated into a common database and easily incorporated into cross-network analyses. Would such profiles provide answers to all of the outstanding questions in biogeochemistry? Of course not, but the data set would be valuable to the analysis and design of a lot of investigator-led, process-based research. In addition, they have the potential to provide ground-truth for remote sensing data collected across network sites.

Coordinating standardized measurements across networks would not be successful without the third cross-network initiative to increase the usability of data resources: an investment in an online data portal that integrates network data streams with individual-investigator studies. Currently, networks have mandates from their funding agencies to make data publicly available via a data portal developed by and for that network, often at the site level. However, the broad community of data users would benefit greatly from online tools that allow the exploration of data streams from across project types, providing easy access and exploratory capacity (e.g., simple visualization, as well as basic statistical and time-series analyses). With respect to representing C cycle processes in ESMs, including data from experimental

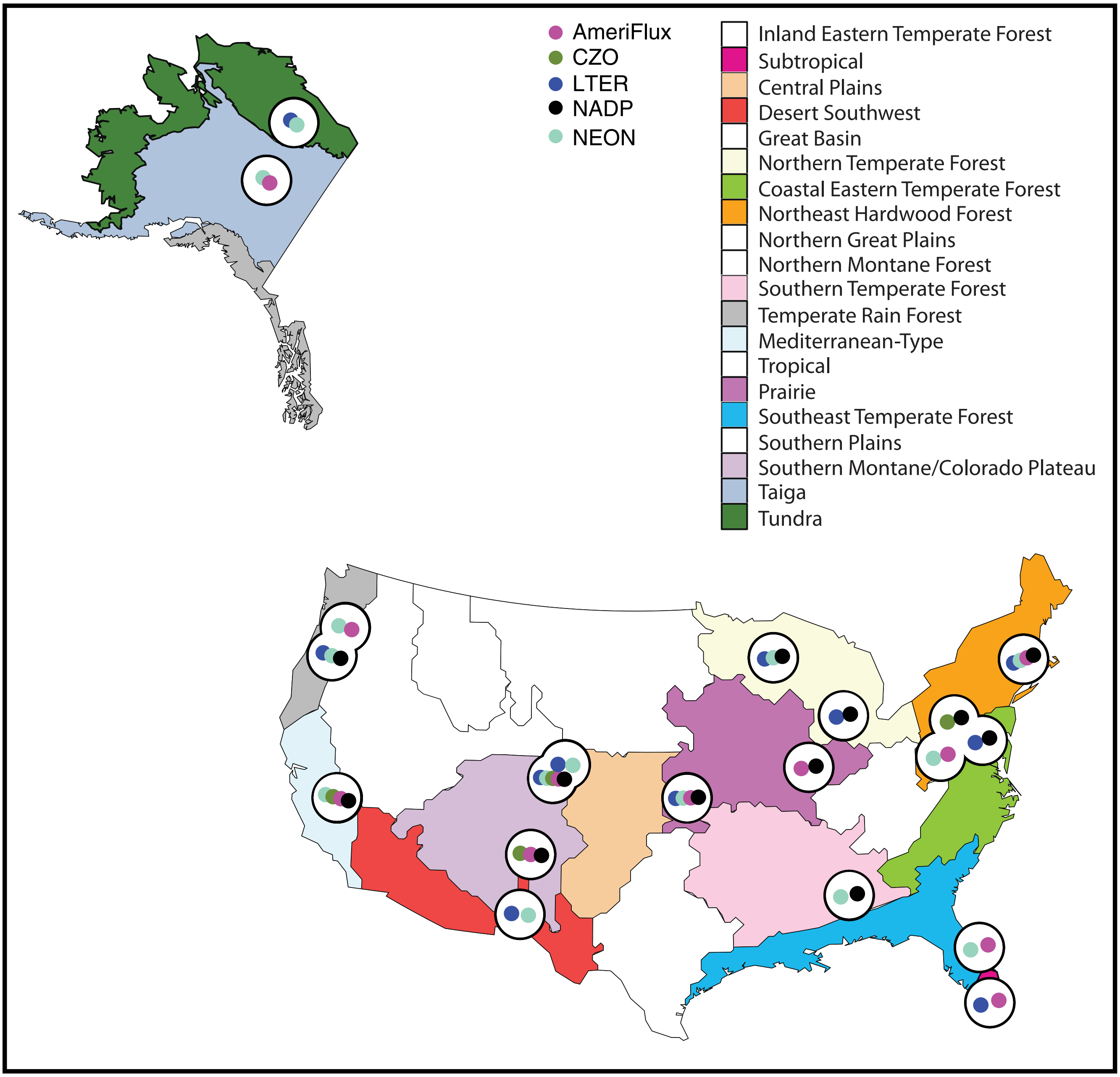

Figure 4. A map of physically co-located sites from five networks making biogeochemical measurements in the United States: The Long-Term Ecological Research (LTER) network, The National Ecological Observatory Network (NEON), the Critical Zone Observatories (CZO) network, AmeriFlux, and The National Atmospheric Deposition Program (NADP). Each of these networks offers different data resources to address questions in biogeochemistry. National Ecological Observatory Network ecoclimatic domains (labeled with the names of biome types) are shown in unique colors. For clarity, we omit site locations within each network that are not co-located with another network, and biomes that do not have overlapping networks are shown in white.

manipulations (e.g., the Free-Air CO_2 Enrichment, FACE; NutNet; and The International Tundra Experiment, ITEX) are crucial to improve and evaluate the accuracy of future model projections. Ideally, a cross-network online data exploration and analysis portal would connect to coordinated experiments; more narrowly focused voluntary data repositories, such as TRY (leaf traits); citizen-science projects; and global, large-scale databases, such as the ISCN. In addition, it could include an interactive map that shows the physical alignment of network data resources. We provide a version of this map, showing overlapping sites within some of the networks highlighted in this article (figure 4).

Online portals that integrate data from different sources and disciplines exist now, in projects such as DataONE (*www.dataone.org*) and its nodes (e.g., KNB Informatics, *https://knb.ecoinformatics.org*), as well as Figshare (*http://figshare.com*). A common criticism of these databases is that there is a trade-off between ease of uploading by investigators and ease of exploration by users. That is, databases that require time-consuming data-entry procedures by investigators result in highly searchable and easily explored resources for users versus those that simply require submission of a data file that users must figure out how to read and compile with other sources. It is worth funding agencies considering the allocation of resources to reduce these barriers to data reporting, access, and exploration and to encourage the development of a cross-network portal, possibly based at one of the existing open databases. If this portal had the capacity to connect with established databases, incorporate individual investigator-led projects, and citizen-science efforts—no current portal does—it would be an extremely useful resource to data users within and outside of science.

Conclusions

Although we have focused on barriers and potential solutions to using network data in studies of the terrestrial C cycle and biogeochemistry more broadly, this discussion is applicable to many research areas within the earth and environmental sciences. Network data are most powerful when structured to allow for comparisons in space and time. Therefore, network development would ideally focus on balancing the standardization of measurements with the flexibility to adopt new methods and integrate innovative investigator-led research into existing operations and infrastructure, both physically and virtually (i.e., online). Many of the suggestions provided here would require a minimal investment of new funds into network science. Instead, they require organizations operating networks and site leads to give careful consideration to how data will ultimately be used by network affiliates and community data users, as well as attention to standardization and making data publicly available quickly after collection.

Acknowledgments

This manuscript was the result of a cross-network workshop (not funded by any agencies) held at the Institute of Arctic and Alpine Research (INSTAAR) at the University of Colorado, Boulder, on 11 November 2014. The attendees/coauthors are affiliated with at least one of the following five networks: AmeriFlux, CZO, LTER, NEON and NutNet, and/or are active network data users. The objective was to discuss the use of network data in the field of biogeochemistry and convened because of the attendees' interest. The authors thank Pierre Glynn and the three anonymous reviewers for helpful comments on earlier versions of this manuscript.

References cited

Anderson SP, Bales RC, Duffy CJ. 2008. Critical Zone Observatories: Building a network to advance interdisciplinary study of Earth surface processes. Mineralogical Magazine 72: 7–10.

Banwart SA, et al. 2013. Sustaining Earth's Critical Zone: Basic Science and Interdisciplinary Solutions for Global Challenges. University of Sheffield.

Bonan GB, Oleson KW, Fisher RA, Lasslop G, Reichstein M. 2012. Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4. Journal of Geophysical Research 117 (art. G02026).

Bonan GB, Williams M, Fisher RA, Oleson KW. 2014. Modeling stomatal conductance in the earth system: Linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. Geoscientific Model Development 7: 2193–2222.

Bond-Lamberty B, Thomson A. 2010. A global database of soil respiration data. Biogeosciences 7: 1915–1926.

Brantley SL, et al. 2015. Designing a suite of measurements to understand the critical zone. Earth Surface Dynamics Discussions 3: 1005–1059.

Bricker OP, Ruggiero MA. 1998. Toward a national program for monitoring environmental resources. Ecological Applications 8: 326–329.

Chorover J, et al. 2015. Common Critical Zone Observatory (CZO) Infrastructure and Measurements: A Guide Prepared By CZO PIs. National Science Foundation. (12 January 2016; *http://criticalzone.org/national/publications/pub/chorover-et-al-2012-common-critical-zone-observatory-czo-infrastructure-and*).

Clark CM, et al. 2007. Environmental and plant community determinants of species loss following nitrogen enrichment. Ecology Letters 10: 596–607.

Collins SL, Childers DL. 2014. Long-term ecological research and network-level science. Eos 95: 293–294.

Driscoll CT, Driscoll KM, Roy KM, Mitchell MJ. 2003. Chemical response of lakes in the Adirondack Region of New York to declines in acidic deposition. Environmental Science and Technology 37: 2036–2042.

Evans CD, Monteith DT, Cooper DM. 2005. Long-term increases in surface water dissolved organic carbon: Observations, possible causes, and environmental impacts. Environmental Pollution 137: 55–71.

Fahey TJ, Knapp AK, eds.. 2007. Principles and Standards for Measuring Primary Production. Oxford University Press.

[FAO] Food and Agriculture Organization of the United Nations, International Institute for Applied Systems Analysis, International Soil Reference and Information Centre, International Solid-State Circuit and System Conference, European Commission Joint Research Centre. 2012. Harmonized World Soil Database (Version 1.2) FAO, IIASA. (12 January 2016; *http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML*)

Frey SD, Lee J, Melillo JM, Six J. 2013. The temperature response of soil microbial efficiency and its feedback to climate. Nature Climate Change 3: 395–398.

German DP, Marcelo KRB, Stone MM, Allison SD. 2012. The Michaelis–Menten kinetics of soil extracellular enzymes in response to temperature: A cross-latitudinal study. Global Change Biology 18: 1468–1479.

Goidts E, Van Wesemael B, Crucifix M. 2009. Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. European Journal of Soil Science 60: 723–739.

Haggerty SB, van Groenigen KJ, Allison SD, Hungate BA, Schwartz E, Koch GW, Kolka RK, Dijkstra P. 2014. Accelerated microbial turnover but constant growth efficiency with warming in soil. Nature Climate Change 4: 903–906.

Jandl R, et al. 2014. Current status, uncertainty, and future needs in soil organic carbon monitoring. Science of the Total Environment 468: 376–383.

Jung M, et al. 2011. Global patterns of land–atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. Journal of Geophysical Research 116 (art. G00J07). doi:10.1029/2010JG001566

Kattge J, et al. 2011. TRY: A global database of plant traits. Global Change Biology 17:2905–2935.

Kattge J, Knorr W, Raddatz T, Wirth C. 2009. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. Global Change Biology 15: 976–991.

Knapp AK, Smith MD. 2001. Variation among biomes in temporal dynamics of aboveground primary production. Science 291: 481–484.
Mervis J. 2015. Ecology's tough climb. Science 349: 1436–1441.
———. 2015b. NEON jobs plentiful but problematic. Science 349: 1441.
———. 2015c. Tragic end for Puerto Rico site. Science 349: 1440.
Murdoch PS, McHale M, Baron JS. 2014. Reflections on a vision for integrated research and monitoring after 15 years. Aquatic Geochemistry 20: 363–380.
O'Neel S, et al. 2015. Icefield-to-ocean linkages across the Northern Pacific Coastal Temperate Rainforest Ecosystem. BioScience 65: 499–512.
Peters DPC, Loescher HW, SanClements MD, Havstad KM. 2014. Taking the pulse of a continent: Expanding site-based research infrastructure for regional- to continental-scale ecology. Ecosphere 5 (art. 29).
Porcal P, Koprivnjak JF, Molot LA, Dillon PJ. 2009. Humic substances—part 7: The biogeochemistry of dissolved organic carbon and its interactions with climate change. Environmental Science and Pollution Research 16: 714–726.
Quevauviller P. 1998. Operationally defined extraction procedures for soil and sediment analysis I. Standardization. TrAC Trends in Analytical Chemistry 17: 289–298.
Robertson GP, Coleman DC, Bledsoe CS, Sollins P. 1999. Standard Soil Methods for Long-Term Ecological Research. Oxford University Press.
Rosenberg MS, Rothstein HR, Gurevitch J. 2013. Effect sizes: Conventional choices and calculations. Pages 61–71 in Koricheva J, Gurevitch J, Mengersen K, eds. Handbook of Meta-Analysis in Ecology and Evolution. Princeton University Press.
Schmidt MW, et al. 2011. Persistence of soil organic matter as an ecosystem property. Nature 478: 49–56.
Schoeneberger PJ, Wysocki DA, Benham EC, Soil Survey Staff. 2012. Field Book for Describing and Sampling Soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center.
Todd-Brown KEO, et al. 2013. Causes of variation in soil carbon predictions from CMIP5 Earth system models and comparison with observations. Biogeosciences 10: 1717–1736.
Torn MS, Trumbore SE, Chadwick OA, Vitousek PM, Hendricks DM. 1997. Mineral control of soil organic carbon storage and turnover. Nature 389: 170–173.
Wieder WR, Bonan GB, Allison SD. 2013. Global soil carbon projections are improved by modelling microbial processes. Nature Climate Change 3: 909–912.
Wieder WR, Grandy AS, Kallenbach CM, Taylor PG, Bonan GB. 2015a. Representing life in the Earth system with soil microbial functional traits in the MIMICS model. Geoscientific Model Development 8: 1789–1808.
Wieder WR, et al. 2015b. Explicitly representing soil microbial processes in Earth system models. Global Biogeochemical Cycles 29: 1782–1800.

Eve-Lyn S. Hinckley (eve.hinckley@colorado.edu) is a fellow at the Institute of Arctic and Alpine Research and assistant professor of environmental studies at the University of Colorado, Boulder. Suzanne P. Anderson is a fellow at the Institute of Arctic and Alpine Research and associate professor of geography at the University of Colorado, Boulder. Jill S. Baron is a senior scientist at the United States Geological Survey, in Fort Collins, Colorado. Peter D. Blanken is an associate professor of geography and Jason C. Neff is a professor of environmental studies at the University of Colorado, Boulder. Gordon B. Bonan is a senior scientist and Danica L. Lombardozzi and William R. Wieder are project scientists at the National Center for Atmospheric Research, in Boulder, Colorado. William R. Wieder is also a research affiliate at the Institute of Arctic and Alpine Research. William D. Bowman is a fellow at the Institute of Arctic and Alpine Research and professor of ecology and evolutionary biology at the University of Colorado, Boulder. Sarah C. Elmendorf, Andrew M. Fox, Keli J. Goodman, Claire K. Lunch, and Michael D. SanClements are staff scientists and Katherine D. Jones is an associate plant ecologist at the National Ecological Observatory Network in Boulder, Colorado. Noah Fierer is a fellow at the Cooperative Institute for Research in Environmental Science and associate professor of ecology and evolutionary biology at the University of Colorado, Boulder. Katharine N. Suding is a fellow at the Institute of Arctic and Alpine Research and professor of ecology and evolutionary biology at the University of Colorado, Boulder.