



## Ideas and perspectives: Addressing environmental challenges using distributed data generation: From collaborative networks to artificial intelligence-enabled science

5 Amy E. Goldman<sup>1</sup>, James C. Stegen<sup>2,3</sup>, Elizabeth T. Borer<sup>4</sup>, Robert Hensley<sup>5</sup>, Allison Myers-Pigg<sup>6</sup>, Mikayla A. Borton<sup>7</sup>, Paytsar Muradyan<sup>8</sup>, Nicholas D. Ward<sup>9</sup>, Jean M. Andino<sup>10</sup>, Andrew D. Richardson<sup>11,12</sup>, Alexandro B. Leverkus<sup>13</sup>, Christopher S. Lowry<sup>14</sup>, Nate G. McDowell<sup>2,3</sup>, Kaizad F. Patel<sup>2</sup>, Scott J. Davidson<sup>15</sup>, Brienne Forbes<sup>1</sup>, Martha R. Downs<sup>16</sup>, W. Robert Bolton<sup>17</sup>, Timothy D. Scheibe<sup>2</sup>, and Avni Malhotra<sup>2</sup>

10

1. Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, 99352, USA

2. Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA

3. School of the Environment, Washington State University, Pullman, WA, 99163, USA

4. Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN, 55108, USA

15

5. National Ecological Observatory Network, Battelle, Boulder, CO, 80301, USA

6. Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Sequim, WA, 98382, USA

7. Department of Food Science and Human Nutrition, Colorado State University, Fort Collins, CO, 80526, USA

8. Environmental Science Division, Argonne National Laboratory, Lemont, IL, 60439, USA

9. Energy and Environment Directorate, Pacific Northwest National Laboratory, Sequim, WA, 98382, USA

20

10. Arizona State University, School for the Engineering of Matter, Transport and Energy, Tempe, AZ, 85226, USA

11. School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, 86011, USA

12. Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, 86011, USA

25

13. Departamento de Ecología, Facultad de Ciencias, Universidad de Granada, Av. Fuentenuova s/n, 18071, Spain

14. University at Buffalo, Department of Earth Sciences, Buffalo, NY, 14260, USA

15. Département des sciences biologiques, Université du Québec à Montréal, Montréal, QC, H2X 1Y4, Canada

16. Long-Term Ecological Research Network, National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, 93101, USA

30

17. Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37830, USA

*Correspondence to:* Amy E. Goldman ([amy.goldman@pnnl.gov](mailto:amy.goldman@pnnl.gov)) and Avni Malhotra ([avni.malhotra@pnnl.gov](mailto:avni.malhotra@pnnl.gov))

35

**Abstract.** Distributed data generation, or data collected from multiple sources and locations using standardized approaches and involving coordination among investigators, has emerged as a powerful approach to meet contemporary demands for scalable environmental knowledge. However, practitioners often lack guidance on best practices for distributed data generation, and a framework classifying its modalities is missing.

40

To address these gaps, we developed a conceptual framework organizing distributed data generation along two axes: participant-based (ranging from highly formalized to highly flexible) and method-based (from experimental to observational). This framework provides common vocabulary across modalities and describes how different approaches affect data generation logistics and outcomes. We propose operational best practices across three critical pillars: outreach, operations, and output (i.e., publications, data), leveraging lessons learned from over 35 existing



distributed data projects. Lastly, we explore how emerging artificial intelligence (AI) capabilities may help address  
45 longstanding challenges in distributed data generation, including in coordination, adaptive sampling, and cross-  
project data integration. This perspective provides strategies and identifies opportunities to advance distributed data  
generation for addressing pressing biogeochemical, environmental, and societal challenges. We underscore the  
transformative potential of distributed data generation for modern, broad-scale environmental research, and provide  
guidance on how to realize that potential.

## 50 **1 Introduction:**

Society faces increasingly complex, multi-scale, and interconnected environmental challenges. Predicting the  
coupled dynamics among human activities, extreme events, and ecosystem functions requires extensive datasets  
capable of capturing variability across space and time (Caldararu et al., 2023; Hanson and Walker, 2020).  
Assembling datasets of this magnitude requires coordination beyond the scope of any single investigator or  
55 traditional research project. Distributed data generation has emerged as a powerful approach to meet contemporary  
scientific and societal demands for scalable, actionable knowledge across broad geographic and ecological domains  
(Borer et al., 2014; Peters et al., 2014a; SanClements et al., 2022; Stegen and Goldman, 2018). Drawing on existing  
efforts from disciplines such as biogeochemistry, ecology, and engineering, we define distributed data generation as  
*data collected from geographically dispersed locations using standardized approaches, and involving coordination*  
60 *among numerous investigators.*

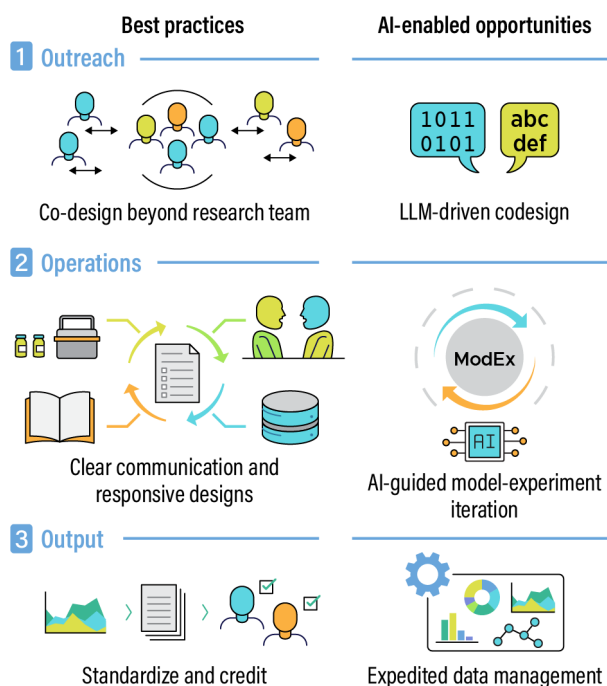
Distributed data generation has transformed the potential to address scientific challenges across scales, from  
microbes to forests, from plots to continents, and from households to regions. Researcher-enabled efforts have  
advanced ecosystem and Earth system predictability, such as understanding global trends in diversity and  
65 productivity via NutNet (Borer et al., 2014), stream biogeochemistry via WHONDRS (Stegen and Goldman, 2018),  
plant phenology via PhenoCam (Young et al., 2025), marine biogeochemistry via GEO-TRACES (Anderson et al.,  
2014), and many more.

While advances have been made in distributed data generation, numerous challenges remain. A lack of shared  
70 vocabulary for distributed efforts, even within a single domain, has limited the communication of lessons learned,  
challenges, and opportunities to integrate projects and fields (SanClements et al., 2022). A conceptual framework of  
the different modalities of distributed data generation would aid in the development of new projects and  
enhancement of existing efforts. Furthermore, systemic issues that influence science in general tend to influence  
distributed science disproportionately. For example, ensuring broad participation and codesign with interested  
75 parties (i.e., stakeholders, organizations, agencies, landowners, land managers, rights holders, and the public)  
remains a key limitation for distributed data generation (Molla et al., 2025; Muraina and Jimoh, 2022). This hinders  
researchers from clearly mapping their goals with community needs. Scaling distributed efforts is further hindered  
by logistical challenges in consistency, data standardization, and transparent expectation setting and engagement  
(Wallace, 2022). Advances in computing, sensors, and artificial intelligence (AI) may ease some barriers, but the



80 large number of contributors makes coordination of data generation especially difficult, while also being one of the  
key aspects that enables progress.

Here, we address a critical gap in distributed data generation, namely the lack of an integrated framework that  
classifies the breadth of distributed data generation approaches and of an overview of best practices and how  
85 emerging technological opportunities may accelerate solutions to environmental and societal challenges.  
Specifically, we (1) define core modalities of distributed data generation to provide a common vocabulary across  
effort types, (2) recommend operational best practices that have proven successful, and (3) propose how AI can  
address longstanding challenges in coordination, adaptive sampling, and cross-project data integration (Fig. 1). We  
limit our discussion of distributed data generation to efforts with human-to-human coordination requiring on-the-  
90 ground efforts. The insights reflected in this Perspective emerged from a workshop of over 25 distributed data  
generation practitioners.



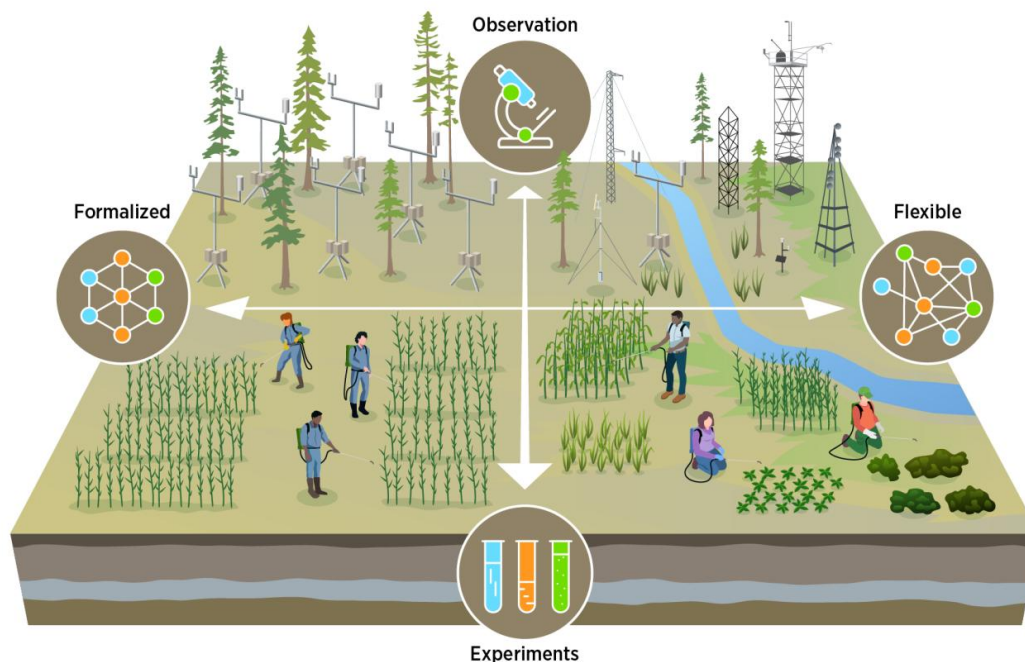
**Figure 1.** Distributed data generation has made transformative progress in cross-scale, cross-disciplinary science and is supported by best practices in (1) outreach, including co-design beyond the research team using multi-model  
95 engagement; (2) operations, including clear communication and expectation setting to allow nimble adjustments; (3) output, including transparent credit and recognition for FAIR products. AI is enabling rapid opportunity development across distributed data generation (Section 4) with large language model (LLM)-driven design, AI-guided model-experiment iteration (ModEx), and efficiency in data management. Acronyms: FAIR: Findable, Accessible, Interoperable and Reusable.



100 **2 A conceptual framework on the modalities of distributed data generation**

An existing body of literature on research networks and collaborative science offers important foundations for distributed data generation approaches, including practical recommendations (Maestre and Eisenhauer, 2019), criteria for successful collaborations (Parker and Kingori, 2016), and models of coordinated distributed experiments (Borer et al., 2014). However, this literature lacks an overarching framework, limiting the possibility to identify and  
105 synthesize such networks. The body of literature resulting from distributed networks is diffuse, spanning a range of network types, from those focused primarily on connecting people to those centered on data generation, and encompassing varying levels of formalization and standardization. This diversity of architectures and the dispersed nature of guidance across disciplines limits the capacity to design or improve distributed data generation efforts. We focus on the mechanics of distributed data generation –the "how" of obtaining data across multiple sites– rather  
110 than on network governance or social connectivity (Lindenmayer et al., 2018; Maestre and Eisenhauer, 2019; Richter et al., 2018; SanClements et al., 2022).

Scientific networks are often classified according to their degree of centralization in terms of people, management, and funding (Novick et al., 2018; Peters et al., 2014b; SanClements et al., 2022). Further, SanClements et al. (2022) proposed a 1:1 relationship between degree of centralization (funding, management, education, cyberinfrastructure)  
115 and degree of standardization (sensors, protocols, algorithms). Our conceptual framework (Fig. 2) builds upon these prior frameworks, collapsing the degree of standardization and centralization into a single axis. However, unlike the SanClements et al. (2022) framework, our focus in the first axis is on the participants generating the distributed data, rather than on the management approach. We add a second axis that has an outsized influence on the practicalities of data generation, namely the degree of observational *versus* experimental methodology. We describe these  
120 interacting axes as modalities of distributed data generation to guide new efforts and offer a shared language for existing efforts. It is worth noting that distributed data generation efforts can be described across many additional continua such as duration of the project, funding, and geographic scope.



125 **Figure 2.** Modalities of distributed data generation conceptualized along two axes. The horizontal axis is  
participant-focused and ranges from highly formalized (left) to highly flexible (right). The vertical axis represents a  
method-based axis, ranging from observational (top) to experimental (bottom). Visual elements illustrate key  
characteristics of integrating these two continua: the left side shows standardized equipment, personnel, and plots  
typical of formalized efforts; the right side shows diverse equipment, personnel, and plots characteristic of flexible  
130 efforts; the top shows observational infrastructure; the bottom shows experimental manipulations. Most distributed  
efforts operate somewhere within this 2D space rather than at the extremes, and some evolve over time (see Section  
2 for details and examples).

### 2.1 Participant-based modalities

Participants in distributed data generation efforts can span a range of technical knowledge (i.e., researchers, non-  
135 researcher domain experts, non-domain experts) and formal engagement (i.e., hired staff, unpaid formal  
collaborators, and unpaid casual volunteers). Distributed data generation efforts often integrate across these different  
types of participants, and it can be helpful to view efforts along a participant-based continuum that ranges from  
highly formalized to highly flexible. Along the entire spectrum, there is almost always a unified body leading  
decision-making for the effort. This body may be small or large, unfunded or funded, but it links participants and  
140 data into a coordinated whole.



### 2.1.1 Highly formalized

Distributed data generation efforts that are structured on the formalized end of the participation continuum often have static or semi-static data types and sites, usually operated by hired staff. Therefore, there are potential limits on new types of data collection and/or new locations being added to such efforts through time. Examples include the U.S. National Ecological Observatory Network (NEON) (Keller et al., 2008), Europe's Integrated Carbon Observation System (ICOS; Heiskanen et al., 2022), and U.S. Geological Survey Streamgaging Network (Eberts et al., 2019). The funding of these entities and employment of hired staff supports frequent release of consistent data. Because they often have a distinct set of data they generate through time coordinated via a core data management team, such efforts provide benchmarks for multi-year trends and extremes, as well as calibration/validation of remote-sensing, low-cost sensors, and models of many types and scales.

### 2.1.2 Highly flexible

On the flexible end of the continuum, new data types and/or sites may be added throughout the life of the project if there is sufficient interest from collaborators and/or the community. Data are often generated by collaborators or unpaid volunteers that may or may not have research or domain expertise. In participatory efforts (i.e., community or crowdsourced science), data generation is accomplished by individual volunteers contributing observations or samples to professionally designed projects under shared protocols. Projects such as eBird (<https://ebird.org/home>) and the USA National Phenology Network's Nature's Notebook (USA-NPN; <https://www.usanpn.org/nn>) work with mostly non-researcher volunteers and demonstrate that many observations can be achieved at modest cost with clear protocols and routine feedback. Conversely, projects may rely on researcher collaborators taking one-time or repeat samples. The Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (Gary et al., 2024; WHONDRS; Stegen and Goldman, 2018), the Exploration of Coastal Hydrobiogeochemistry Across a Network of Gradients (EXCHANGE) consortium (Myers-Pigg et al., 2023), the Molecular Observation Network (MONet; Bowman et al., 2023), Phenocam (Young et al., 2025), and The PeatPic Project (Davidson et al., 2025; Young et al., 2025) provide examples of efforts in which geographically distributed researchers collect samples and/or measurements for a centralized project. Highly flexible efforts with distributed implementation are well-suited for testing generalizable hypotheses across environmental gradients and rapidly applying protocols to new regions or environmental disturbances, often supporting iteration with model development. They can provide high spatial density and/or large spatial extent of data, depending on the project's goals and participants.

### 2.1.3 Intermediates along the participant-continuum

There are many intermediate states possible between the two described endpoints. An example of a midpoint is data generation efforts that have static sites run by researchers who are not hired staff of the distributed data generation entity. In these cases, participation is unrestricted, so anyone with the necessary skills and resources can add a new site by engaging with the centralized guiding body and following the same protocols. Examples include AmeriFlux (Novick et al. 2018; Chu et al. 2023) and FLUXNET (Baldocchi et al. 2001; Delwiche et al. 2024), which use sites established/funded by individual researchers that have access to centralized data hosting, guidelines, and processing



pipelines. Globally replicated experiments such as Nutrient Network (NutNet; Borer et al., 2014) and Disturbance and Recovery Across Global Grasslands Network (DRAGNet; Borer and Seabloom, 2025) similarly have sites established by individual researchers that follow identical protocols for experimental manipulation, leading to community-generated datasets that enable broad inference.

## 180 **2.2 Method-based modalities**

Much like the participant-based modalities described above, methods of distributed data generation efforts can be considered along a continuum. We describe the endpoints as (1) experimental and (2) observational approaches.

### **2.2.1 Experimental**

185 Coordinated distributed experiments (Fraser et al., 2013) apply common manipulations across many sites to test general rules and drivers of variation across heterogeneous environmental conditions. With standardized treatments, synchronized timing, and shared metadata, initiatives such as NutNet, DRAGNet, and DroughtNet (Poyatos et al., 2021) enable causal inference beyond any one location (Borer et al., 2014; Borer and Seabloom, 2025). These standards are codified in publicly-available protocols specifying treatment formulations, sampling schedules, etc. Other examples of distributed experiments include efforts to target large-scale drivers of wood decomposition  
190 (Seibold et al., 2021), cellulose decomposition in streams (Tiegs et al., 2024), arthropod use of leaf shelters (Romero et al., 2022), invertebrate impacts on plant communities (Kempel et al., 2025) or effectiveness of revegetation methods (Leverkus et al., 2021). Core elements of distributed experiments are long-term coordination, consistent interventions, fidelity checks, and researcher engagement.

### **2.2.2 Observational**

195 Observational networks generate data from fixed and mobile platforms, as well as spatial data from hundreds to thousands of locations. While some efforts focus on single-timepoint measurements across a broad geographic range (e.g. WHONDRS, MONet), others also include a temporal component of repeated sampling/measurements, to get long-term time series (e.g. NEON, PhenoCam). As described above, AmeriFlux has individual researcher-led distributed sites that contain sensor installations (i.e., flux towers) generating time series data. In contrast, NEON  
200 exemplifies a staffed observatory with cross-site standardization (Elmendorf et al., 2016; Meier et al., 2023), but it similarly operates through established distributed sites with integrated airborne, sensor-based, and observational sampling efforts. In the Mountain Invasion Research Network (MIREN; Kueffer et al., 2014), a similar approach is applied to monitoring mountain species distributions (Haider et al., 2022). These networks underpin process studies, systematic model evaluation and data assimilation using continuous, comparably processed time series and/or  
205 interoperable, highly distributed spatial observations.

### **2.2.3 Intermediates along the method-continuum**

Along the continuum of methods, there are examples of efforts that use both approaches simultaneously, such as the Long-Term Ecological Research (LTER) Network and the National Estuarine Research Reserve System (NERRS)



210 Science Collaborative, (Elizabeth L. Wenner Maggie Geist, 2001; Knapp et al., 2012; Raposa et al., 2023) and also  
efforts that use one approach initially and then transition to the other (e.g., observational networks collecting  
samples that are then used for laboratory experiments; locations for experiments identified after initial observations).  
EXCHANGE (Myers-Pigg et al., 2023) is an example of a hybrid methods modality, wherein a network of coastal  
researchers provide observational data on soils (Pennington et al., 2023) across sites and community members have  
the opportunity to propose experiments that can then be conducted in a standardized way across the network. Many  
215 distributed experiments also collect control plot data, producing a standardized observational dataset within the  
framework of an experiment (Borer and Seabloom, 2025).

### 3 Best practices

In this section, we provide best practices gleaned from the perspectives of attendees of a virtual workshop held in  
2025, as well as from assessing common successful approaches across over 35 distributed data generation efforts.  
220 Our recommendations span three critical aspects of distributed data generation projects: outreach, operations, and  
output (Fig. 1). We highlight some corresponding examples of distributed data generation projects in Table 1.

**Outreach** best practices include engaging with people beyond the immediate research team for education, data  
generation opportunities, and co-design, with a critical eye to ensuring broad engagement (Maestre and Eisenhauer,  
225 2019). Co-designed projects often yield outputs that communities can trust and act on (National Academies of  
Sciences, Engineering, and Medicine et al., 2018; Norström et al., 2020). However, meaningful co-design requires  
substantial time investment, resources for community engagement, and institutional commitment (Table 1). Not all  
distributed efforts have the capacity for deep co-design, and practitioners must balance aspirational goals with  
practical constraints. One way to maximize engagement across interested parties is to use multi-modal engagement.  
230 For example, AmeriFlux uses workshops, annual meetings, and virtual seminars to engage their community. Multi-  
modal engagement also helps build trust across participants and reduces burden on one set of participants to  
continuously engage, thus minimizing attrition. Maximizing geographical coverage of distributed data generation  
efforts is also often a goal, especially given that many of our existing networks and databases are heavily biased to  
western Europe and North America (Dwivedi et al., 2022; Hughes et al., 2021; Malhotra et al., 2026; Skaldina and  
235 Blande, 2025). This requires extensive and targeted searches for participants. Engaging the general public can help  
with this but is not always practical. If considering this route, reducing the barriers for meaningful engagement is a  
best practice that several efforts have reported (Table 1). For example, enabling physical sample collections or  
public access to project data in user-friendly formats, are both ways that can improve engagement from the general  
public and also help maximize spatial coverage.

240

**Operations** includes practices for clear communication, governance, and implementation across collaborators.  
Strong operational planning acknowledges that good science cannot be separated from the infrastructure and  
communities that support it. Collaborative science is most effective when it is embedded in the research design from  
the outset (U.S. DOE 2026a). Best practices in operations include open communication (Iversen et al., 2020), clearly



245 defined roles and responsibilities followed by appropriate recognition (Leverkus et al., 2021), clear mechanisms for  
 engagement (Borer et al., 2023), and others (Table 1). Clear communication also enables the scientific goals to  
 remain somewhat nimble and responsive to emerging questions, conditions and needs of interested parties, though  
 the degree of responsiveness varies substantially depending on project structure and resources. For example,  
 WHONDRS adapts their sampling approach based on environmental uniqueness of their existing sites and regions  
 250 that have high error in predictive models (Hoffman et al., 2013; Malhotra et al., 2025).

**Output** includes generating and publishing data, scientific papers, models, and other products. Best practices around  
 output can transform distributed projects from ad hoc collections of data into enduring resources that amplify  
 scientific impact and may enable effective use of data in decision-making by communities that engage in research  
 255 co-design. In highly formalized efforts such as NEON, established standardized protocols ensure that data and  
 metadata collection remains consistent and comparable, both across sites, and through time (Atkins et al., 2025). For  
 less formalized efforts that rely on distributed volunteers to make measurements (e.g., EXCHANGE, AmeriFlux),  
 establishing common sampling kits, training materials, and metadata standards can help maximize the potential of  
 the data and make broad-scale comparisons possible (Myers-Pigg et al., 2023). Centralized and relational data  
 260 repositories can further aid in data findability and scientific discovery (Chu et al., 2023). It is critical in all cases to  
 ensure transparent methods for recognition and credit for all contributors in data and publications (Borer et al.,  
 2023). Last but not the least, making data publicly available as early as possible, ideally before manuscript  
 development, ensures that data are equally available to the entire team and beyond (Pennington et al., 2023).

265 **Table 1:** Demonstrated best practices in distributed data generation with example resources. Examples are not  
 intended to be an exhaustive list or to indicate a preference for one project over another but rather provide a few  
 tangible resources for each best practice.

Best Practice	Example
Outreach	
Co-design with interested parties to identify research questions and local priorities, make design decisions, and focus protocols on feasibility	The CROCUS ( <a href="https://www.anl.gov/crocus">https://www.anl.gov/crocus</a> ; Collis et al., 2025) SW-IFL Urban Integrated Field Laboratories ( <a href="https://sw-iftl.asu.edu">https://sw-iftl.asu.edu</a> ) Tracking the Colour of Peatlands (Color of Peat; Davidson, 2021)
Leverage multi-modal communications: Webinar series, workshops, annual meeting	Ameriflux: <a href="https://ameriflux.lbl.gov/community/ameriflux-meetings-workshops/">https://ameriflux.lbl.gov/community/ameriflux-meetings-workshops/</a> NEON: <a href="https://www.neonscience.org/resources/learning-hub">https://www.neonscience.org/resources/learning-hub</a> LTER: <a href="https://lternet.edu/events/categories/community-calls/">https://lternet.edu/events/categories/community-calls/</a>



Distributed engagement across the research lifecycle (from ideation to manuscript writing)	WHONDRS(Borton et al., 2022; Malhotra et al., 2025)
Ensuring wide geographical coverage and broad opportunities	NutNet (Borer et al., 2014) Phenocam (Young et al., 2025) μSudAqua (Fermani et al., 2024) Argo (Wong et al., 2020)
Reducing barriers to public engagement	Community Snow Observations: <a href="https://communitysnowobs.org/about-us/">https://communitysnowobs.org/about-us/</a> CoCoRaHS: <a href="https://www.cocorahs.org/">https://www.cocorahs.org/</a>
Enabling members of the public to engage with scientific data collection via environmental photograph submission	Lagunas de Sierra Nevada: <a href="https://obsnev.es/en/participacion/">https://obsnev.es/en/participacion/</a> Aurorasaurus: <a href="https://www.aurorasaurus.org/">https://www.aurorasaurus.org/</a>
Enabling members of the public to collect data at pre-established sites	CrowdHydrology/CrowdWater: <a href="http://www.crowdhydrology.com">www.crowdhydrology.com</a> <a href="https://crowdwater.ch/en/start/">https://crowdwater.ch/en/start/</a> NPS Dragonfly Mercury Project: <a href="https://www.nps.gov/subjects/citizenscience/dragonfly-mercury-project.htm">https://www.nps.gov/subjects/citizenscience/dragonfly-mercury-project.htm</a> Tracking the Colour of Peatlands: <a href="https://peatcolours.github.io/tracking.html">https://peatcolours.github.io/tracking.html</a>
Journal special issues/ conference themes	WHONDRS (Borton et al., 2022) NEON (Nagy et al., 2021) LTER (Jones and Driscoll, 2022) GEOTRACES (Buck et al., 2024)
Conduct extensive and targeted search for participants	Zooniverse: <a href="https://www.zooniverse.org/">https://www.zooniverse.org/</a> MyGardenOfTrees (Bison et al., 2025)
Operations	
Open communication rooted in psychological safety	NGEE Arctic (Iversen et al., 2020)



Clearly defined roles and responsibilities	Seeding vs Planting Experiment (Leverkus et al., 2021) Global Seismograph Network (Davis, 2024)
Recognizing and crediting roles beyond the research team (i.e., data generators, domain experts, volunteers and community members, data managers, repository curators, and coordinators)	WHONDRS: <a href="https://data.ess-dive.lbl.gov/portals/WHONDRS">https://data.ess-dive.lbl.gov/portals/WHONDRS</a> NGEE Arctic (Iversen et al., 2020)
Established model for shared leadership	Dry Rivers Research Coordination Network (Allen et al., 2025) LTER: <a href="https://lternet.edu/committees/">https://lternet.edu/committees/</a>
Clear expectations and mechanisms for engagement (e.g. authorship)	NutNet (Borer et al., 2023), <a href="https://nutnet.org/authorship">https://nutnet.org/authorship</a> Bug-Net: <a href="https://www.bug-net.org/test/">https://www.bug-net.org/test/</a>
Clear policies for data use and citation	Ameriflux: <a href="https://ameriflux.lbl.gov/data/data-policy/">https://ameriflux.lbl.gov/data/data-policy/</a> ICOS (ICOS, 2024)
Standardized supplies (e.g., described in protocols or via distributed sampling kits)	EXCHANGE (Myers-Pigg et al., 2023) MONet: <a href="https://github.com/EMSL-MONet/MONet-Protocols-">https://github.com/EMSL-MONet/MONet-Protocols-</a> TeaTime4Science: <a href="https://www.teatime4science.org/">https://www.teatime4science.org/</a> Global deadwood decomposition experiment (Seibold et al., 2021)
Detailed written observational protocols	NEON: <a href="https://www.neonscience.org/data-collection/protocols-standardized-methods">https://www.neonscience.org/data-collection/protocols-standardized-methods</a> The PeatPic Project (Davidson et al., 2025) CoCoRaHS: <a href="https://www.cocorahs.org/">https://www.cocorahs.org/</a>
Detailed written experimental protocols	NutNet (Borer et al., 2014) MONet: <a href="https://github.com/EMSL-MONet/MONet-Protocols-">https://github.com/EMSL-MONet/MONet-Protocols-</a> BugNet (Kempel et al., 2025)
Detailed video protocols	WHONDRS: <a href="https://youtu.be/9fHKCeVCOWc?si=nn7YseD7DehrNc1c">https://youtu.be/9fHKCeVCOWc?si=nn7YseD7DehrNc1c</a> CoCoRaHS: <a href="https://www.youtube.com/playlist?list=PL86DC4C330F518387">https://www.youtube.com/playlist?list=PL86DC4C330F518387</a>



Responsive study design linked to emerging questions, conditions, and needs	WHONDRS(Malhotra et al., 2025) FrogID (Callaghan et al., 2023) NGEE Arctic (Hoffman et al., 2013)
Output	
Use of metadata and data standards	Ameriflux: <a href="https://ameriflux.lbl.gov/data/badm/badm-standards/">https://ameriflux.lbl.gov/data/badm/badm-standards/</a> MIREN (Kueffer et al., 2014) TeaComposition (Djukic et al., 2021) GROWdb (Borton et al., 2025)
Common vocabularies, time formats, file types and structures, and standard units	Ameriflux: <a href="https://ameriflux.lbl.gov/data/aboutdata/data-variables/">https://ameriflux.lbl.gov/data/aboutdata/data-variables/</a> ; <a href="https://ameriflux.lbl.gov/data/badm/badm-standards/">https://ameriflux.lbl.gov/data/badm/badm-standards/</a> LTER: <a href="https://bioportal.bioontology.org/ontologies/LTER-CV">https://bioportal.bioontology.org/ontologies/LTER-CV</a>
Format for data and metadata, quality control checks	Ameriflux (Chu et al., 2023) FLUXNET-CH4(Pastorello et al., 2020) NEON: <a href="https://www.neonscience.org/data-samples/data-management/data-quality-program">https://www.neonscience.org/data-samples/data-management/data-quality-program</a>
Centralized relational database; centralized data and metadata hosting and DOI management	NutNet (Lind, 2016) Ameriflux (Chu et al., 2023) NEON: <a href="https://www.neonscience.org/data">https://www.neonscience.org/data</a> SeedingVsPlanting: <a href="https://seedvsplant.wixsite.com/seedvsplant">https://seedvsplant.wixsite.com/seedvsplant</a>
Ability to view or request usage metrics	Ameriflux: <a href="https://ameriflux.lbl.gov/sites/siteinfo/US-PFa#data-use-log">https://ameriflux.lbl.gov/sites/siteinfo/US-PFa#data-use-log</a> NGEE Arctic: <a href="https://data.ess-dive.lbl.gov/portals/NGEE-Arctic">https://data.ess-dive.lbl.gov/portals/NGEE-Arctic</a>
Using a defined metadata and data structure and repeatable workflows to manage standardization	NutNet(Lind, 2016) MONet: <a href="https://sc.emsl.pnnl.gov/">https://sc.emsl.pnnl.gov/</a> GEOTRACES (Aguilar-Islas et al., 2024)
Publish data prior to manuscript development such that all data are open access to anyone within and outside of the network	WHONDRS: <a href="https://data.ess-dive.lbl.gov/portals/WHONDRS">https://data.ess-dive.lbl.gov/portals/WHONDRS</a> CROCUS: <a href="https://data.ess-dive.lbl.gov/portals/crocus/Data">https://data.ess-dive.lbl.gov/portals/crocus/Data</a> EXCHANGE (Pennington et al., 2023)



270 Across all of these categories, trust is crucial and is strengthened when participants experience mutual benefit, when they know their contributions will be recognized and that decisions will support both individual and collective goals (Harris et al., 2021; Jagosh et al., 2015). Successful distributed data generation depends on making these considerations explicit from the outset. A single effort need not adopt all of these best practices, and many best practices described here require sustained funding, dedicated personnel, or institutional support. Projects operating with limited resources may need to prioritize a subset of practices based on their specific goals and constraints.

#### 275 **4 AI-enabled opportunities in distributed data generation:**

The best practices for distributed data generation described in Sect. 3 require overcoming logistical and technological barriers, usually due to the large number of project participants. The rise of AI brings opportunities to better incorporate best practices across large teams, and magnify the impact of distributed data generation (Fig. 1). However, concrete examples of AI-enabled distributed efforts remain limited and significant steps (governance, 280 training, standards, etc.) are necessary for AI-readiness of environmental sciences in general (Mishra, 2026). Below we propose three growth areas for distributed data generation facilitated by AI advancements.

##### **4.1 Outreach: LLM-driven enhancements in efficiency and scalability of trans-disciplinary co-design**

The outreach best practice of effective co-design, and integration of disparate data and knowledge, requires communication and collaboration among humans with varying background knowledge and motivations (Chambers 285 et al., 2021). Whether it is a soil scientist communicating with a computational scientist or a process modeler conveying ecosystem management scenarios to a farmer, co-design and knowledge sharing across people and projects is hampered by communication barriers and delays in science-to-action pipelines.

290 Large language models (LLMs) offer one solution to effective use of co-production and science-to-action pipelines that require human collaboration across disparate domains. LLMs could improve collaborations across similar or connected distributed data generation projects by expediting cross-domain translation, similar to LLM use in cross-language translation (Gui et al., 2019). Linking related studies and disciplines across the globe has added benefits of maximizing impact and discovery, while reducing repetition and pressure on collaborators that are asked repeatedly for involvement. Such improvements can be applied outside of distributed data generation efforts, but are 295 particularly valuable in this domain due to the challenges of maintaining consistent terminology and shared understanding across dozens to hundreds of participants who may span multiple disciplines, institutions, and career stages. Given this scale, traditional one-on-one communication approaches become increasingly inefficient and error-prone.

300 LLM-based collaboration tools could be used throughout a project lifecycle to translate plans, findings, and implications across domains and team members. LLMs can also be used to convert highly technical content into



content relevant for nontechnical entities that require the information in a certain format to, for example, convert to policy. Thus, LLMs could play an important role in expediting the process of converting technical findings into actionable pipelines (Nie and Liu, 2025). More broadly, applications of LLMs could help achieve the key tenets of  
305 co-production, such as researching solutions together and empowering voices (Chambers et al., 2021), and informing future research priorities (Ji et al., 2025). In this way, AI-enabled co-design could ensure that fundamental process understanding derived from distributed data can be applied to broad questions and applications.

However, the effectiveness of LLMs for distributed data generation remains largely untested. Current LLMs show  
310 variable accuracy in technical domains, can introduce biases or errors in domain-specific translations, and can struggle with context that requires deep domain expertise (Cong-Lem et al., 2024; K. Rostam and Kertész, 2025). For example, using LLMs to automate data-to-action pipelines in an urban ecology use case suggested that automation could only reduce 10% of the labor cost of writing reports for urban planning, and that this was primarily due to inaccuracies in LLM inference (Richards et al., 2026). Thus, large-scale implementation of LLMs  
315 would require substantial validation, particularly for ensuring technical accuracy across disciplines. Additionally, automated discipline translation tools risk reducing meaningful human engagement if deployed without careful attention to maintaining collaborative relationships that are central to successful distributed efforts (Richards et al., 2026).

#### 4.2 Operations: AI-guided distributed data generation to achieve rapid adaptability and iteration

320 The operational best practice of study designs that are responsive to emerging questions, conditions, and needs could also be enabled by AI approaches. For example, a fundamental challenge is the persistent issue of sampling bias. AI-guided approaches can systematically reduce spatial and participant selection biases by autonomously identifying high-priority sampling locations based on *a priori* user-defined needs (e.g., model uncertainty, environmental distinctiveness) (Gary et al., 2024; Zhu et al., 2024). AI-guided distributed data generation enables rapid, iterative  
325 cycles between data collection, prediction, and model improvements through optimization of sampling locations (Malhotra et al., 2025; McDowell et al. *in review*). In such efforts, AI models can continuously refine sampling strategies to maximize information gain while minimizing redundant effort, combining AI with approaches such as submodular optimization (Gary et al., 2024).

330 The integration of adaptive sampling in near-real time with automated model updating would also create unprecedented opportunities for responsive, hypothesis-driven data generation that can adapt to emerging scientific questions and evolving environmental conditions (e.g., Reichstein et al., 2025). Distributed data generation allows such adaptive approaches to be scalable and nimble across large spatial domains (regional to global). This would be particularly powerful when implemented through hybrid models that combine the pattern-recognition capabilities of  
335 AI with the mechanistic understanding embedded in process-based models, creating systems that can both capture complex environmental heterogeneity and provide interpretable insights into underlying biological, chemical, and/or physical processes (e.g., Liu et al., 2024).



340 Implementing AI-guided adaptive sampling requires substantial computational infrastructure, real-time data  
transmission capabilities, and expertise in both data and domain science. Many distributed efforts, particularly those  
relying on volunteer participation or operating in remote locations, may lack the technical infrastructure for near-  
real-time data integration and model updating. These approaches are most feasible for well-funded initiatives with  
existing cyberinfrastructure, though costs may decrease as technologies mature.

345 AI agents –autonomous software systems that can plan and execute complex functions with minimal human help–  
will also be a key opportunity as they are capable of guiding sampling decisions at both global scales (optimal site  
selection across continents) and local scales (fine-scale spatial and temporal sampling guidance) (U.S. DOE 2026b.  
Such agents could be centralized (e.g., classic high-performance computing or cloud compute) and/or decentralized  
via edge computing. An evolution towards autonomous (or semi-autonomous) scientific workflows could help  
350 democratize participation in distributed data generation efforts while simultaneously improving data quality through  
standardized protocols and real-time quality control. Agents could also relieve some of the participation burden of  
repetitive tasks such as summarizing meetings and action items. Thus, the development of AI-based iterative data  
generation and modeling methods represents a transformative opportunity to address fundamental challenges in  
science enabled by distributed data generation.

### 355 **4.3 Output: AI-facilitated data harmonization within and across distributed efforts**

A single distributed project can require harmonization among data of disparate types and even stemming from  
different disciplines. New methods for expediting data harmonization could greatly enhance the impact of  
distributed projects. Here, too, LLMs can help expedite harmonization across disciplinary boundaries. LLMs could  
aid in an initial harmonization of similar data types, a task that has traditionally been done manually and is labor  
360 intensive and error prone, keeping in mind that automated harmonization requires careful validation by domain  
experts to detect semantic mismatches, unit conversion errors, or inappropriate harmonization that algorithms may  
miss (Hodgson et al., 2026). Open-source LLMs (e.g., Llama-family models) can be particularly valuable for such a  
task, because they can be fine-tuned using project-specific metadata and trained for specific tasks (Sai et al., 2025).  
Agentic workflows could also be developed, wherein agents discover and harmonize relevant external data with data  
365 from a distributed project (Parkinson et al., 2026). As is the gold standard now, future data would also be available  
from centralized data hubs that follow open data standards and are findable, accessible, interoperable, and reusable  
(FAIR; Wilkinson et al., 2016). However, in the future, maximizing the AI-readiness of these data would further  
lower the barrier of entry for adding new data and data types by automating data update and harmonization actions.  
Distributed data hubs will have to be scalable and contain machine-readable documents (Stocker et al., 2025) to  
370 harness the power of AI, both for analysis and coproduction. To realize this transition, the integration of data across  
trans-disciplinary distributed data generation efforts will benefit from new AI methods for quality control and  
harmonization. One promising direction is fine-tuning LLM's for specific (multi)domains because this has been  
demonstrated to improve LLM performance for domain-specific applications (Zhang et al., 2025).



375 Agents could also keep track of data policies and ensure proper credit and recognition, likely in a less biased way  
than humans do (Herz et al., 2020), if provided with unbiased rules. These are particularly important considerations  
for distributed data generation efforts that may involve hundreds to thousands of participants, and AI could provide  
scalable solutions. However, current AI agent capabilities for complex policy management remain limited (Richards  
et al., 2026), and human-in-the-loop oversight remains essential to validate outputs and ensure appropriate credit  
380 attribution and policy compliance (Larosa et al., 2025; Mammides et al., 2026; Reynolds et al., 2025). For  
distributed efforts with hundreds to thousands of participants, this oversight can remain practical through audit logs  
and exception-based escalation, where agents apply standard rules at scale but route uncertain or contested cases to  
human reviewers (Mammides et al., 2026; Plaat et al., 2025).

## 385 **5 Conclusions**

The framework presented here aims to enable expansion in the use of distributed data generation by establishing  
shared vocabulary, defining core modalities, proposing best practices, and identifying opportunities for future  
development. Distributed data modalities arise from integrating across two continua: participants, ranging from  
highly formalized to highly flexible approaches, and methods, ranging from observational to experimental. This  
390 framework captures diverse data generation strategies, and our examples show that many successful efforts integrate  
multiple modalities.

Three pillars underpin successful distributed data generation. Outreach includes co-design and multimodal  
engagement that facilitate broad opportunities. Operations include aligning decisions and designs with safety, trust,  
395 clear expectations. Output includes generating and openly sharing rigorous data and manuscripts, supported by  
standardization and repeatable workflows. These practices help mitigate persistent challenges such as inadequate  
recognition and data fragmentation, which can limit the impact of distributed data efforts.

Emerging AI technologies offer potential opportunities to transform distributed data generation, though  
400 implementation barriers remain. LLMs may expedite co-design by bridging disciplinary communication barriers and  
accelerate science-to-action pipelines, while AI-guided sampling approaches can reduce bias and enable responsive,  
iterative data collection. AI-facilitated data harmonization and management could integrate diverse datasets across  
projects and disciplines, creating a more comprehensive understanding of ecological and environmental systems.  
Operationalizing these AI-enabled opportunities faces implementation barriers. As these technologies mature and  
405 become more accessible, careful attention to validation, human oversight, and broadening access will be essential.

We encourage distributed data generators to use the shared vocabulary proposed here, adopt the best practices we  
outline, and build collaborative bridges between existing efforts with the help of AI. By implementing these  
approaches and fostering broader communication about the value of distributed data generation, the community can  
410 collectively advance both scientific understanding and societal applications. The convergence of established best



practices with AI-enabled opportunities positions distributed data generation as a central tool to address some of the most pressing social, ecological, and environmental challenges through scalable, mutually beneficial, and scientifically rigorous approaches.

#### 415 **Code and data availability**

No code or datasets were generated during the current study.

#### **Author contribution**

This statement uses CrediT taxonomy. Andino, Bolton, Borer, Borton, Davidson, Downs, Goldman, Hensley, Leverkusen, Lowry, Malhotra, McDowell, Muradyan, Richardson, Scheibe, Stegen and Ward contributed to  
420 conceptualization of the manuscript. Scheibe and Stegen contributed to funding acquisition. Goldman, Malhotra, Scheibe, and Stegen contributed to methodology. Forbes, Goldman, and Malhotra contributed to project administration. Goldman, Malhotra, Muradyan, and Richardson contributed to visualization. Andino, Borer, Borton, Davidson, Goldman, Hensley, Leverkusen, Lowry, Malhotra, McDowell, Muradyan, Richardson, Stegen and Ward contributed to writing the original draft. All co-authors contributed to writing review and editing. All co-authors  
425 contributed critically to the drafts and gave final approval for publication.

#### **Competing interests**

The authors declare that they have no conflict of interest.

#### **Acknowledgements**

The co-authors thank the other participants in the 2025 workshop that precipitated this manuscript: John Bargar,  
430 Danielle Christianson, Sara Knox, Richard Marinos, and Greg Maurer. NGM, TDS, AEG, BF, JCS, and AM were supported by the Earth & Biological Sciences Program Development Office at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the US Department of Energy. ADR acknowledges support from the US National Science Foundation (DEB 2142144 and 2224545) and the US Department of Energy (DE-SC0024218). ABL acknowledges grants SMART-OAK, funded by the European Union  
435 through FORWARDS (Horizon Europe Project No. 101084481) grants to third parties managed by European Forest Institute; and RESISTRES (C-EXP267-UGR23) granted by University of Granada and funded by Consejería de Universidad, Investigación e Innovación and by ERDF Andalusia Program 2021-2027. ETB acknowledges support of US National Science Foundation DEB-2311608 and 2425352. RTH was supported by the National Ecological Observatory Network operated by Battelle for the U.S. National Science Foundation (DBI 2217817). MRD  
440 acknowledges support of U.S. National Science Foundation 2419138. JMA acknowledges support from the US Department of Energy (DE-SC0023520). KFP was supported through through the COMPASS-FME project (<https://compass.pnnl.gov/>) and the Molecular Observation Network (MONet) at the Environmental Molecular Sciences Laboratory (<https://ror.org/04rc0xn13>), a US Department of Energy Office of Science User Facility



sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830.

445 SJD's work is supported by the Government of Québec through the Plan pour une économie verte (PEV). AMP acknowledges support from the US Department of Energy through the River Corridor Science Focus Area and the COMPASS-FME projects. NDW acknowledges support from the US Department of Energy through the COMPASS-FME project. PM's work is supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research's Urban Integrated Field Laboratories CROCUS project research activity  
450 (DE-AC02- 06CH11357). WRB's work is supported by the US Department of Energy, Office of Biological and Environmental Research's Next Generation Ecosystem Experiment in the Arctic project. Claude 4.5 Sonnet (Anthropic) provided editorial suggestions during manuscript development; all scientific content and final decisions remain the responsibility of the authors.

## References

- 455 Aguilar-Islas, A., University of Alaska Fairbanks, Planquette, H., Lohan, M., Geibert, W., and Cutter, G.: Intercalibration: A cornerstone of the success of the GEOTRACES program, *Oceanography* (Wash. D. C.), 37, <https://doi.org/10.5670/oceanog.2024.404>, 2024.
- Allen, D. C., Burgin, A. J., Seybold, E. C., Dodds, W. K., Busch, M. H., Bergstrom, A., Krabbenhoft, C. A., Boersma, K. S., Stegen, J. C., Olden, J. D., Atkinson, C. L., Jones, C. N., Datry, T., Godsey, S. E., Shogren, A. J., Walters, A. W., Plont, S., Walker, R. H., Shanafield, M., Mims, M. C., Price, A. N., Smith, C. R., You, Y., Bogan, M. T., Burrows, R. M., Messenger, M. L., Stubbington, R., and Zimmer, M. A.: Shared leadership can promote success in collaborative research networks in ecology, *Funct. Ecol.*, <https://doi.org/10.1111/1365-2435.70109>, 2025.
- 460 Anderson, R., Mawji, E., Cutter, G., Measures, C., and Jeandel, C.: GEOTRACES: Changing the way we explore ocean chemistry, *Oceanography* (Wash. D. C.), 27, 50–61, 2014.
- 465 Atkins, J. W., Aho, K. S., Chen, X., Elmore, A. J., Fiorella, R., Luo, W., Lombardozzi, D., Lunch, C., Manak, L., de Pablo, L. X., Myers-Pigg, A. N., Record, S., Qiu, T., Reed, S., Ruddell, B., Strange, B., Torrens, C. L., Yule, K., and Richardson, A. D.: Recommendations for developing, documenting, and distributing data products derived from NEON data, *Ecosphere*, 16, <https://doi.org/10.1002/ecs2.70159>, 2025.
- 470 Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.*, 82(11), 2415-2434, [https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2), 2001
- 475 Bison, M., Ponta, N., Daniella Maria Schweizer, and Csilléry, K.: Harnessing foresters' engagement for climate change adaptation: the emerging tool of next-generation citizen science, *bioRxiv*, <https://doi.org/10.1101/2025.09.15.676198>, 2025.
- Borer, E. T. and Seabloom, E. W.: Maximizing inference from distributed experimental networks via “add-on” studies, *Front. Ecol. Environ.*, 23, <https://doi.org/10.1002/fee.70007>, 2025.
- 480 Borer, E. T., Harpole, W. S., Adler, P. B., Lind, E. M., Orrock, J. L., Seabloom, E. W., and Smith, M. D.: Finding generality in ecology: a model for globally distributed experiments, *Methods in Ecology and Evolution*, 5, 65–73, 2014.
- Borer, E. T., MacDougall, A. S., Stevens, C. J., Sullivan, L. L., Wilfahrt, P. A., and Seabloom, E. W.: Writing a massively multi-authored paper: Overcoming barriers to meaningful authorship for all, *Methods Ecol. Evol.*,



- 485 <https://doi.org/10.1111/2041-210x.14096>, 2023.
- Borton, M. A., Collins, S. M., Graham, E. B., Garayburu-Caruso, V. A., Goldman, A. E., de Melo, M., Renteria, L., Stegen, J. C., and WHONDERS Crowdsourced Consortium: It takes a village: Using a crowdsourced approach to investigate organic matter composition in global rivers through the lens of ecological theory, *Front. Water*, 4, <https://doi.org/10.3389/frwa.2022.870453>, 2022.
- 490 Borton, M. A., McGivern, B. B., Willi, K. R., Woodcroft, B. J., Mosier, A. C., Singleton, D. M., Bambakidis, T., Pelly, A., Daly, R. A., Liu, F., Freiburger, A., Edirisinghe, J. N., Faria, J. P., Danczak, R., Leleiwi, I., Goldman, A. E., Wilkins, M. J., Hall, E. K., Pennacchio, C., Roux, S., Eloë-Fadrosh, E. A., Good, S. P., Sullivan, M. B., Wood-Charlson, E. M., Miller, C. S., Ross, M. R. V., Henry, C. S., Crump, B. C., Stegen, J. C., and Wrighton, K. C.: A functional microbiome catalogue crowdsourced from North American rivers, *Nature*, 637, 103–112, 2025.
- 495 Bowman, M. M., Heath, A. E., Varga, T., Battu, A. K., Chu, R. K., Toyoda, J., Cheeke, T. E., Porter, S. S., Moffett, K. B., LeTendre, B., Qafoku, O., Bargar, J. R., Mans, D. M., Hess, N. J., and Graham, E. B.: One thousand soils for molecular understanding of belowground carbon cycling, *Front. Soil Sci.*, 3, <https://doi.org/10.3389/fsoil.2023.1120425>, 2023.
- 500 Buck, C., Skidaway Institute of Oceanography, Fietz, S., Hamilton, D., Ho, T.-Y., Perron, M., and Shelley, R.: GEOTRACES: Fifteen years of progress in marine aerosol research, *Oceanography (Wash. D. C.)*, 37, <https://doi.org/10.5670/oceanog.2024.409>, 2024.
- Caldararu, S., Rolo, V., Stocker, B. D., Gimeno, T. E., and Nair, R.: Ideas and perspectives: Beyond model evaluation – combining experiments and models to advance terrestrial ecosystem science, *Biogeosciences*, 20, 3637–3649, 2023.
- 505 Callaghan, C. T., Thompson, M., Woods, A., Poore, A. G. B., Bowler, D. E., Samonte, F., Rowley, J. J. L., Roslan, N., Kingsford, R. T., Cornwell, W. K., and Major, R. E.: Experimental evidence that behavioral nudges in citizen science projects can improve biodiversity data, *Bioscience*, 73, 302–313, 2023.
- 510 Chambers, J. M., Wyborn, C., Ryan, M. E., Reid, R. S., Riechers, M., Serban, A., Bennett, N. J., Cvitanovic, C., Fernández-Giménez, M. E., Galvin, K. A., Goldstein, B. E., Klenk, N. L., Tengö, M., Brennan, R., Cockburn, J. J., Hill, R., Munera, C., Nel, J. L., Österblom, H., Bednarek, A. T., Bennett, E. M., Brandeis, A., Charli-Joseph, L., Chatterton, P., Curran, K., Dumrongrojwathana, P., Durán, A. P., Fada, S. J., Gerber, J.-D., Green, J. M. H., Guerrero, A. M., Haller, T., Horcea-Milcu, A.-I., Leimona, B., Montana, J., Rondeau, R., Spierenburg, M., Steyaert, P., Zaehring, J. G., Gruby, R., Hutton, J., and Pickering, T.: Six modes of co-production for sustainability, *Nat. Sustain.*, 4, 983–996, 2021.
- 515 Chu, H., Christianson, D. S., Cheah, Y.-W., Pastorello, G., O'Brien, F., Geden, J., Ngo, S.-T., Hollowgrass, R., Leibowitz, K., Beekwilder, N. F., Sandesh, M., Dengel, S., Chan, S. W., Santos, A., Delwiche, K., Yi, K., Buechner, C., Baldocchi, D., Papale, D., Keenan, T. F., Biraud, S. C., Agarwal, D. A., and Torn, M. S.: AmeriFlux BASE data pipeline to support network growth and data sharing, *Sci Data*, 10, 614, 2023.
- 520 Collis, S., Muradyan, P., O'Brien, J., Anderson, G., Wagner, T.J., Blind-Doskocil, L., Sullivan, R., Tuftedal, M., Pal, S., Zvolanek, E. and Jackson, R.: The CROCUS Measurement Strategy (No. ICUC12-438). Copernicus Meetings, <https://doi.org/10.5194/icuc12-438>, 2025.
- Cong-Lem, N., Soyoo, A., and Tsering, D.: A systematic review of the limitations and associated opportunities of ChatGPT, *Int. J. Hum. Comput. Interact.*, 1–16, 2024.
- 525 Davidson, S. J.: Tracking the Colour of Peatlands, Citizen science project, available at: <https://peatcolours.github.io/tracking.html>, last access: 6/16/2026, 2021–ongoing.
- Davidson, S. J., Malhotra, A., Jassey, V. E. J., Strack, M., Aitova, E., Anderson, R., Atkinson, L. J., Barel, J. M., Bird, M., Brehier, C., Donaldson-Selby, G., Duley, E., Eklof, J., de Eyto, E., Granath, G., Grant, A., Hartmann, A., Holland, A., Huth, V., Jones, C. P., Lee, S.-C., Lopatin, J., Milner, A. M., Peacock, M., Peichl, M., Perez-Quezada,



- 530 J. F., Perryman, C. R., Pickard, A., Rautakoski, H., Silvester, E., Virkkala, A.-M., and Wegener, E.: The PeatPic project: predicting plot-scale green leaf phenology across peatlands, *Environ. Res. Lett.*, 20, 114002, 2025.
- Davis, P.: Development and operation of a global-scale seismographic network: The IRIS/USGS GSN, *Perspectives of Earth and Spac*, 5, <https://doi.org/10.1029/2023cn000225>, 2024.
- 535 Delwiche, K.B., Nelson, J., Kowalska, N., Moore, C.E., Shirkey, G., Tarin, T., Cleverly, J.R. and Keenan, T.F.: Charting the Future of the FLUXNET Network. *Bull. Am. Meteorol. Soc.*, 105(3), E466-E473, <https://doi.org/10.1175/BAMS-D-23-0316.1>, 2024.
- 540 Djukic, I., Kepfer-Rojas, S., Kappel Schmidt, I., Steenberg Larsen, K., Beier, C., Berg, B., Verheyen, K., Trevathan-Tackett, S. M., Macreadie, P. I., Bierbaumer, M., Patoine, G., Eisenhauer, N., Guerra, C. A., Maestre, F. T., Hagedorn, F., Oggioni, A., Bergami, C., Magagna, B., Kwon, T., Shibata, H., and TeaComposition initiative: The TeaComposition initiative: Unleashing the power of international collaboration to understand litter decomposition, <https://doi.org/10.25674/SO93ISS1PP73>, 2021.
- 545 Dwivedi, D., Santos, A. L. D., Barnard, M. A., Crimmins, T. M., Malhotra, A., Rod, K. A., Aho, K. S., Bell, S. M., Bomfim, B., Brearley, F. Q., Cadillo-Quiroz, H., Chen, J., Gough, C. M., Graham, E. B., Hakkenberg, C. R., Haygood, L., Koren, G., Lilleskov, E. A., Meredith, L. K., Naeher, S., Nickerson, Z. L., Pourret, O., Song, H.-S., Stahl, M., Taş, N., Vargas, R., and Weintraub-Leff, S.: Biogeosciences Perspectives on Integrated, Coordinated, Open, Networked (ICON) Science, *Earth Space Sci*, 9, e2021EA002119, 2022.
- Eberts, S. M., Woodside, M. D., Landers, M. N., and Wagner, C. R.: Monitoring the pulse of our Nation’s rivers and streams—The U.S. Geological Survey streamgaging network, <https://doi.org/10.3133/fs20183081>, 2019.
- Elizabeth L. Wenner Maggie Geist: The national estuarine research reserves program to monitor and preserve estuarine waters, *Coast. Manage.*, 29, 1–17, 2001.
- 550 Elmendorf, S. C., Jones, K. D., Cook, B. I., Diez, J. M., Enquist, C. A. F., Hufft, R. A., Jones, M. O., Mazer, S. J., Miller-Rushing, A. J., Moore, D. J. P., Schwartz, M. D., and Weltzin, J. F.: The plant phenology monitoring design for The National Ecological Observatory Network, *Ecosphere*, 7, e01303, 2016.
- 555 Fermani, P., Martyniuk, N., Saraceno, M., Gereá, M., Sabio y García, C., Schiaffino, M. R., Sánchez, M. L., Allen Dohle, S., Alonso, C., Barrera, M. A., Bastidas Navarro, M., Bernal, M. C., Burgueño Sandoval, G. M., Cetra, N., De Melo, M. L., García, P. E., Gómez, B. M., Gómez Lugo, S., Gómez-Ramírez, E., Griffiero, L., Hünicken, L. A., Lagomarsino, L., Lozada, M., Mansilla Ferro, C., Modenutti, B., Mateus-Barros, E., Ojeda, D. A., Padulles, M. L., Pereira, E., Porcel, S., Quiroga, M. V., Saad, J. F., Salas, M. C., Santucho, J., Sarmento, H., Segade, M. E., Soto Cárdenas, C., Torremorell, A., Unrein, F., Zabala, M. S., Zanetti, J., and Graziano, M.: A Latin American network of microbial observatories for monitoring aquatic ecosystems, *Ecol. Austral*, 636–646, 2024.
- 560 Fraser, L. H., AL Henry, H., Carlyle, C. N., White, S. R., Beierkuhnlein, C., Cahill, J. F., Casper, B. B., Cleland, E., Collins, S. L., Dukes, J. S., Knapp, A. K., Lind, E., Long, R., Luo, Y., Reich, P. B., Smith, M. D., Sternberg, M., and Turkington, R.: Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology and environmental science, *Frontiers in Ecology and the Environment*, 11, 147–155, 2013.
- 565 Gary, S. F., Scheibe, T. D., Rexer, E., Torreira, A. V., Garayburu-Caruso, V. A., Goldman, A., and Stegen, J. C.: Prediction of distributed river sediment respiration rates using community-generated data and machine learning, *Journal of Geophysical Research: Machine Learning and Computation*, 1, <https://doi.org/10.1029/2024jh000199>, 2024.
- Gui, Q., Liu, C., and Du, D.: Globalization of science and international scientific collaboration: A network perspective, *Geoforum*, 105, 1–12, 2019.
- 570 Haider, S., Lembrechts, J. J., McDougall, K., Pauchard, A., Alexander, J. M., Barros, A., Cavieres, L. A., Rashid, I., Rew, L. J., Aleksanyan, A., Arévalo, J. R., Aschero, V., Chisholm, C., Clark, V. R., Clavel, J., Daehler, C., Dar, P. A., Dietz, H., Dimarco, R. D., Edwards, P., Essl, F., Fuentes-Lillo, E., Guisan, A., Gwate, O., Hargreaves, A. L.,



- Jakobs, G., Jiménez, A., Kardol, P., Kueffer, C., Larson, C., Lenoir, J., Lenzner, B., Padrón Mederos, M. A., Mihoc, M., Milbau, A., Morgan, J. W., Müllerová, J., Naylor, B. J., Nijs, I., Nuñez, M. A., Otto, R., Preuk, N., Ratier Backes, A., Reshi, Z. A., Rumpf, S. B., Sandoya, V., Schroder, M., Speziale, K. L., Urbach, D., Valencia, G., Vandvik, V., Vitková, M., Vorstenbosch, T., Walker, T. W. N., Walsh, N., Wright, G., Zong, S., and Seipel, T.: Think globally, measure locally: The MIREN standardized protocol for monitoring plant species distributions along elevation gradients, *Ecol Evol*, 12, e8590, 2022.
- 575
- Hanson, P. J. and Walker, A. P.: Advancing global change biology through experimental manipulations: Where have we been and where might we go?, *Glob Chang Biol*, 26, 287–299, 2020.
- 580
- Harris, L. A., Garza, C., Hatch, M., Parrish, J., Posselt, J., Alvarez Rosario, J. P., Davidson, E., Eckert, G., Wilson Grimes, K., Garcia, J. E., Haacker, R., Horner-Devine, M. C., Johnson, A., Lemus, J., Prakash, A., Thompson, L., Vitousek, P., Martin Bras, M. P., and Reyes, K.: Equitable Exchange: A framework for diversity and inclusion in the geosciences, *AGU Adv.*, 2, <https://doi.org/10.1029/2020av000359>, 2021.
- 585
- Heiskanen, J., Brümmer, C., Buchmann, N., Calfapietra, C., Chen, H., Gielen, B., Gkritzalis, T., Hammer, S., Hartman, S., Herbst, M., Janssens, I. A., Jordan, A., Juurola, E., Karstens, U., Kasurinen, V., Kruijt, B., Lankreijer, H., Levin, I., Linderson, M.-L., Loustau, D., Merbold, L., Myhre, C. L., Papale, D., Pavelka, M., Pilegaard, K., Ramonet, M., Rebmann, C., Rinne, J., Rivier, L., Saltikoff, E., Sanders, R., Steinbacher, M., Steinhoff, T., Watson, A., Vermeulen, A. T., Vesala, T., Vitková, G., and Kutsch, W.: The Integrated Carbon Observation System in Europe, *Bull. Am. Meteorol. Soc.*, 103, E855–E872, 2022.
- 590
- Herz, N., Dan, O., Censor, N., and Bar-Haim, Y.: Opinion: Authors overestimate their contribution to scientific work, demonstrating a strong bias, *Proc Natl Acad Sci U S A*, 117, 6282–6285, 2020.
- Hodgson, R. Y., Robinson, S. A., Boutin, A. C., Chan, F. K., Bennett, J. R., Buxton, R. T., Caufield, J. H., Hanna, D. E. L., and Alamenciak, T.: Assessing the effectiveness of ontology-grounded AI term extraction using OntoGPT for environmental evidence synthesis, *Environ. Evid.*, 15, 1, 2026.
- 595
- Hoffman, F. M., Kumar, J., Mills, R. T., and Hargrove, W. W.: Representativeness-based sampling network design for the State of Alaska, *Landsc. Ecol.*, 28, 1567–1586, 2013.
- Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., and Qiao, H.: Sampling biases shape our view of the natural world, *Ecography (Cop.)*, 44, 1259–1269, 2021.
- 600
- ICOS: ICOS Handbook 2024, <https://doi.org/10.18160/28AV-80QR>, 2024.
- Iversen, C., Bolton, W., Rogers, A., Wilson, C., and Wullschleger, S.: Building a culture of safety and trust in team science, *Eos (Washington DC)*, 101, <https://doi.org/10.1029/2020eo143064>, 2020.
- Jagosh, J., Bush, P. L., Salsberg, J., Macaulay, A. C., Greenhalgh, T., Wong, G., Cargo, M., Green, L. W., Herbert, C. P., and Pluye, P.: A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects, *BMC Public Health*, 15, 725, 2015.
- 605
- Ji, X., Wu, X., Deng, R., Yang, Y., Wang, A. and Zhu, Y.: Utilizing large language models for identifying future research opportunities in environmental science, *Journal of Environmental Management*, 373, 123667, 2025.
- Jones, J. A. and Driscoll, C. T.: Long-term ecological research on ecosystem responses to climate change, *Bioscience*, 72, 814–826, 2022.
- 610
- K. Rostam, Z. R. and Kertész, G.: Advances in pre-trained language models for domain-specific text classification: A systematic review, *ACM Trans. Intell. Syst. Technol.*, 16, 1–41, 2025.
- Keller, M., Schimel, D. S., Hargrove, W. W., and Hoffman, F. M.: A continental strategy for the National Ecological Observatory Network, *Front. Ecol. Environ.*, 6, 282–284, 2008.
- Kempel, A., Adamidis, G. C., Anadón, J. D., Atkinson, J., Auge, H., Avtzis, D., Bachelot, B., Bashirzadeh, M.,



- 615 Bota, J. L., Classen, A., Constantinou, I., Crawley, M., de Bellis, T., Dostal, P., Ebeling, A., Eisenhauer, N.,  
Eldridge, D. J., Encina, G., Estrada, C., Everingham, S., Fanin, N., Feng, Y., Gaspar, M., Gooriah, L., Graff, P.,  
Montalván, E. G., Montalván, P. G., Hartke, T. R., Huang, L., Jochum, M., Kaljund, K., Karmiris, I., Koorem, K.,  
Korell, L., Laine, A.-L., le Provost, G., Lessard, J.-P., Liu, M., Liu, X., Liu, Y., Llançabure, J., Loïez, S., Loydi, A.,  
Marrero, H., Gockel, S., Montoya, A., Münzbergová, Z., Niu, Y., Ott, D., Oyarzabal, M., Panitsa, M.,  
620 Papatheodorou, E., Piper, F. I., Püssa, K., Rand, K., Saiz, H., Sanders, N. J., Schädler, M., Scherber, C., Semchenko,  
M., Sepp, S.-K., Shah, M. A., Shaheen, I., Stein, C., Stewart, J., Tang, Z., Tschan, G., van Nouhuys, S.,  
Vandegehuchte, M. L., Vernon, M., R, S., V., Wang, J., Xiao, Y., Xystrakis, F., Yang, J., Yang, S., Zografou, K.,  
and Allan, E.: The Bug-Network (BugNet): A Global Experimental Network Testing the Effects of Invertebrate  
625 15, e72111, 2025.
- Knapp, A. K., Smith, M. D., Hobbie, S. E., Collins, S. L., Fahey, T. J., Hansen, G. J. A., Landis, D. A., La Pierre, K.  
J., Melillo, J. M., Seastedt, T. R., Shaver, G. R., and Webster, J. R.: Past, present, and future roles of long-term  
experiments in the LTER network, *Bioscience*, 62, 377–389, 2012.
- 630 Kueffer, C., Daehler, C., Dietz, H., McDougall, K., Parks, C., Pauchard, A., and Rew, L.: The mountain invasion  
research network (MIREN). Linking local and global scales for addressing an ecological consequence of global  
change, *Gaia*, 23, 263–265, 2014.
- Larosa, F., Hoyas, S., Alberto Conejero, H., Garcia-Martinez, J., Nerini, F. F., and Vinuesa, R.: Large language  
models in climate and sustainability policy: limits and opportunities, arXiv [cs.CY], arXiv,  
<https://doi.org/10.48550/arXiv.2502.02191>, 2025.
- 635 Leverkus, A. B., Levy, L., Andivia, E., Annighöfer, P., De Cuyper, B., Ivetic, V., Lazdina, D., Löff, M., and Villar-  
Salvador, P.: Restoring oak forests through direct seeding or planting: Protocol for a continental-scale experiment,  
*PLoS One*, 16, e0259552, 2021.
- Lind, E. M.: Unified data management for distributed experiments: A model for collaborative grassroots scientific  
networks, *Ecol. Inform.*, 36, 231–236, 2016.
- 640 Lindenmayer, D. B., Likens, G. E., and Franklin, J. F.: Earth Observation Networks (EONs): Finding the Right  
Balance, *Trends Ecol Evol*, 33, 1–3, 2018.
- Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., Zhu, Q., Till, J., Jia, X., Jiang, C., Wang, S., Qin, Z., Kong,  
H., Grant, R., Mezbahuddin, S., Kumar, V., and Jin, Z.: Knowledge-guided machine learning can improve carbon  
cycle quantification in agroecosystems, *Nat Commun*, 15, 357, 2024.
- 645 Maestre, F. T. and Eisenhauer, N.: Recommendations for establishing global collaborative networks in soil ecology,  
*Soil Org*, 91, 73–85, 2019.
- Malhotra, A., Forbes, B., Gary, S. F., Goldman, A. E., Waterman, B. R., Garayburu-Caruso, V., Fluet-Chouinard,  
E., Mehan, S., Bruen, M., Taylor, M., Ardon, M., Cardenas, M. B., Dodds, W. K., Lonborg, C., McDowell, W. H.,  
Moustapha, M., Myers-Pigg, A. N., Regier, P., Rubin, T., Song, H., Stewart, R. D., Villa, J., Ward, N. D., Scheibe,  
650 T. D., and Stegen, J. C.: Harnessing artificial intelligence to automate environmental predictions, *bioRxiv*,  
<https://doi.org/10.1101/2025.11.06.684583>, 2025.
- Malhotra, A., von Fromm, S. F., Bond-Lamberty, B., Doetterl, S., Georgiou, K., Graham, E. B., Heckman, K. A.,  
Jia, R., Patel, K. F., Rod, K. A., Santos, F., Terrer, C., Todd-Brown, K., Zheng, J., Hofmockel, K., and Bailey, V.:  
Do we have globally representative data to understand soil processes?, *Biogeochemistry*,  
655 <https://doi.org/10.1007/s10533-025-01301-z>, 2026.
- Mammides, C., Gu, H., Nimalrathna, T. S., Avicena, N. R., Papadopoulos, H., and Campos-Arceiz, A.: Emerging  
applications of large language models in ecology and conservation science, *Conserv. Biol.*, e70287, 2026.
- McDowell, N.G., Li, L., Malhotra, A., Myers-Pigg, A., Bond-Lamberty, B., Goldman, A.E., Vernon, C.R., Hetland,



- 660 R., Bailey, V., Chen, X., Rice, J., Scheibe, T., and Stegen, J.: Multi-scale experimental designs for transferable ecosystem research, *Ecosphere*, in review.
- Meier, C. L., Thibault, K. M., and Barnett, D. T.: Spatial and temporal sampling strategy connecting NEON Terrestrial Observation System protocols, *Ecosphere*, 14, <https://doi.org/10.1002/ecs2.4455>, 2023.
- Mishra, G.: Ecology is not yet ready for AI-and why that matters, *Proc. Natl. Acad. Sci. U. S. A.*, 123, e2533976123, 2026.
- 665 Molla, A., Flores, A. B., and Sailor, D. J.: Addressing data gaps and disparities in access to air quality networks: a case study of Maricopa County, Arizona, *Population and Environment*, 47, 1–21, 2025.
- Muraina, T. O. and Jimoh, S. O.: Global climate change and the hidden fate of African grasslands, *Bioscience*, 72, 805–805, 2022.
- 670 Myers-Pigg, A. N., Pennington, S. C., Homolka, K. K., Lewis, A. M., Otenburg, O., Patel, K. F., Regier, P., Bowe, M., Boyanov, M. I., Conroy, N. A., Day, D. J., Norris, C. G., O’Loughlin, E. J., Roebuck, J. A., Jr, Stetten, L., Bailey, V. L., Kemner, K. M., Ward, N. D., and EXCHANGE Consortium: Biogeochemistry of upland to wetland soils, sediments, and surface waters across Mid-Atlantic and Great Lakes coastal interfaces, *Sci. Data*, 10, 822, 2023.
- 675 Nagy, R. C., Balch, J. K., Bissell, E. K., Cattau, M. E., Glenn, N. F., Halpern, B. S., Ilangakoon, N., Johnson, B., Joseph, M. B., Marconi, S., O’Riordan, C., Sanovia, J., Swetnam, T. L., Travis, W. R., Wasser, L. A., Woolner, E., Zarnetske, P., Abdulrahim, M., Adler, J., Barnes, G., Bartowitz, K. J., Blake, R. E., Bombaci, S. P., Brun, J., Buchanan, J. D., Chadwick, K. D., Chapman, M. S., Chong, S. S., Chung, Y. A., Corman, J. R., Couret, J., Crispo, E., Doak, T. G., Donnelly, A., Duffy, K. A., Dunning, K. H., Duran, S. M., Edmonds, J. W., Fairbanks, D. E., Felton, A. J., Florian, C. R., Gann, D., Gebhardt, M., Gill, N. S., Gram, W. K., Guo, J. S., Harvey, B. J., Hayes, K.
- 680 R., Helmus, M. R., Hensley, R. T., Hondula, K. L., Huang, T., Hundertmark, W. J., Iglesias, V., Jacinthe, P.-A., Jansen, L. S., Jarzyna, M. A., Johnson, T. M., Jones, K. D., Jones, M. A., Just, M. G., Kaddoura, Y. O., Kagawa-Vivani, A. K., Kaushik, A., Keller, A. B., King, K. B. S., Kitzes, J., Koontz, M. J., Kouba, P. V., Kwan, W.-Y., LaMontagne, J. M., LaRue, E. A., Li, D., Li, B., Lin, Y., Liptzin, D., Long, W. A., Mahood, A. L., Malloy, S. S., Malone, S. L., McGlinchy, J. M., Meier, C. L., Melbourne, B. A., Mietkiewicz, N., Morisette, J. T., Moustapha, M., Muscarella, C., Musinsky, J., Muthukrishnan, R., Naithani, K., Neely, M., Norman, K., Parker, S. M., Perez Rocha, M., Petri, L., Ramey, C. A., Record, S., Rossi, M. W., SanClements, M., et al.: Harnessing the NEON data
- 685 revolution to advance open environmental science with a diverse and data-capable community, *Ecosphere*, 12, <https://doi.org/10.1002/ecs2.3833>, 2021.
- 690 National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Board on Science Education, and Committee on Designing Citizen Science to Support Science Learning: *Learning Through Citizen Science: Enhancing Opportunities by Design*, edited by: Dibner, K. A. and Pandya, R., National Academies Press (US), Washington (DC), 2018.
- Nie, Q. and Liu, T.: Large language models: Tools for new environmental decision-making, *J Environ Manage*, 375, 124373, 2025.
- 695 Norström, A. V., Cvitanovic, C., Löf, M. F., West, S., Wyborn, C., Balvanera, P., Bednarek, A. T., Bennett, E. M., Biggs, R., de Bremond, A., Campbell, B. M., Canadell, J. G., Carpenter, S. R., Folke, C., Fulton, E. A., Gaffney, O., Gelcich, S., Jouffray, J.-B., Leach, M., Le Tissier, M., Martín-López, B., Louder, E., Loutre, M.-F., Meadow, A. M., Nagendra, H., Payne, D., Peterson, G. D., Reyers, B., Scholes, R., Speranza, C. I., Spierenburg, M., Stafford-Smith, M., Tengö, M., van der Hel, S., van Putten, I., and Österblom, H.: Principles for knowledge co-production in
- 700 sustainability research, *Nat. Sustain.*, 3, 182–190, 2020.
- Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., and Torn, M. S.: The AmeriFlux network: A coalition of the willing, *Agric. For. Meteorol.*, 249, 444–456, 2018.
- Parker, M. and Kingori, P.: Good and bad research collaborations: Researchers’ views on science and ethics in



global health research, *PLoS One*, 11, e0163579, 2016.

- 705 Parkinson, R. H., Cerbone, H., Mieskolainen, M., Cao, S., Wilson, A. D., Albacete, S., Armstrong, E. B., Bass, C., Botías, C., Brown, A., Hayward, A. J., Herbertsson, L., Jones, A. K., Nagloo, N., Nicholls, E., Rigosi, E., Sgolastra, F., Siviter, H., Stanley, D. A., Straub, L., Straw, E. A., Tadei, R., Walter, K., Stevance, H. F., Daniels, R. K., Lambert, B., and Roberts, S.: MetaBeeAI: An AI pipeline for structured evidence extraction from biological literature, *Ecol. Inform.*, 96, 103813, 2026.
- 710 Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D.,  
715 Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. de Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. di, Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A.,  
720 Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, *Sci. Data*, 7, 225, 2020.
- Pennington, S. C., Alford, S., Back, M. P., Bailey, V., Baldwin, A., Bolinger, J., Bowe, M., Boyanov, M. I., Cianci-Gaskill, J. A., Conroy, N. A., Cooper, M. J., Day, D., Demeo, A., Derby, K., Detweiler, D., Devres-Zimmerman, S.,  
725 Eberhard, E., Gedan, K., Haaf, L., Homolka, K. K., Johnson, E., Kemner, K. M., Khan, A., Kirwan, M., Kittaka, P., Koontz, E., Langley, A., Leff, R., Lerberg, S., Lewis, A. M., Malkin, S., Marcarelli, A. M., McMurray, S. E., Messerschmidt, T., Michael, T. C., Michael, H. A., Minor, E. C., Moye, B., Mozdzer, T. J., Neubauer, S., Norris, C. G., O'Loughlin, E. J., Otenburg, O., Pain, A., Patel, K. F., Philben, M., Phillips, E., Pratt, D., Regier, P., Roebuck, J. A., Jr, Sage, L., Sandborn, D., Smith, S., Smith, A., Soin-Voshell, S., Song, B., Sprague-Gettsy, A., St. Laurent, K.,  
730 Staver, L., Stearns, A., Stetten, L., Swerida, R., Theuerkauf, E. J., Tully, K., Vargas, R., Ward, N. D., Watson, E., Weilminster, C., and Myers-Pigg, A. N.: EXCHANGE Campaign 1: A community-driven baseline characterization of soils, sediments, and water across coastal Gradients, <https://doi.org/10.15485/1960313>, 2023.
- Peters, D. P. C., Loescher, H. W., SanClements, M. D., and Havstad, K. M.: Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology, *Ecosphere*, 5, 1–23, 2014a.
- 735 Peters, D. P. C., Loescher, H. W., SanClements, M. D., and Havstad, K. M.: Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology, *Ecosphere*, 5, 1–23, 2014b.
- Plaat, A., van Duijn, M., van Stein, N., Preuss, M., van der Putten, P., and Batenburg, K. J.: Agentic Large Language Models, a survey, *arXiv [cs.AI]*, arXiv, <https://doi.org/10.48550/arXiv.2503.23037>, 2025.
- 740 Poyatos, R., Werner, C., and Martínez-Vilalta, J.: Global Transpiration Data from Sap Flow Measurements: the SAPFLUXNET Database, 2021.
- Raposa, K. B., Woolfolk, A., Endris, C. A., Fountain, M. C., Moore, G., Tyrrell, M., Swerida, R., Lerberg, S., Puckett, B. J., Ferner, M. C., Hollister, J., Burdick, D. M., Champlin, L., Krause, J. R., Haines, D., Gray, A. B., Watson, E. B., and Wasson, K.: Evaluating thin-layer sediment placement as a tool for enhancing tidal marsh resilience: A coordinated experiment across eight US national estuarine research reserves, *Estuaries Coast.*,  
745 <https://doi.org/10.1007/s12237-022-01161-y>, 2023.
- Reichstein, M., Benson, V., Blunk, J., Camps-Valls, G., Creutzig, F., Fearnley, C. J., Han, B., Kornhuber, K., Rahaman, N., Schölkopf, B., Tárraga, J. M., Vinueza, R., Dall, K., Denzler, J., Frank, D., Martini, G., Nganga, N., Maddix, D. C., and Weldemariam, K.: Early warning of complex climate risk with integrated artificial intelligence, *Nat. Commun.*, 16, 2564, 2025.



- 750 Reynolds, S. A., Beery, S., Burgess, N., Burgman, M., Butchart, S. H. M., Cooke, S. J., Coomes, D., Danielsen, F., Di Minin, E., Durán, A. P., Gassert, F., Hinsley, A., Jaffer, S., Jones, J. P. G., Li, B. V., Mac Aodha, O., Madhavapeddy, A., O'Donnell, S. A. L., Oxbury, W. M., Peck, L., Pettorelli, N., Rodríguez, J. P., Shuckburgh, E., Strassburg, B., Yamashita, H., Miao, Z., and Sutherland, W. J.: The potential for AI to revolutionize conservation: a horizon scan, *Trends Ecol. Evol.*, 40, 191–207, 2025.
- 755 Richards, D., Worden, D., and Lavorel, S.: Generative artificial intelligence use in automated urban ecological assessments requires substantial human oversight, *Landsc. Urban Plan.*, 270, 105615, 2026.
- Richter, D. D., Billings, S. A., Groffman, P. M., Kelly, E. F., Lohse, K. A., McDowell, W. H., White, T. S., Anderson, S., Baldocchi, D. D., Banwart, S., Brantley, S., Braun, J. J., Brecheisen, Z. S., Cook, C. W., Hartnett, H. E., Hobbie, S. E., Gaillardet, J., Jobbagy, E., Jungkunst, H. F., Kazanski, C. E., Krishnaswamy, J., Markewitz, D., O'Neill, K., Riebe, C. S., Schroeder, P., Siebe, C., Silver, W. L., Thompson, A., Verhoef, A., and Zhang, G.: Ideas and perspectives: Strengthening the biogeosciences in environmental research networks, *Biogeosciences*, 15, 4815–4832, 2018.
- 760 Romero, G. Q., Gonçalves-Souza, T., Roslin, T., Marquis, R. J., Marino, N. A. C., Novotny, V., Cornelissen, T., Orivel, J., Sui, S., Aires, G., Antoniazzi, R., Dáttilo, W., Breviglieri, C. P. B., Busse, A., Gibb, H., Izzo, T. J., Kadlec, T., Kemp, V., Kersch-Becker, M., Knapp, M., Kratina, P., Luke, R., Majnarić, S., Maritz, R., Mateus Martins, P., Mendesil, E., Michalko, J., Mrazova, A., Novais, S., Pereira, C. C., Perić, M. S., Petermann, J. S., Ribeiro, S. P., Sam, K., Trzcinski, M. K., Vieira, C., Westwood, N., Bernaschini, M. L., Carvajal, V., González, E., Jausoro, M., Kaensin, S., Ospina, F., Cristóbal-Pérez, E. J., Quesada, M., Rogy, P., Srivastava, D. S., Szpryngiel, S., Tack, A. J. M., Teder, T., Videla, M., Viljur, M.-L., and Koricheva, J.: Climate variability and aridity modulate the role of leaf shelters for arthropods: A global experiment, *Glob Chang Biol*, 28, 3694–3710, 2022.
- 765 Sai, C. V., Elmitwally, N., Rice, I., Mahmoud, H., Vickers, I., and Schmoor, X.: Fine-tuning large language models for domain-specific tasks: A comparative study on supervised and parameter-efficient fine-tuning techniques for cybersecurity and IT support, in: *Lecture Notes on Data Engineering and Communications Technologies*, Springer Nature Switzerland, Cham, 150–160, 2025.
- 775 SanClements, M. D., Record, S., Rose, K. C., Donnelly, A., Chong, S. S., Duffy, K., Hallmark, A., Heffernan, J. B., Liu, J., Mitchell, J. J., Moore, D. J. P., Naithani, K., O'Reilly, C. M., Sokol, E. R., Whitney, K. S., Weintraub-Leff, S. R., and Yang, D.: People, infrastructure, and data: A pathway to an inclusive and diverse ecological network of networks, *Ecosphere*, 13, e4262, 2022.
- 780 Seibold, S., Rammer, W., Hothorn, T., Seidl, R., Ulyshen, M. D., Lorz, J., Cadotte, M. W., Lindenmayer, D. B., Adhikari, Y. P., Aragón, R., Bae, S., Baldrian, P., Barimani Varandi, H., Barlow, J., Bäessler, C., Beauchêne, J., Berenguer, E., Bergamin, R. S., Birkemoe, T., Boros, G., Brandl, R., Brustel, H., Burton, P. J., Cakpo-Tossou, Y. T., Castro, J., Cateau, E., Cobb, T. P., Farwig, N., Fernández, R. D., Firn, J., Gan, K. S., González, G., Gossner, M. M., Habel, J. C., Hébert, C., Heibl, C., Heikkala, O., Hemp, A., Hemp, C., Hjältén, J., Hotes, S., Kouki, J., Lachat, T., Liu, J., Liu, Y., Luo, Y.-H., Macandog, D. M., Martina, P. E., Mukul, S. A., Nachin, B., Nisbet, K., O'Halloran, J., Oxbrough, A., Pandey, J. N., Pavlíček, T., Pawson, S. M., Rakotondranary, J. S., Ramanamanjato, J.-B., Rossi, L., Schmidl, J., Schulze, M., Seaton, S., Stone, M. J., Stork, N. E., Suran, B., Sverdrup-Thygeson, A., Thorn, S., Thyagarajan, G., Wardlaw, T. J., Weisser, W. W., Yoon, S., Zhang, N., and Müller, J.: The contribution of insects to global forest deadwood decomposition, *Nature*, 597, 77–81, 2021.
- 785 Skaldina, O. and Blande, J. D.: Global biases in ecology and conservation research: Insight from pollinator studies, *Ecol. Lett.*, 28, e70050, 2025.
- 790 Stegen, J. C. and Goldman, A. E.: WHONDRS: A community resource for studying dynamic river corridors, *mSystems*, 3, <https://doi.org/10.1128/msystems.00151-18>, 2018.
- 795 Stocker, M., Snyder, L., Anfuso, M., Ludwig, O., Thießen, F., Farfar, K. E., Haris, M., Oelen, A., and Jaradeh, M. Y.: Rethinking the production and publication of machine-readable expressions of research findings, *Sci Data*, 12, 677, 2025.



Tiegs, S. D., Capps, K. A., Costello, D. M., Schmidt, J. P., Patrick, C. J., Follstad Shah, J. J., LeRoy, C. J., CELLDEX Consortium<sup>‡</sup>, and CELLDEX Consortium: Human activities shape global patterns of decomposition rates in rivers, *Science*, 384, 1191–1195, 2024.

800 U.S. DOE. Lessons Learned from Ecosystem-Scale Experimental Field Studies, DOE/SC-2024. U.S. Department of Energy Office of Science, <https://doi.org/10.2172/3021280>, 2026a.

U.S. DOE. Envisioning Frontiers in AI and Computing for Biological Research Workshop Report, DOE/SC-2022. U.S. Department of Energy Office of Science, <https://doi.org/10.2172/2566158>, 2026b.

805 Wallace, L.: Intercomparison of PurpleAir Sensor Performance over Three Years Indoors and Outdoors at a Home: Bias, Precision, and Limit of Detection Using an Improved Algorithm for Calculating PM, *Sensors (Basel)*, 22, <https://doi.org/10.3390/s22072755>, 2022.

810 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data*, 3, 160018, 2016.

815 Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., Yashayaev, I., Barker, P. M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, M. O., Schmid, C., Lyman, J. M., McTaggart, K. E., Purkey, S. G., Zilberman, N., Alkire, M. B., Swift, D., Owens, W. B., Jayne, S. R., Hersh, C., Robbins, P., West-Mack, D., Bahr, F., Yoshida, S., Sutton, P. J. H., Cancouët, R., Coatanoan, C., Dobbler, D., Juan, A. G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., 820 Claustre, H., D'Ortenzio, F., Le Reste, S., Le Traon, P.-Y., Rannou, J.-P., Saout-Grit, C., Speich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I. M., Klein, B., Notarstefano, G., Poulain, P.-M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaka, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida, T., Cowley, R., Lovell, J. L., Oke, P. R., van Wijk, E. M., Carse, F., Donnelly, M., Gould, W. J., Gowers, K., King, B. A., Loch, S. G., Mowat, M., Turton, J., Rama Rao, E. P., Ravichandran, M., Freeland, H. J., Gaboury, I., Gilbert, D., Greenan, B. J. W., Ouellet, M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K., Jo, H., et al.: 825 Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats, *Front. Mar. Sci.*, 7, <https://doi.org/10.3389/fmars.2020.00700>, 2020.

830 Young, A. M., Milliman, T., Hufkens, K., Ballou, K., Coffey, C., Begay, K., Fell, M., Javadian, M., Post, A. K., Schädel, C., Vladich, Z., Zimmerman, O., Browning, D. M., Florian, C. R., Moon, M., SanClements, M. D., Seyednasrollah, B., Friedl, M. A., and Richardson, A. D.: Tracking vegetation phenology across diverse biomes using Version 3.0 of the PhenoCam Dataset, , <https://doi.org/10.5194/essd-2025-120>, 2025.

Zhang, Y., Lin, S., Xiong, Y., Li, N., Zhong, L., Ding, L., and Hu, Q.: Fine-tuning large language models for interdisciplinary environmental challenges, *Environ. Sci. Ecotechnol.*, 27, 100608, 2025.

835 Zhu, Q., Yuan, K., Li, F., Riley, W. J., Hoyt, A., Jackson, R., McNicol, G., Chen, M., Knox, S. H., Briner, O., Beerling, D., Gedney, N., Hopcroft, P. O., Ito, A., Jain, A. K., Jensen, K., Kleinen, T., Li, T., Liu, X., McDonald, K. C., Melton, J. R., Miller, P. A., Müller, J., Peng, C., Poulter, B., Qin, Z., Peng, S., Tian, H., Xu, X., Yao, Y., Xi, Y., Zhang, Z., Zhang, W., Zhu, Q., and Zhuang, Q.: Critical needs to close monitoring gaps in pan-tropical wetland CH<sub>4</sub> emissions, *Environ. Res. Lett.*, 19, 114046, 2024.