Open science, reproducibility, and transparency in ecology

Article *in* Ecological Applications · January 2019

DOI: 10.1002/eap.1822

2 authors:

Stephen M. Powers
Washington State University

Stephanie E. Hampton
Washington State University

Ecological Applications, 29(1), 2019, e01822

Open science, reproducibility, and transparency in ecology

Stephen M. Powers [1] and Stephanie E. Hampton

School of the Environment, Washington State University, Pullman, Washington 99164 USA

Citation: S. M. Powers, and S. E. Hampton. 2019. Open science, reproducibility, and transparency in ecology. Ecological Applications 29(1):e01822. 10.1002/eap.1822

Abstract. Reproducibility is a key tenet of the scientific process that dictates the reliability and generality of results and methods. The complexities of ecological observations and data present novel challenges in satisfying needs for reproducibility and also transparency. Ecological systems are dynamic and heterogeneous, interacting with numerous factors that sculpt natural history and that investigators cannot completely control. Observations may be highly dependent on spatial and temporal context, making them very difficult to reproduce, but computational reproducibility can still be achieved. Computational reproducibility often refers to the ability to produce equivalent analytical outcomes from the same data set using the same code and software as the original study. When coded workflows are shared, authors and editors provide transparency for readers and allow other researchers to build directly and efficiently on primary work. These qualities may be especially important in ecological applications that have important or controversial implications for science, management, and policy. Expectations for computational reproducibility and transparency are shifting rapidly in the sciences. In this work, we highlight many of the unique challenges for ecology along with practical guidelines for reproducibility and transparency, as ecologists continue to participate in the stewardship of critical environmental information and ensure that research methods demonstrate integrity.

Key words: collaborative tools; data policy; data science; ecoinformatics; ecosystem; environmental science; open science; repeatability; replicability; reproducible; transparent; workflows.

Ecologists have long faced a novel challenge not routinely encountered in less field-oriented sciences: repeated testing is fundamental to the scientific method (Popper 1934) yet it is impossible to perfectly repeat observational studies of the natural world (Vanderbilt and Blankman 2017). This issue is timely as scientists across disciplines increasingly recognize the challenges of reproducing published results, and the threats that irreproducible results pose to the scientific process (Munafo et al. 2017). Reproducibility and transparency issues are particularly important for scientists engaged in actionable science and ecological applications; this work often feeds back rapidly and directly on biota, ecosystems, and people who have stakes in conservation or management outcomes, in turn affecting perceptions about the integrity of our field.

For ecologists engaged in observational and field-based studies, data are often highly dependent on the spatial and temporal context of the specific system (Huang 2014, Schnitzer and Carson 2016, LaDeau et al. 2017, Peters and Okin 2017). The weather does not recycle itself. Gradual changes over time, regime shifts, and legacies of past events can greatly influence how natural systems work, as well as our perceptions about how they work (Magnuson 1990). With this complexity, ecology has relied on deep understanding of natural history as a source of ideas about pattern and process (Anderson 2017), often through long periods of intensive observation that could be argued as irreproducible. Even if we could bring back our predecessors, many of their study systems now bear little resemblance to earlier states.

Such irregularities in the natural world can torment experimental ecologists. Repeated or replicated sequences in time are elusive and, likewise, neighboring populations, communities, or ecosystems observed within the same day or year can still differ in important ways that affect the outcomes of studies. Thus, while broad guidelines for reproducible research in the sciences are available (Sandve et al. 2013), ecologists face novel challenges that complicate adoption of such general practices. These challenges are linked to the heterogeneity of the systems we study, as well as the approaches and information we use. Strategies that increase the reproducibility of ecological studies are being pursued (Milcu et al. 2018) and it is important that ecologists address these issues if we are to continue serving a critical role in understanding the complex dynamics of the biosphere in the Anthropocene.

Computational Reproducibility

While field-oriented observations may not ever be completely or perfectly repeated, replicated, or reproduced (Plesser 2018), ecologists still can achieve

Manuscript received 4 May 2018; revised 21 July 2018; accepted 30 July 2018. Corresponding Editor: David Schimel.
[1] E-mail: steve.powers@wsu.edu

computational reproducibility. It is imperative that we prepare young researchers for the computational expectations of the future by engaging them in the process now, and fostering the relationships and career tracks that reinforce such efforts.

Computational reproducibility frequently refers to the ability to generate equivalent analytical outcomes from the same data set using the same code and software (Peng 2011, Stodden et al. 2013, Leek and Peng 2015, Fidler et al. 2017). One practice is for authors to share both code and data that generated the specific results shown in their peer-reviewed publications (Poisot et al. 2016). Such aids for the peer review process may be essential as analyses become more complex (Ellison 2010), allowing reviewers and other readers to follow the decision-making processes of authors in careful reviews of analytical workflows. In this publication model, authors should be confident at the time of manuscript submission that a reviewer could sit down with the code and data, and generate the same results that are in the paper; this process requires not just the raw code and data, but also good documentation of both (see *Practical Guidance for Openness* below). When data are shared without code, an independent investigator can attempt to regenerate results nearly from scratch, but typically with considerable costs in time and effort. When both code and data are published, a reader can directly test alternative analysis techniques and improve the existing one by modifying the author's code. Thus, sharing code ensures that less time is wasted, for reviewers, editors, and authors alike, as we adjust to shifting expectations for transparency and reproducibility. About linking publications more directly with underlying analyses Donoho (2010), went so far to write: "An article about computational results is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result."

As in other scientific fields, repeated or replicated studies in ecology and environmental science are relatively uncommon (Schnitzer and Carson 2016), as are published tests involving computational reproducibility. With professional reward systems focused on novelty throughout the sciences, there have been few incentives for scientists to repeat each other's work, leading to unsettling questions about the reliability of results and conclusions. However, we now live in era where problems can be caught and prevented by embracing the goal of reproducibility in our work habits, peer review systems, and overall culture (Reichman et al. 2011, Michener and Jones 2012, Hampton et al. 2013, Borregaard and Hart 2016). Key aspects of computational reproducibility can be achieved using open source tools for tasks such as data integration, data wrangling, statistics and modeling, handling of multiple users, visualization, workflow pipelines (e.g., Kepler, noWorkflow for Python), creation of process metadata (e.g., RDataTracker), and version control. Such tools are now widely accessible (Jones et al. 2006, Hampton et al. 2015, Borregaard and Hart 2016, Ellison 2017, Rocchini et al. 2017) and attitudes and norms increasingly favor data reuse (Curty et al. 2017).

Transparency

Simultaneously, computational reproducibility can aid in transparency (Lowndes et al. 2017). In ecological applications such as natural resource management and policy issues, high transparency may be essential for scientists and scientific institutions to maintain public trust (Reichman et al. 2011, Michener and Jones 2012, Michener 2015, Schimel and Keller 2015). With scripted analyses and detailed metadata accompanying data sets, methods and decisions are explicitly documented and can be shared at any point in the project. Scripted analyses require researchers to document their workflows from the very beginning, making the products easier to share later on, sometimes even prior to completion if necessary. This upfront work can greatly minimize post-project workloads aimed at publicizing and sharing the data products (Gil et al. 2016) or addressing inquiries about contentious results. Computational reproducibility, with data and code available to readers, obviates the need for stakeholders to challenge researchers and managers for access to these resources, for example, through legal mechanisms such as the Freedom of Information Act (Huettmann 2005). Such legal mechanisms are important tools for increasing transparency of the democratic process, but the logistics can be burdensome to all parties. Meanwhile, if we can foster the relationships, skills, and activities that coincide with publishing of data and code, we might also suppress some of the anxieties (e.g., fear of a "data disaster" or analysis error) that relate to mental health of young researchers (Evans et al. 2018, Woolston 2018).

Yet again, ecology has particulars in how sensitive topics, such as endangered species, are handled when facilitating transparency. A researcher who encounters a small population of rare orchids might effectively condemn a species to extinction by revealing its location. Similarly, a broad-scale environmental study may identify specific locations with otherwise unforeseen hazards that compromise property values, or conversely suggest economic opportunities that attract international attention.

Publishing location specifics may exaggerate the consequences of the "digital divide" across diverse cultures that have variable access to computing tools (Ess and Sudweeks 2001), disadvantaging local people if they lack capacity to access or interpret such data. Difficult ethical issues are raised by publishing scientific information that is not equally accessible, particularly in light of international declarations on human rights to the "benefits of science," an aspect of the international human rights framework that remains unfulfilled (Duke et al. 2011).

Fortunately, ecologists have developed standard practices for protecting organisms and environments that may be applied in sensitive situations. For example, it is customary to obfuscate the location of endangered species or the location of intake pipes for municipal drinking water supplies when publishing data. While these practices are common for such issues of endangered species and human safety, other situations involving potentially sensitive information may have fewer guidelines. Complicating matters, researchers and policy makers have recently debated a new idea: that only studies based on open data can be used in decision making. The outcome of this debate could have widespread impact, including situations where useful scientific information could be excluded from decision-making processes if the work does not meet the evolving standards of openness. We encourage authors to be mindful of the changes in data policies, transparency, and reproducibility that are upon us, and to consult with experienced data managers and editors when they anticipate potential for ethical dilemmas arising from data publication.

Rapid Scientific Progress

Many calls for open data and open software have been less focused on reproducibility and transparency, and more focused on the opportunities for rapid scientific advancements associated with open science (Nielsen 2011). When data and methods are easily accessible for repurposing, researchers can build directly on previous research (Carpenter et al. 2009), avoid blind alleys, and not "reinvent the wheel" (Hampton et al. 2015). Further, open data and open methods enable the repurposing of existing data toward new questions not originally envisioned at the time data were collected (Carpenter et al. 2009). Additional benefits include increased opportunities for teamwork and rapid formation of research teams (Milojevic 2014), as well as enhanced citation rates (Piwowar and Vision 2013, McKiernan et al. 2016). Further, because empirical relationships of the past may not necessarily hold in the future (Gustafson 2013), vigilance is required to continually confront data sets with the best available information as well as models; these endeavors to update the state of the science on a topic can be greatly accelerated if the data and methods exist in well-documented, accessible, reusable, and machine-readable form.

Practical Guidance for Openness

Recognition and adoption of open research practices has rapidly risen, and has been associated with policies that increase public access to data (Stodden et al. 2013, Heimstadt et al. 2014, Michener 2015, McKiernan et al. 2016, Culina et al. 2018) and code (Stodden et al. 2013). Such policies are motivated by a combination of ethical, moral, or utilitarian arguments (Willinksy 2006, Duke and Porter 2013, Soranno et al. 2015). Ecologists can expect more journals, including ESA journals, to encourage if not mandate the archiving of code in their data policies (Simmons 2016, Collins and Verdier 2017, Schimel 2018), further pushing the community toward open science approaches.

What does "open science" mean? Key features include transparency of process and sharing of data products, code, and metadata. Completely open science includes public communication throughout a project from idea generation to post-publication discussion (Hampton et al. 2015). In an open setting, enhanced collaboration and rapid idea vetting can occur through what Hackett et al. (2008) call "peer review on the fly." A researcher practicing completely open science might generate discussion about new ideas through social media, invite online collaboration from anyone at any stage, execute work in an open online notebook with evolving code and data sets (e.g., JuPyter notebooks; Szitenberg et al. 2015), post pre-prints to elicit feedback, publish in journals and repositories that have open-access or open review formats (e.g., F1000Research; Hollister and Stachelek 2017), and engage in post-publication discussion through online forums. Hampton et al. (2015) provide more detailed examples of open science workflows and options.

Open science and computational reproducibility go hand in hand if well documented versions of code and data used to generate the results of a study are made public. Code is frequently shared through web-based services that provide version control and clearly document evolution of code, such as GitHub (Hampton et al. 2015). Data repositories familiar to ecologists and other environmental scientists include DataONE member nodes such as Environmental Data Initiative and Knowledge Network for Biocomplexity (KNB). These repositories host thousands of data sets related to patterns and processes, including those from the US Long-term Ecological Research (LTER; Michener 2016), providing diverse and frequently updated examples that demonstrate emerging best practices.

In addition, some investigators' institutions such as universities, agencies and non-profit organizations offer services for housing public data (Borgman et al. 2015), and at the least formal level investigators can publish data and code as Supplements. A caution in using institutional repositories or Supplements is that investigators should consider (1) the standardization of data and metadata into formats that an individual organization may not be prepared to provide but that an international repository will guide, frequently with specifics useful in one's own field (e.g., the KNB was developed in ecology), and (2) the discoverability of the data and code that international repositories can facilitate. With the rise of this variety of options for individual investigators, providing data sets on personal websites or "e-mail me for the data" are no longer acceptable practices (Stodden et al. 2018), given the low expectations for longevity or consistency of these solutions.

Producing high-quality, standardized, and machine-readable metadata is an extremely important aspect of data sharing (Jones et al. 2006). It is necessary for reproducibility and transparency, and ultimately it is the backbone for executing data integrations across multiple data sets. In ecology, the Ecological Metadata Language

(EML) is commonly used (e.g., DataONE member nodes) as it can accommodate a broad variety of data types, and there are free tools are available for documenting data in EML. However, with the heterogeneity of data involved in ecological research (Jones et al. 2006), from satellite imagery to genomics, EML is not the only useful standard. More important than which metadata standard an investigator chooses is to simply decide on an existing standard and implement it. Rather than pushing researchers toward a single metadata language, the ecoinformatics field in the past decade has moved primarily toward developing crosswalks for metadata standards that will aid integration and allow multiple popular metadata standards to be maintained (Michener and Jones 2012).

Engaging in completely open science may be unrealistic for many researchers, particularly for those unfamiliar with the full suite of tools or those working in environments with strictly regulated workflows, and we urge investigators to simply learn and adopt elements that are helpful or important to them. Hampton et al. (2015) detail a spectrum of such situations, from the researcher who hasn't yet learned the tools for open science but can be a knowledgeable advocate, to the seasoned open scientist doing work in a completely public forum. Many ecologists currently work somewhere in between. The rise of open science has created broad availability of useful tools that support modern needs for computational reproducibility and transparency.

We're Running Out Of Excuses

Three years have passed since *Ecological Applications* initiated a data policy that mandated data publication. In this policy (Box 1), all data associated with

Box 1. Data policy of the journal Ecological Applications, accessed May 2018.

https://esajournals.onlinelibrary.wiley.com/hub/journal/19395582/resources/data-policy-eap

As a condition for publication of a manuscript in Ecological Applications, all data associated with the results must be made available in a permanent, publicly accessible, data archive or repository.

Authors are strongly encouraged to deposit the data underlying their manuscripts in the Dryad data repository or Figshare, which both provide flexible platforms for a wide variety of digital data. Other permanent repositories include GenBank for DNA sequences, ORNL-DAAC for biogeochemical data, Knowledge Network for Biocomplexity and the LTER Data Portal, as well as institutional repositories such as that at the University of Illinois.

Archived data should be sufficiently complete so that subsequent users can repeat tables, graphs, and statistical analyses reported in the original publication, and derive summary statistics for new or meta- analyses. Thus, the normal resolution of the data that are archived will be at the level of individual observations.

Publication in Ecological Applications constitutes publication of the data, which are then citable, and the desire of authors to control additional research with these data shall not generally be grounds for withholding published data. Sensitive information including but not limited to precise locality data for rare, threatened, or endangered species, or identity of human subjects, should be redacted as required.

Sufficient metadata should accompany the data file so that others can readily use files and interpret variables, including their units. Such metadata can usually be provided in a short text file. Data must be registered and available at the time of publication, although in specific cases, data registration and metadata availability at the time of acceptance, with a firm subsequent date for release of primary data may be acceptable.

By depositing data prior to publication of a manuscript, a permanent link can be made to and from the published paper.

Wiley Online Library can be used for this purpose, but only if the material is submitted with the original submission for peer review. Data must be deposited in other depositories following acceptance and prior to publication.

Advantages of depositing data in a permanent repository include:

- Visibility: Making your data available online (and linking it to the publication) provides a new pathway for others to learn about your work.
- Citability: All data you deposit will receive a persistent, resolvable identifier that can be used in a citation as well as listed on your CV.
- Workload reduction: If you receive individual requests for data, you can simply direct them to files in the archive.
- Preservation: Your data files will be permanently and safely archived in perpetuity.
- Impact: You will garner citations through the reuse of your data.

Authors will be responsible for any fees charged by external data repositories in order to comply with the data archiving requirement.

Box 2. NASA's Earth Science Data and Information Policy, accessed May 2018.

https://earthdata.nasa.gov/earth-science-data-systems-program/policies/data-information-policy

NASA's Earth Science program was established to use the advanced technology of NASA to understand and protect our home planet by using our view from space to study the Earth system and improve prediction of Earth system change. To meet this challenge, NASA promotes the full and open sharing of all data with the research and applications communities, private industry, academia, and the general public. The greater the availability of the data, the more quickly and effectively the user communities can utilize the information to address basic Earth science questions and provide the basis for developing innovative practical applications to benefit the general public.

A common set of carefully crafted data exchange and access principles was created by the Japanese, European, and U.S. International Earth Observing System (IEOS) partners during the 1990s and the early years of the 21st century. From these principles, NASA has adopted the following data policy (in this context, the term "data" includes observation data, metadata, products, information, algorithms, including scientific source code, documentation, models, images, and research results):

- NASA will plan and follow data acquisition policies that ensure the collection of long-term data sets needed to satisfy the research requirements of NASA's Earth science program.
- NASA commits to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available.
- There will be no period of exclusive access to NASA Earth science data. Following a post-launch checkout period, all data will be made available to the user community. Any variation in access will result solely from user capability, equipment, and connectivity.
- NASA will make available all NASA-generated standard products along with the source code for algorithm software, coefficients, and ancillary data used to generate these products.
- All NASA Earth science missions, projects, and grants and cooperative agreements shall include data management plans to facilitate the implementation of these data principles.
- NASA will enforce a principle of non-discriminatory data access so that all users will be treated equally. For data products supplied from an international partner or another agency, NASA will restrict access only to the extent required by the appropriate Memorandum of Understanding (MOU).
- In keeping with the Office of Management and Budget (OMB) Circular A-130, NASA will charge for distribution of data no more than the cost of dissemination. In cases where such dissemination cost would unduly inhibit use, the distribution charge will generally be below that cost.
- Through MOUs and agreements with appropriate interagency partners, NASA will ensure that all data required for Earth system science research are archived. Data archives will include easily accessible information about the data holdings, including quality assessments, supporting relevant information, and guidance for locating and obtaining data.
- NASA will engage in ongoing partnerships with other Federal agencies to increase the effectiveness and reduce the cost of the NASA Earth science program. This interagency cooperation shall include: sharing of data from satellites and other sources, mutual validation and calibration data, and consolidation of duplicative capabilities and functions.
- NASA will, in compliance with applicable Federal law and policy, negotiate and implement arrangements with its international partners, with an emphasis on meeting the data acquisition, distribution, and archival needs of the United States.
- NASA will collect a variety of metrics intended to measure or assess the efficacy of its data systems and services, and assess user satisfaction. Consistent with applicable laws, NASA will make those data available for review.

The data collected by NASA represent a significant public investment in research. NASA holds these data in a public trust to promote comprehensive, long-term Earth science research. Consequently, NASA developed policy consistent with existing international policies to maximize access to data and to keep user costs as low as possible. These policies apply to all data archived, maintained, distributed or produced by NASA data systems.

manuscripts must be made available in a permanent, publicly accessible, data archive or repository. Large projects also often necessitate their own data policies, such as the NASA Earth Science Data and Information Policy (Box 2), and the data access policy of the North Temperate Lakes Long-term Ecological Research (LTER) project (Box 3). Similarly, it has been more than seven years since the U.S. National Science Foundation mandated data management plans in submitted grant proposals. These changes were initially met with skepticism by many in the scientific community, a reaction that was understandable because efficient pathways for accomplishing these tasks did not yet permeate our culture and training. But a lot has changed in a short period of time.

A growing proportion of ecologists have risen to the challenges presented by the open science movement. There has been a proliferation of tools, infrastructure, and people to support this transition (Michener and Jones 2012, Hampton et al. 2015, Soranno et al. 2015, Ellison 2017, Rocchini et al. 2017). Existing tools and approaches can be grouped into five key areas (Hampton et al. 2017): data management and processing, analysis, software skills, visualization, and communication methods for collaboration and dissemination. These resources, and the people who use them, are becoming central to the culture of ecology. If we incorporate the lessons and practices of reproducibility and openness into undergraduate curricula and training of young researchers now, we can increase our capacity to address the current and future challenges. Evidence is mounting that the move toward open science and data sharing has positive career outcomes (Piwowar et al. 2007), and indeed the most influential scientists in some fields also appear to be the most prominent data sharers (Dai et al. 2018). Ecology has now moved more fully into the "big science" era. High-resolution and broad-scale ecological data are emerging from macrosystems research, in situ and remote sensing, large-scale environmental observatories such as the National Ecological Observatory Network, the Ocean Observatory Initiative, and the National Critical Zone Observatory, as well as international grass-roots collaborations such as the Global Lake Ecological Observatory Network and Nutrient

Box 3. Data access policy of the North Temperate Lakes Long-term Ecological Research (LTER) project, member of the US LTER network, accessed May 2018.

https://lter.limnology.wisc.edu/about/ntl-lter-data-access-policy

The North Temperate Lakes (NTL) LTER Data Access Policy is designed to make data from the NTL-LTER database as freely available as possible for academic, research, education, and other professional purposes and aligns with the Creative Commons license CC-BY 4.0 (see https://creativecommons.org/licenses/by/4.0/). Each data set published on the web page is accompanied by documentation (metadata). We encourage the use of our data sets but ask that users read and agree to our Data Use Agreement.

Data use agreement: Permission to download NTL-LTER data sets is granted to the Data User subject to the following terms:

- The Data User must realize that these data sets are being actively used by others for ongoing research and that coordination may be necessary to prevent duplicate publication. The Data User is urged to contact Emily Stanley, lead Principal Investigator, (ehstanley@wisc.edu) to check on other uses of the data. Where appropriate, the Data User may be encouraged to consider collaboration and/or co-authorship with original investigators.
- The Data User must realize that the data may be misinterpreted if taken out of context. We request that you provide Emily Stanley, ATTN: Data Access, Center for Limnology, University of Wisconsin-Madison, 680 North Park Street, Madison, Wisconsin 53706 USA with a copy of any manuscript using the data so that she may review and provide comments on the presentation of our data.
- The Data User must acknowledge use of the data by an appropriate citation (see Citation) of the NTL-LTER database.

By using these data, the Data User agrees to abide by the terms of this agreement. Thank you for your cooperation.

Citation: Use of the data in publications should acknowledge the North Temperate Lakes LTER project. A generic citation for our databases is: <name of data set>, North Temperate Lakes Long Term Ecological Research program (http://lter.limnology.wisc.edu), NSF, <contact person for data set>, Center for Limnology, University of Wisconsin-Madison. The data set name and contact person for each data set can be found in the metadata header of the online data sets.

Data Availability: Our goal is to release all long term data associated with core research areas within 2 years of collection. These data and accompanying metadata will be available for download from the NTL-LTER web site.

Disclaimer: While substantial efforts are made to ensure the accuracy of data and documentation, complete accuracy of data sets cannot be guaranteed. All data are made available "as is." The North Temperate Lakes LTER shall not be liable for damages resulting from any use or misinterpretation of data sets. Data users should be aware that we periodically update data sets.

Network. These developments allow us to confront our ideas with unprecedented power and produce valuable information and tools that foster critical inquiry and liberate new knowledge to the benefit of society amid global change.

Acknowledgments

This manuscript benefited from lively discussion in workshops supported by NSF awards EF-1358900 and ACI-1216894, and constructive comments from Aaron Galloway, Corinna Gries, Jordan Read, and Kara Woo.

Literature Cited

Anderson, J. G. T. 2017. Why ecology needs natural history. American Scientist 105:290.

Borgman, C. L., P. T. Darch, A. E. Sands, I. V. Pasquetto, M. S. Golshan, J. C. Wallis, and S. Traweek. 2015. Knowledge infrastructures in science: data, diversity, and digital libraries. International Journal on Digital Libraries 16:207–227.

Borregaard, M. K., and E. M. Hart. 2016. Towards a more reproducible ecology. Ecography 39:349–353.

Carpenter, S. R., et al. 2009. Accelerate synthesis in ecology and environmental sciences. BioScience 59:699–701.

Culina, A., M. Baglioni, T. W. Crowther, M. E. Visser, S. Woutersen-Windhouwer, and P. Manghi. 2018. Navigating the unfolding open data landscape in ecology and evolution. Nature Ecology & Evolution 2:420–426.

Collins, S. L., and J. M. Verdier. 2017. The coming era of open data. BioScience 67:191–192. https:/doi.org./10.1093/biosci/bix023

Curty, R. G., K. Crowston, A. Specht, B. W. Grant, and E. D. Dalton. 2017. Attitudes and norms affecting scientists' data reuse. PLoS ONE 12:e0189288.

Dai, S.-Q., H. Li, J. Xiong, J. Ma, H. Q. Guo, X. Xiao, and B. Zhao. 2018. Assessing the extent and impact of online data sharing in eddy covariance flux research. Journal of Geophysical Research: Biogeosciences 123:129–137.

Donoho, D. L. 2010. An invitation to reproducible computational research. Biostatistics 11:385–388.

Duke, C. S., and J. H. Porter. 2013. The ethics of data sharing and reuse in biology. BioScience 63:483–489.

Duke, C. S., G. Middendorf, and J. Wyndham. 2011. Science as a human right: ESA and the AAAS science and human rights coalition. Bulletin of the Ecological Society of America 92:61–63.

Ellison, A. M. 2010. Repeatability and transparency in ecological research. Ecology 91:2536–2539.

Ellison, A. M. 2017. Best practices for replicable research. Pages 250–263 *in* A. Shavit and A. M. Ellison, editors. Stepping into the same river twice: replication in biological research. Yale University Press, New Haven and London.

Ess, C., and F. Sudweeks. 2001. Culture, technology, communication: Towards an intercultural global village. State University of New York Press, Albany, New York.

Evans, T. M., L. Bira, J. B. Gastelum, L. T. Weiss, and N. L. Vanderford. 2018. Evidence for a mental health crisis in graduate education. Nature Biotechnology 36:282–284.

Fidler, F., Y. E. Chee, B. C. Wintle, M. A. Burgman, M. A. Mccarthy, and A. Gordon. 2017. Metaresearch for evaluating reproducibility in ecology and evolution. BioScience 67:282–289.

Gil, Y., et al. 2016. Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. Earth and Space Science 3:388–415.

Gustafson, E. J.. 2013. When relationships estimated in the past cannot be used to predict the future using mechanistic models to predict landscape ecological dynamics in a changing world. Landscape Ecology 28:1429–1437.

Hackett, E. J., J. N. Parker, D. Conz, D. Rhoten, and A. Parker. 2008. Ecology transformed: The National Center for ecological analysis and synthesis and the changing patterns of ecological research. MIT Press, Cambridge, Massachusetts.

Hampton, S. E., C. A. Strasser, and J. J. Tewksbury. 2013. Growing pains for ecology in the Twenty-First Century. BioScience 63:69–71.

Hampton, S. E., et al. 2015. The Tao of open science for ecology. Ecosphere 6:120.

Hampton, S. E., et al. 2017. Skills and knowledge for data-intensive environmental research. BioScience 67:546–557.

Heimstadt, M., F. Saunderson, and T. Heath. 2014. From toddler to teen: growth of an open data ecosystem. eJournal of eDemocracy Open Government 6:123–135.

Hollister, J., and J. Stachelek. 2017. Lakemorpho: Calculating lake morphometry metrics in R [version 1, referees: 2 approved]. F1000Research 6:1718.

Huang, X. 2014. Reproducibility in ecological research. Science 346:1307–1307.

Huettmann, F. 2005. Databases and science-based management in the context of wildlife and habitat: toward a certified ISO standard for objective decision-making for the global community by using the internet. Journal of Wildlife Management 69:466–472.

Jones, M. B., M. P. Schildhauer, O. J. Reichman, and S. Bowers. 2006. The new bioinformatics: Integrating ecological data from the gene to the biosphere. Annual Review of Ecology, Evolution, and Systematics 37:519–544.

LaDeau, S. L., B. A. Han, E. J. Rosi-Marshall, and K. C. Weathers. 2017. The next decade of big data in ecosystem science. Ecosystems 20:274–283.

Leek, J. T., and R. D. Peng. 2015. Reproducible research can still be wrong: adopting a prevention approach. Proceedings of the National Academy of Sciences of USA 112:1645–1646.

Lowndes, J. S. S., B. D. Best, C. Scarborough, J. C. Afflerbach, M. R. Frazier, C. C. O'Hara, N. Jiang, and B. S. Halpern. 2017. Our path to better science in less time using open data science tools. Nature Ecology & Evolution 1:0160.

Magnuson, J. J. 1990. Long-term ecological research and the invisible present. BioScience 40:495–501.

McKiernan, E. C., et al. 2016. How open science helps researchers succeed. Elife 5:e16800.

Michener, W. K. 2015. Ecological data sharing. Ecological Informatics 29:33–44.

Michener, W. K. 2016. Advances in managing long term ecological research data. Ecological Informatics 36:199–200.

Michener, W. K., and N. B. Jones. 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology & Evolution 27:85–93.

Milcu, A., et al. 2018. Genotypic variability enhances the reproducibility of an ecological study. Nature Ecology & Evolution 2:279–287.

Milojevic, S. 2014. Research team formation and evolution. Proceedings of the National Academy of Sciences USA 111:3984–3989.

Munafo, M. R., et al. 2017. A manifesto for reproducible science. Nature Human Behavior 1:0021.

Nielsen, M.. 2011. Reinventing discovery: The new era of networked science. Princeton University Press, Princeton, New Jersy, USA.

Peng, R. D.. 2011. Reproducible research in computational science. Science 334:1226–1227.

Peters, D. P. C., and G. S. Okin. 2017. A toolkit for ecosystem ecologists in the TME of big science. Ecosystems 20:259–266.

Piwowar, H. A., and T. J. Vision. 2013. Data reuse and the open data citation advantage. Peerj 1:UNSP e175.

Piwowar, H. A., R. S. Day, and D. B. Fridsma. 2007. Sharing detailed research data is associated with increased citation rate. PLoS ONE 2:e308.

Plesser, H. 2018. Reproducibility vs replicability: A brief history of a confused terminology. Frontiers in Neuroinformatics 11:76.

Poisot, T., D. Gravel, S. Leroux, S. A. Wood, M. J. Fortin, B. Baiser, A. Cirtwill, M. B. Araujo, and D. B. Stouffer. 2016. Synthetic datasets and community tools for the rapid testing of ecological hypotheses. Ecography 39:402–408.

Popper, K. R. 1934. The logic of scientific discovery. Verlag von Julius Springer, Vienna, Austria.

Reichman, O. J., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. Science 331:703–705.

Rocchini, D., V. Petras, A. Petrasova, N. Horning, L. Furtkevicova, M. Neteler, B. Leutner, and M. Wegmann. 2017. Open data and open source for remote sensing training in ecology. Ecological Informatics 40:57–61.

Sandve, G. K., A. Nekrutenko, J. Taylor, and E. Hovig. 2013. Ten simple rules for reproducible computational research. PLoS Computational Biology 9:e1003285.

Schimel, D. 2018. Editorial. Ecological Applications 28:869–870.

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

Schnitzer, S. A., and W. P. Carson. 2016. Would ecology fail the repeatability test? BioScience 66:98–99.

Simmons, L. W. 2016. Mandatory data archiving in Behavioral Ecology. Behavioral Ecology 27:1.

Soranno, P. A., K. S. Cheruvelil, K. C. Elliott, and G. M. Montgomery. 2015. It's good to share: Why environmental scientists' ethics are out of date. BioScience 65:69–73.

Stodden, V., P. Guo, and Z. Ma. 2013. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. PLoS ONE 8:e67111.

Stodden, V., J. Seiler, and Z. Ma. 2018. An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences USA 115:2584–2589.

Szitenberg, A., M. John, M. L. Blaxter, and D. H. Hunt. 2015. ReproPhylo: An environment for reproducible phylogenomics. PLoS Computational Biology 11:e10004447.

Vanderbilt, K., and D. Blankman. 2017. Reliable metadata and the creation of trustworthy, reproducible, and re-usable data sets. Pages 179–194 *in* A. Shavit and A. M. Ellison, editors. Stepping into the same river twice: replication in biological research. Yale University Press, New Haven, Connecticut and London.

Willinksy, J. 2006. Why open access to research and scholarship? Journal of Neuroscience 26:9078–9079.

Woolston, C. 2018. Feeling overwhelmed by academia? You are not alone. Nature 557:129–131.