

The Ethics of Data Sharing and Reuse in Biology

CLIFFORD S. DUKE AND JOHN H. PORTER

Recent increases in capabilities for gathering, storing, accessing, and sharing data are creating corresponding opportunities for scientists to use data generated by others in their own research. Although sharing data and crediting sources are among the most basic of scientific ethical principles, formal ethical guidelines for data reuse have not been articulated in the biological sciences community. This article offers a framework for developing ethical principles on data reuse, addressing issues such as citation and coauthorship, with the aim of stimulating a conversation in the science community and with the goal of having professional societies formally incorporate considerations of data reuse into their codes of ethics.

Keywords: citation, coauthorship, data reuse, data sharing, ethics

The so-called “data deluge” in the sciences in recent years has brought with it enormous increases in our ability to gather, store, access, and share data (Hey et al. 2009). These capabilities have in turn focused discussions on the opportunities for scientists to enhance their research through the use and analysis of data sets other than their own. Sharing data in its simplest sense—that of providing the primary data that underlie a published paper to those who wish to examine the original observations that support the authors’ conclusions—would appear to be among the most basic of scientific ethical principles, and a number of journals have adopted policies that mandate the sharing of data linked to publications (e.g., a consortium of evolutionary biology journals, *Ecological Monographs*, and *Science*; Whitlock et al. 2010, Ellison and Baldwin 2011, Hanson et al. 2011). However, the biological science community has not yet articulated clear ethical guidelines for data sharing and reuse, including appropriate citation and coauthorship practices (figure 1). The National Research Council (NRC) provides a set of principles for data sharing, including the responsibilities of authors, scientific societies, and publishers (NRC 2009), but does not present guidelines on data reuse. In this article, we take principles for data sharing as a starting point and offer some considerations for the development of ethical guidelines for data reuse. Practical solutions will also ultimately need to include consideration of legal constraints, but because of their complexity and the availability of other publications that address them (e.g., Miller et al. 2008, Reichman and Uhlir 2003), we will not further address them here.

Making the data that support publications publicly available benefits the data generator, the scientific community, and the public (table 1; Uhlir and Schröder 2007). Archived data are less likely to be lost as a result of accidents or computer failures, and the data are accessible from multiple locations. The scientific community gains by being able to discover, access, and analyze multiple data sets (e.g., Hampton et al. 2013). Furthermore, making data available allows papers to be checked for errors (Whitlock et al. 2010) and may reduce the potential for scientific fraud (e.g., South and Duke 2010, Molloy 2011). Similarly, making procedures and data available for public examination may be a key factor for public trust in scientists and the results of scientific research, particularly in the face of controversy (Beardsley 2010).

Data sharing also has implications for human rights. Article 15 of the International Covenant on Economic, Social and Cultural Rights requires governments to recognize the right to “enjoy the benefits of scientific progress and its applications”; to take the steps necessary for the “conservation, the development, and the diffusion of science”; to “respect the freedom indispensable for scientific research”; and to “recognize the benefits to be derived from the encouragement and development of international contacts and cooperation” in science.

Although work to define the detailed implications of article 15 is ongoing (e.g., Duke et al. 2011), the implementation of this right could imply an obligation to make data produced through publicly funded research publicly available. This interpretation is supported by funding agencies’

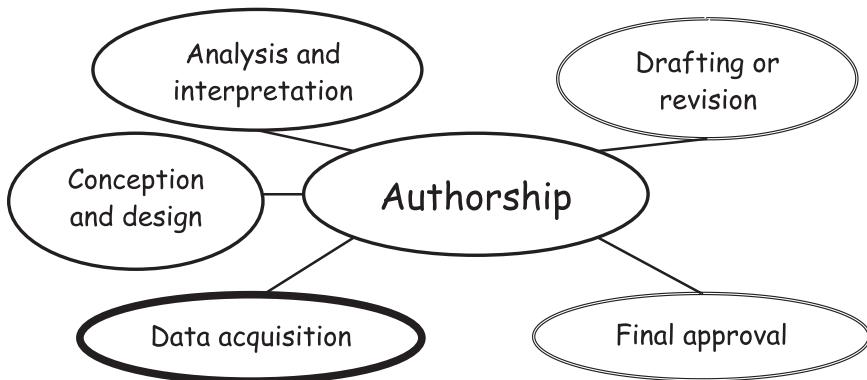


Figure 1. Criteria for authorship. One or more of the activities on the left (single border), plus two required activities (double border, on the right) are required for authorship. The bolded “Data acquisition” is not included in the criteria for authorship in most ecological journals, although it is in most medical journals.

Table 1. Why scientists have an ethical obligation to share data.

Obligation	Rationale
Replication	The ability to replicate results of analyses is a fundamental element of the scientific process, but for many types of time-sensitive environmental data, replication is possible only if the original data are available for reanalysis.
Human rights	Article 15 of the International Covenant on Economic, Social and Cultural Rights dictates that everyone has the right to “enjoy the benefits of scientific progress and its applications,” but failure to share data restricts that right.
Data preservation	Data that are not systematically archived are frequently lost as a result of computer failures, software obsolescence, or simple neglect.
Scientific progress	The scientific community, as a whole, benefits from the ability to discover, access, and analyze diverse data sets.
Data integrity	Multiple users make it more likely that deficiencies in data will come to light.
Public trust	In the face of controversy, the willingness to share data can be a key factor in increasing public trust in the results of scientific research.

growing emphasis on data management—for example, the US National Science Foundation’s recent requirement that a data management plan be included in proposals.

Obligations of data generators

In the past, the ethical obligation of scientists to share primary data has been constrained by technology; data were contained in notebooks or stored locally in various magnetic storage forms. Making data available to others involved copying paper or electronic files. Technological advances have largely eliminated these barriers, with the establishment of public archives such as GenBank, for genetic sequence data; the Protein Data Bank, for data on protein structures; and Dryad, for ecological and evolutionary biology data. However, the ethical codes and publication standards for biological sciences have only begun to integrate that recognition. For example, proteomics researchers have developed a set of principles for sharing data that support publications, analogous to those used by the genomic community (Rodriguez et al. 2009), and the journal *Molecular*

and Cellular Proteomics requires that primary data from the publication of genome- and proteome-scale analyses be published as supplemental data on the journal’s Web site. The Ecological Society of America (ESA) instructs contributors to its journals that “the editors and publisher expect authors to make the data underlying published articles available” and encourages authors to register their data sets. However, data archiving is required for only one (*Ecological Monographs*) of the society’s five peer-reviewed journals. The ESA’s Code of Ethics states that “Researchers will not submit for publication any manuscript containing data they are not authorized to use,” but does not make reference to data sharing per se.

The failure of many existing ethics policies to address data sharing, combined with the increasing demand for access to original data to address global issues, creates a need for professional societies to revise and extend their ethical guidelines. This is in keeping with recommendation 6 of the NRC (2009) that standards for sharing research data “be developed through a process that involves researchers, research institutions, research sponsors, professional societies, journals, representatives of other research fields, and representatives of public interest organizations, as appropriate for each

particular field” (p. 7). In the present article, we assume, as a starting point, that researchers have an ethical obligation to make public the data underlying their publications. Elaborating on this, we recommend that data generators provide those data in a timely fashion (preferably before publication) and ensure the accuracy and clarity of the associated metadata. This recommendation takes into account that a journal may permit an embargo period of up to 1 year (e.g., Whitlock et al. 2010), during which the data need not be exposed to the general public. When making data available, generators should not place unreasonable constraints on the use of data—for example, multiyear embargo periods for data linked to publications—especially in cases in which these constraints would restrict the scope of future use.

Obligations of data users

Researchers have the ethical obligation to share their data, but data users also have ethical obligations, which take several forms. The first and foremost is appropriate attribution

(assigning credit for the data), in the form of a statement of acknowledgment (unstructured attribution) or a citation (formal reference to data) (NRC 2012). Data citations, similar to bibliographic citations, provide a concise way to point readers to specific resources and to acknowledge the contributions of the data providers (Parsons et al. 2010, Ball and Duke 2012). The increasing availability of public data repositories (Nelson 2009, Ball and Duke 2012) and the use of persistent digital object identifiers make citation an increasingly attractive option (Brase 2009, Green 2009). As was noted by Schofield and colleagues (2009), “a mechanism, such as a digital object identifier for resources in public repositories, would allow ready searching of the literature for specific bioresources, which is currently extremely difficult. It would also add incentives for complying with data release and deposition policies by attributing credit to researchers who do share” (p. 171).

However, not all data that might be shared are citable. Loosely structured data informally exchanged among researchers may lack the cohesion and persistence required for a citation. In such a case, acknowledgment—either in the “Methods” section of the paper or in a separate “Acknowledgments” section—can be used. Regardless of the form of the citation or acknowledgment, data users have an ethical obligation to attribute data sources, both to provide credit to the data providers and to provide readers with the references required to replicate the analyses in the published work. The use of data sets without citation or acknowledgment is akin to plagiarism—taking credit for work done by others.

In addition to citation, the ethical obligations of data users may include offers of coauthorship to data providers whose data are integral to the final work. For most ecologists, this is a new area that may require revising authorship guidelines in order to maintain the core principles that author lists include individuals who made significant contributions to a publication and who are responsible for the quality and accuracy of the work. Currently, guidelines for authorship for ecological journals generally ignore the providers of data, perhaps assuming that authors are analyzing only data that they collected themselves. This assumption is less tenable when ecological data are increasingly shared. In the medical field, the International Committee of Medical Journal Editors publishes a document called the “Uniform Requirements for Manuscripts,” which describes the criteria for authorship. The third edition of this document introduced criteria for authorship that were similar to those used by most ecological journals: “Authorship credit should be based only on substantial contributions to (a) [the] conception and design, or analysis and interpretation of data; and to (b) drafting the article or revising it critically for important intellectual content; and on (c) final approval of the version to be published” (ICMJE 1988, p. 402). However, when the medical journal *The Lancet* began to require authors to self-identify their contributions, it was discovered that authorship lists frequently included

authors whose primary contribution was providing data (Garcia 2004). In recognition of this practice, a special statement was issued in 2000, and the 2004 edition of the requirements modified the criteria for authorship so that the first criterion included the “acquisition of data” (i.e., “substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data”; ICMJE 2004, p. 2). Although the statements of acceptable forms of intellectual contribution have broadened, the requirements to contribute to interpretation and to approve the final version of the manuscript remain intact.

During the long history of coauthorship, common practices have emerged that may vary by community or discipline (Weltzin et al. 2006). The core principle that authors contribute to and are responsible for their work is subject to some interpretation and is occasionally a source of bitter disagreement. For example, coauthorship of a manuscript is often not offered to technicians, who may have done the majority of the data collection but did not contribute to the design of the data-collection protocols or to the analysis. In other cases, coauthorship may be offered to individuals who, through help in obtaining funding or facilities, made the research possible but did not otherwise substantively participate in the production of the manuscript (e.g., heads of laboratory groups), although this practice has been increasingly challenged (Kassirer and Angell 1991).

Authorship guidelines such as the “Uniform requirements for manuscripts submitted to biomedical journals” (ICMJE 2004) emphasize “substantial contributions,” but when does a contribution of data from a data provider rise to the level of being “substantial”? There are several factors that could be considered in reaching a decision regarding coauthorship, including the degree to which the shared data are integral to the manuscript, unique or novel characteristics of the data, and the availability or willingness of the potential coauthor to participate in the drafting and revision of the manuscript and to take responsibility for the final product.

How integral a particular data set is to a manuscript is often a function of the number of other data sources used in the research. If all the data in a manuscript come from a single source, a strong case can be made for an offer of coauthorship. If all the data were removed from the manuscript, it would probably become unpublishable. In contrast, if the data analyzed come from a large number of similar data sets from multiple sources, citation of the data sources rather than an offer of coauthorship would be most appropriate. The rationale for this choice is twofold. First, removal of a single data set among many is unlikely to affect the publishability of the manuscript. Second, most journals balk at extremely long lists of authors (e.g., Kassirer and Angell 1991, Weeks et al. 2004).

A manuscript could be dependent on a specific data set either because it would not be possible to perform a particular analysis without the data or because the results of the analysis would be different were those data omitted.

In determining the desirability of coauthorship, the former cause should be weighted most heavily, because it focuses on the ability to perform an analysis, which could yield either positive or negative results. Such data are integral to the manuscript. The latter cause of dependency is less compelling, because whereas data providers are responsible for the availability and accuracy of their data, the specific data values observed are a function of the phenomenon being measured. Therefore, coauthorship should be offered for making integral data available (allowing an analysis to proceed), but coauthorship might not be offered just because one data set among many had unusual data values that were important in determining the results of a statistical test.

How integral a data set is to a manuscript becomes harder to assess when there is an intermediate number of similar data sets or if the data used vary widely in importance within the context of a manuscript. Just as there are gray areas surrounding what constitutes the “substantial contributions” of researchers to a manuscript, there are gray areas regarding how “central” or “integral” (NRC 2003) a given data set is to a manuscript. Therefore, additional factors should be considered.

A second factor is how novel the data are. Collecting novel types of data constitutes a greater intellectual contribution. This intellectual contribution may take the form of identifying new ways to measure a phenomenon, the recognition that spatially or temporally extensive data are needed, or the data’s integration in a specific way that is particularly valuable. There is a range in how novel different types of data are. Some data are collected using standardized tools in prescribed sampling regimes. Other data are collected using innovative methodologies or sampling schemes. Similarly, some data describe single locations at a point in time, whereas other data may consist of long-term observations over a large area. In some cases, a given data set may be derived from other data that have been integrated or harmonized in such a way as to provide additional value to the set. Coauthorship is more appropriate when the data

are novel in some way, be it the methodology with which they are collected, the sampling design, the spatial or temporal extent, or the level of integration.

A final and determinative factor in whether a data provider serves as a coauthor is his or her availability or willingness to fulfill the other requirements for authorship: participation in the drafting or revision of the manuscript and final approval of the version to be published. In some cases, a data provider may be deceased, in which case he or she would be unable to meet the additional criteria for coauthorship. In other cases, the data provider may be an institution, which—again—would not be able to fulfill the additional conditions for authorship. A special case arises when the data providers are asked whether they would be willing to be coauthors but decline, either because they do not want to dedicate the time and effort required to aid in the drafting and revision of the manuscript or because they disagree with the findings or opinions expressed in it. In such cases, publication can proceed without the data providers as coauthors, and the authors can simply cite or acknowledge the data product.

Some data sets require, as a condition of their use, a license or material transfer agreement that stipulates that the data set providers be included as coauthors on any publication using the data. This practice was condemned by the NRC (2003) because such restrictions would permit data providers to prevent publication of views that conflict with their own. However, although coauthorship should not be a mandatory adjunct to using data, it may still be ethically correct to offer coauthorship when the data are integral to a manuscript, provided that there is a clear understanding that publication of the manuscript will proceed regardless of whether a data provider accepts the offer. Apart from ethical considerations, including data providers as coauthors can have benefits for the quality of the manuscript as a result of the providers’ intimate familiarity with the data that may not be fully captured by metadata or other documentation. This follows the recommendation of

Wolkovich and colleagues (2012) that “Ideally, researchers repurposing data will involve data holders throughout the project—from hypothesis refinement to publication—since data holders usually have additional insights and understanding of their data that is difficult to capture in metadata” (p. 2104).

In applying these factors, it is useful to consider some examples (box 1a–1c). An alternative model discussed by Weltzin and colleagues (2006) is replacing authorship with “contributorship,” wherein everyone contributing to a manuscript is listed along with a description of the roles that each individual played in its production.

Box 1a. Criteria for determining whether an offer of coauthorship is appropriate: Example 1.

A data provider has developed extremely small Global Positioning System trackers and has used them to collect high-resolution data on the locations of many small mammals over an extended period. Following a publication on the methods and basic movement statistics, the data are made available through a repository to other researchers. A researcher downloads those data and applies spatial statistics to characterize the social dynamics of a species. Should the researcher contact the data provider regarding coauthorship? (See figure 2.)

The data are essential to the success of a manuscript, because they are the only data used in the paper, probably because of their novel characteristics (high-resolution locations). Therefore, a strong case can be made that the data providers should be offered an opportunity to collaborate on the manuscript as coauthors. However, even if they decline to participate, publication can proceed, and the authors can cite or acknowledge the data set.

Box 1b. Criteria for determining whether an offer of coauthorship is appropriate: Example 2.

A researcher has assembled data on nutrient levels from three streams and used them to test a hypothesis regarding the seasonality of nutrient loading. Data for each of the streams come from different sources. One data set is from a graduate student; another is from a senior professor, now deceased; and the third comes from a government agency. Should the researcher contact the data providers regarding coauthorship? (See figure 2.)

Here, the decision is harder than in example 1 (see box 1a). The researcher needs to decide whether the paper would be publishable if only two of the data sources were used, whether there are other stream nutrient data sets that could equally well be substituted, or whether there are novel characteristics about these particular data that make them essential to the paper. If the researcher concludes that all three data sets are essential to the success of the paper, he or she may wish to tender an offer of coauthorship to the graduate student. The deceased professor can no longer meet the additional requirements for authorship. Similarly, the government agency is probably not in a position to participate in the drafting and revision of the manuscript. If the researcher concludes that the manuscript would be publishable using only two of the data sets or that there are many similar data sets that could be substituted, citation but not coauthorship may be appropriate. As was the case above (box 1a), even if the potential coauthor declines to participate, publication of the manuscript should proceed.

Box 1c. Criteria for determining whether an offer of coauthorship is appropriate: Example 3.

A researcher develops a model for predicting storm movement and tests it using meteorological data from 250 different stations, each operated independently. Should the researcher contact the data providers regarding coauthorship? (See figure 2.)

On the basis of our criteria, the answer is *certainly not*. Removing any one of the 250 stations might affect the results of the model validation, but it would not render the testing of the model impossible; therefore, none of the individual data sets is integral to the paper. Similarly, meteorological station data are relatively standardized, so there is little case to be made for any specific data set being novel or unique. Finally, most journals would not accept a paper with 251 authors. However, the sources of the data should still be cited or acknowledged.

A special challenge for proper citation is posed by data sets that integrate, aggregate, or otherwise alter existing data sets in order to provide additional value. For example, a researcher might create a data set by integrating a large number of stream-chemistry data sets, each in their own format, with different units and taken at different time steps, into a single data set with consistent format, units, and time steps and with additional quality control and assurance testing. The resulting data set is substantively different from the data used to create it, especially where aggregation or interpolation is used, because it may be impossible to re-create or extract the original data from the integrated data set. This integrated data set may itself be published in a public data repository. In determining the authorship of the integrated data set, the previously discussed criteria can be applied. However, there is an additional obligation that the metadata associated with the integrated data set include information about the provenance of the data, such as the

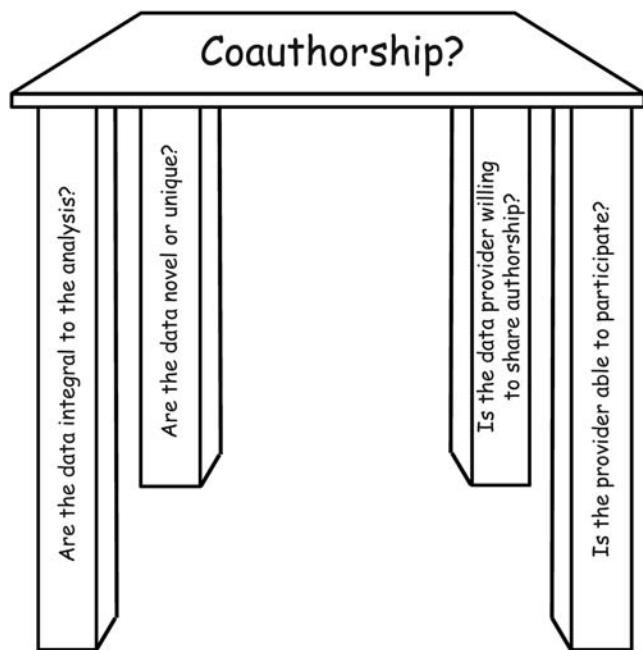


Figure 2. Criteria for coauthorship by a data provider.
All criteria must be met for a provider of data to be included as a coauthor. Regardless of whether an offer of coauthorship is made or accepted, publication can proceed by incorporating citations or acknowledgments for the data source.

list of data sets that were integrated. Computer systems used in repositories are increasingly supporting automated or semiautomated recording of provenance information (Servilla et al. 2008, Michener et al. 2011). However, this topic is sufficiently technical that we will not pursue it further, except to note that, as in review papers, researchers using such integrated or value-added data sets typically cite the integrated data set rather than all the individual data sets of which it is composed.

Box 2. Ethical principles for data reuse.

Publishers should revise authorship guidelines to address the contribution of data providers, and professional societies should explore formally incorporating the consideration of data reuse into their codes of ethics.

Data generators should make public the data underlying their publications in a timely fashion and should ensure the accuracy and clarity of the associated metadata.

Conditions for the use of data should never include restrictions that limit the range of scientific conclusions that can be reached.

Data users have an obligation to cite data sources, through formal citations when that is possible or through acknowledgment otherwise.

Coauthorship should be offered to data providers if their data are integral to the manuscript and if they can meet the other conditions of authorship, such as participation in the preparation of the manuscript and acceptance of responsibility for the conclusions reached.

If a data provider declines to accept coauthorship, perhaps because of scientific disagreements, the use of the data and publication should proceed anyway.

Sharing data is an ethical obligation for scientists, because the timely sharing of scientific data benefits both science and society.

Conclusions

The development of both greater demand and greater resources for public data sharing raises important ethical questions for data generators and for those who use data generated by others (box 2). Sharing data is becoming increasingly important to science, both for the integrity of science (i.e., replication) and for the development of synthetic data products that allow existing data to be applied to new problems. However, the degree to which researchers are willing to share data depends to a large extent on the responsible conduct of their colleagues who reuse that data. It is crucial that data users honor the trust placed in them by those who make their data available, by providing proper attribution or coauthorship and accepting reasonable constraints on data use. Formal ethical guidelines (for any purpose) developed by a community, for the community, can help determine the boundaries of such constraints and can facilitate decisionmaking by individuals about how to act in particular circumstances. Extending those ethical guidelines to address the issues associated with data reuse will help support a culture of data sharing. We hope that the guidelines proposed here will stimulate a conversation in the science community, with the goal of having professional societies formally incorporate considerations of data reuse into their codes of ethics.

Acknowledgments

This article is a product of the DataONE Community Engagement and Education Working Group, which was supported by US National Science Foundation DataONE award no. OCI-0830944. JP was supported by US National Science Foundation grant no. DEB-0621014. Wendy Gram, Stephanie Hampton, Carly Strasser, and other members of the DataONE Community Education and Engagement Working Group and three anonymous reviewers provided many helpful and insightful comments.

References cited

- Ball A, Duke M. 2012. How to cite datasets and link to publications. Digital Curation Centre. (12 March 2013; www.dcc.ac.uk/resources/how-guides/cite-datasets)
- Beardsley T. 2010. The biologist's burden. *BioScience* 60: 483.
- Brase J. 2009. DataCite: A global registration agency for research data. Pages 257–261 in Proceedings of the Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology (COINFO 2009). Institute of Electrical and Electronics Engineers.
- Ellison AM, Baldwin JD. 2011. Editorial. *Ecological Monographs* 81: 1–2.
- Duke CS, Middendorf G, Wyndham J. 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.
- Garcia AM. 2004. Sixth version of the Uniform Requirements for Manuscripts Submitted to Biomedical Journals. *Journal of Epidemiology and Community Health* 58: 731–733. doi:10.1136/jech.2004.024802
- Green T. 2009. We need publishing standards for datasets and data tables. *Organisation for Economic Co-operation and Development*. doi:10.1787/787355886123
- Hampton SE, Strasser SA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH. 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11: 156–162. doi:10.1890/120103
- Hanson B, Sugden A, Alberts B. 2011. Making data maximally available. *Science* 331: 649.
- Hey T, Tansley S, Tolle K, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.
- [ICMJE] International Committee of Medical Journal Editors. 1988. Uniform requirements for manuscripts submitted to biomedical journals. *British Medical Journal* 296: 401–405.
- . 2004. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. ICMJE. (12 March 2013; www.icmje.org/2004_urm.pdf)
- Kassirer JP, Angell M. 1991. On authorship and acknowledgments. *New England Journal of Medicine* 325: 1510–1512. doi:10.1056/NEJM19911213252112
- Michener WK, Porter J, Servilla M, Vanderbilt K. 2011. Long term ecological research and information management. *Ecological Informatics* 6: 13–24.
- Miller P, Styles R, Heath T. 2008. Open data commons, a licence for open data. In Bizer C, Heath T, Idehen K, Berners-Lee T, eds. *Proceedings of the Linked Data on the Web Workshop*, Beijing, China, April 22, 2008, CEUR Workshop Proceedings, ISSN 1613-0073. (20 March

- 2013; <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS-Vol-369/paper08.pdf>)
- Molloy JC. 2011. The Open Knowledge Foundation: Open data means better science. *PLOS Biology* 9 (art. e1001195).
- Nelson B. 2009. Data sharing: Empty archives. *Nature* 461: 160–163.
- [NRC] National Research Council. 2003. *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. National Academies Press.
- . 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. National Academies Press.
- . 2012. *For Attribution: Developing Data Attribution and Citation Practices and Standards*. National Academies Press.
- Parsons MA, Duerr R, Minster J-B. 2010. Data citation and peer-review. *Eos, Transactions, American Geophysical Union* 91: 297–298.
- Reichman JH, Uhlig PF. 2003. Contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law and Contemporary Problems* 66: 315–462. (12 March 2013; www.jstor.org/stable/20059181)
- Rodriguez H, et al. 2009. Recommendations from the 2008 international summit on proteomics data release and sharing policy: The Amsterdam principles. *Journal of Proteome Research* 8: 3689–3692.
- Schofield PN, et al. 2009. Post-publication sharing of data and tools. *Nature* 461: 171–173.
- Servilla M, Costa D, Laney C, San Gil I, Brunt J. 2008. The ECOTRENDS Web portal: An architecture for data discovery and exploration. Pages 139–144 in Gries C, Jones MB, eds. *Proceedings of the Environmental Information Management Conference 2008 (EIM 2008)*. US National Center for Ecological Analysis and Synthesis. (12 March 2013; <https://eim.ecoinformatics.org/eim2008/cp-eim2008/eim-proceedings-2008>)
- South DB, Duke CS. 2010. Will a data registry increase professional integrity? *Journal of Forestry* 108: 370–371.
- Uhlig PF, Schröder P. 2007. Open data for global science. *Data Science Journal* 6: OD36–OD53.
- Weeks WB, Wallace AE, Surott Kimberly BC. 2004. Changes in authorship patterns in prestigious US medical journals. *Social Science and Medicine* 59: 1949–1954.
- Weltzin JF, Belote RT, Williams LT, Keller JK, Engel EC. 2006. Authorship in ecology: Attribution, accountability, and responsibility. *Frontiers in Ecology and the Environment* 4: 435–441.
- Whitlock MC, McPeek MA, Rausher MD, Rieseberg L, Moore AJ. 2010. Data archiving. *American Naturalist* 175: 145–146.
- Wolkovich EM, Regetz J, O'Connor MI. 2012. Advances in global change research require open science by individual researchers. *Global Change Biology* 18: 2102–2110.

Clifford S. Duke (csduke@esa.org) is the director of science programs at the Ecological Society of America, in Washington, DC. John H. Porter (jhp7e@virginia.edu) is a research associate professor in the Department of Environmental Sciences at the University of Virginia, in Charlottesville.