

Scientific Journals can Advance Tropical Biology and Conservation by Requiring Data Archiving

Emilio M. Bruna¹

Department of Wildlife Ecology and Conservation and Center for Latin American Studies, University of Florida, PO Box 110430, Gainesville, Florida 32611-0430, U.S.A.

DATA ARE THE FOUNDATION FOR SCIENTIFIC PROGRESS. While researchers often report statistical summaries of the data they collect in their publications, there is widespread recognition that the raw data themselves are increasingly useful for meta-analyses, to address novel questions, or as baselines for studies evaluating temporal changes in biological processes (Clark 2007, Whitlock *et al.* 2010). In many fields, the archiving of raw data as a precondition for publication of research papers has been common practice for decades. For instance, many journals require that the DNA sequences used in evolutionary or genomic studies be deposited in archives such as the National Center for Biotechnology Information's GenBank (<http://www.ncbi.nlm.nih.gov>) or The Arabidopsis Information Resource (<http://www.arabidopsis.org>), making them publicly available to other researchers. Similarly, fields such as oceanography and atmospheric sciences also have a well-developed tradition of data sharing (*e.g.*, the British Oceanographic Data Center [<http://www.bodc.ac.uk>]).

Despite admirable efforts to promote data archiving by several international research consortia (*e.g.*, the Center for Tropical Forest Science and the Large-Scale Biosphere Atmosphere Program) and academic societies (*e.g.*, the Ecological Society of America's Data Registry program), this practice has yet to become widespread among ecologists (Nelson 2009). Consequently, many datasets are unavailable to the scientific community even years after they were last used by the researchers that collected them. Most of these datasets are eventually lost permanently—they are stored on media that are outdated or unreliable (*e.g.*, punch cards, floppy disks, hard drives, cassette tapes) or are discarded as investigators pursue new avenues of research, change careers, retire, or die (Michener *et al.* 1997). Even datasets that are still accessible can be effectively lost because they are poorly organized or lack the metadata necessary for their interpretation (Michener *et al.* 1997, Borer *et al.* 2009).

The loss of these datasets—usually collected at considerable expense and great personal effort—is an impediment to scientific progress that could be avoided if researchers placed datasets and associated metadata in permanent digital data repositories. Awareness of this problem has led to an increasingly loud call for researchers to clarify where and how the data used in their publications are archived, as well as for funding agencies and journals to develop mechanisms to promote this practice (*e.g.*, Clark 2007, Anonymous

2009, Schofield *et al.* 2009). Now a suite of prestigious journals that includes *Evolution* and *The American Naturalist* has taken the bold step of requiring that data used in papers they publish be archived and made freely available in a digital data repository (Whitlock *et al.* 2010). The editors of these journals argue that in addition to minimizing the loss of priceless data, archiving advances the field because it encourages novel studies or meta-analyses, promotes a culture of transparency, and makes it easier for others to verify results (Whitlock *et al.* 2010).

I propose that the Association for Tropical Biology and Conservation (ATBC) consider adopting a similar policy for papers published in *Biotropica*. In addition to the potential benefits highlighted by Whitlock *et al.* (2010), the archiving of data on which *Biotropica*'s papers are based will bring with several other tangible benefits. First, tropical ecosystems are undergoing myriad, rapid, and unprecedented environmental changes. The data collected by *Biotropica*'s authors could provide an invaluable resource to the scientists and decision-makers studying global change phenomena and designing conservation and management strategies. For example, Clark *et al.* (2010) used publicly available datasets of tree growth (Clark & Clark 2006), atmospheric CO₂ levels (Keeling *et al.* 2005, available at <http://scrippsco2.ucsd.edu>), and local climate (available from the Organization for Tropical Studies at <http://www.ots.duke.edu>) to study wood production in tropical forests; it is only because they were able to integrate these archived datasets that they were able to conclude that wood production in lowland rain forests may be severely reduced in future climates that are only slightly drier or warmer than present-day ones. Second, archived data could play a critical role in capacity building—undergraduates and graduate students will be able to use previously inaccessible datasets to conduct novel studies or meta-analyses for their theses. Because research using publicly available data requires no expenditure beyond stipends and the purchase of computers, they are also a very cost-efficient means of increasing a country's scientific productivity and developing a cadre of highly trained and productive postdoctoral scientists (*sensu* the postdoctoral fellows of the National Center for Ecological Analysis and Synthesis, <http://www.nceas.ucsb.edu>). Third, publicly archiving data could help assuage concerns regarding the export of intellectual property, biopiracy, and the failure to include local scientists in data collection efforts that often plague foreign scientists working in tropical countries (Stocks *et al.* 2008). Finally, it is the policy of many government agencies—as stated in a recent request for proposals of the United States Department of

Received 18 March 2010; revision accepted 28 March 2010.

¹Corresponding author; e-mail: embruna@ufl.edu

Energy (USDE)—that ‘data obtained through public funding are a public trust’ (USDE 2010) and must therefore be made publicly available by grantees. By encouraging and facilitating the archiving of data, journals can play a role in helping researchers meet the reporting requirements of funding agencies and the ethical obligations they have to the citizens that finance their scientific endeavors.

Requiring data be archived as a precondition for publication is obviously not without legitimate concerns that must be addressed. First and foremost, researchers will be justifiably hesitant about making data freely available to others when these data are still being used for analyses to be published in subsequent articles. For that reason, some journals have a policy allowing archived data to remain embargoed following an article’s publications. For *The American Naturalist* the embargo period is 1 yr (Whitlock *et al.* 2010); this length of time is arbitrary and should probably be longer for *Biotropica*. Of course, scientists interested in gaining access to data before the end of the embargo could always contact the authors of studies directly.

Second, scientists might be reluctant to archive datasets—even those that have not been used in years—because of they are concerned they will not receive citations, co-authorship, or other forms of attribution for their efforts. Such attribution is important—demonstrating one’s work is cited by the scientific community is playing an increasingly prominent role in professional evaluation and advancement, and the productivity resulting from previous grants is often a key metric used in evaluating subsequent applications. Although thornier, this obstacle can be overcome by, for instance, allowing authors to request co-authorship on papers using certain datasets or by developing a mechanism whereby datasets can be cited (*e.g.*, the Data Papers published in *Ecology* [<http://esapubs.org/archive>]).

Third, there may be cases in which making data publicly available is politically, culturally, or biologically problematic. For example, scientists investigating endangered species or those subject to poaching may not wish to reveal the precise location in which these species were found, and studies involving human subjects or indigenous groups often require permission before data can be disseminated. In cases such as these, the Editor-in-Chief may make exceptions or allow for data to be archived in such a way that it is only accessible after approval by the study’s authors.

Fourth, obligatory data archiving requires an infrastructure that ensures long-term data integrity, access, and flexible data entry structure suitable for the often idiosyncratic data collected in field-based studies. Fortunately, it is not necessary for *Biotropica*’s publisher or the ATBC to develop and maintain this infrastructure—several excellent options for the long-term archiving of ecological data already exist. These include the Knowledge Network for Biocomplexity (<http://knb.ecoinformatics.org>) and the Dryad (<http://www.datadryad.org>), both of which are designed for easily archiving ecological datasets and facilitating access to them by other researchers.

Finally, authors may choose to publish their papers in journals that do not require data archiving. One study, however, found that the papers of authors that archive their data are cited much more frequently than those that do not (Piwowar *et al.* 2007). In light of the increasing emphasis placed on impact factors by universities and

funding agencies, this is a clear benefit of an archiving policy to both the journal and the authors.

All of these concerns regarding archiving have been successfully addressed by numerous journals and academic societies; there is no reason to expect any differently from the membership of the ATBC. Although I expect there would be some initial and perhaps vocal resistance to such a policy, I believe that on reflection many of our members will conclude that any potential drawbacks are far outweighed by the benefits of permanently archiving data. In addition to minimizing the permanent and deplorable loss of thousands of datasets, I believe the most important consequence of making data broadly available is the demystification of the scientific process and the means by which scientists arrive at their conclusions. A lack of transparency, failure to share data, and misunderstanding of the scientific method were core elements of the recent ‘climategate’ scandal (Heffernan 2009), and one of the sadder things to emerge from the debacle is that an erosion of public confidence in scientists stemming from a perception they manipulate data to suit their foregone conclusions. Because our field sites are not only biological hotspots but also political and economic ones, our conclusions can be similarly contentious and we are therefore not immune to such accusations. Requiring authors to archive data will increase public confidence in the integrity of the scientific process, advance our discipline, and ensure *Biotropica*’s reputation a premier and innovative journal.

LITERATURE CITED

- ANONYMOUS. 2009. Data’s shameful neglect. *Nature* 461: 145.
- BORER, E. T., E. W. SEABLOOM, M. B. JONES, AND M. SCHILDHAUER. 2009. Some simple guidelines for effective data management. *Bull. Ecol. Soc. Am.* 90: 205–214.
- CLARK, D. A. 2007. Detecting tropical forests’ responses to global climatic and atmospheric change: Current challenges and a way forward. *Biotropica* 39: 4–19.
- CLARK, D. B., AND D. A. CLARK. 2006. Tree growth, mortality, physical condition, and microsite in an old-growth lowland tropical rain forest. *Ecology* 87: 2132.
- CLARK, D. B., D. A. CLARK, AND S. F. OBERBAUER. 2010. Annual wood production in a tropical rain forest in NE Costa Rica linked to climatic variation but not to increasing CO₂. *Global Change Biol.* 16: 747–759.
- HEFFERNAN, O. 2009. Climate data spat intensifies. *Nature* 460: 787.
- KEELING, C. D., S. C. PIPER, R. B. BACASTOW, M. WAHLEN, T. P. WHORF, M. HEIMANN, AND H. A. MEIJER. 2005. Atmospheric CO₂ and 13CO₂ exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications. In J. R. EHLERINGER, T. E. CERLING, AND M. D. DEARING (Eds.). *A history of atmospheric CO₂ and its effects on plants, animals, and ecosystems*, pp. 83–113. Springer Verlag, New York, New York.
- MICHENER, W. K., J. W. BRUNT, J. J. HELLY, T. B. KIRCHNER, AND S. G. STAFFORD. 1997. Nongeospatial metadata for the ecological sciences. *Ecol. Appl.* 7: 330–342.
- NELSON, B. 2009. Data sharing: Empty archives. *Nature* 461: 160–163.
- 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.
- SCHOFIELD, P. N., T. BUBELA, T. WEAVER, L. PORTILLA, S. D. BROWN, J. M. HANCOCK, D. EINHORN, G. TOCCHINI-VALENTINI, M. HRABE DE ANGELIS,

- AND N. ROSENTHAL. 2009. Post-publication sharing of data and tools. *Nature* 461: 171–173.
- STOCKS, G., L. SEALES, F. PANIAGUA, E. MAEHR, AND E. M. BRUNA. 2008. The geographical and institutional distribution of ecological research in the tropics. *Biotropica* 40: 397–404.
- USDE. 2010. Program Announcement to DOE National Laboratories LAB 10-04. Available at http://www.er.doe.gov/grants/LAB10_04.html
- WHITLOCK, M. C., M. A. MCPEEK, M. D. RAUSHER, L. RIESEBERG, AND A. J. MOORE. 2010. Data archiving. *Am. Nat.* 175: 145–146.