

Biodiversity Data Sharing: Will Peer-Reviewed Data Papers Work?

XIAOLEI HUANG, BRADFORD A. HAWKINS, AND GEXIA QIAO

The public availability of primary biodiversity data is essential to biodiversity and ecological research, natural resource management, and decision-making. To this end, there is an urgent need for primary data in the public domain (Hampton et al. 2012). In February 2012, the Global Biodiversity Information Facility (GBIF) released a position paper advocating the peer-reviewed publication of biodiversity data, which was actually an update of their recent advocacy of *data papers*—the “scholarly publication of a searchable metadata document describing a particular online accessible dataset or a group of datasets” (Chavan and Penev 2011, p. 3)—in conjunction with Pensoft journals (Chavan and Penev 2011, Penev et al. 2011). The justifications for data papers are related to professional recognition (e.g., they are citable) and data quality control (e.g., they are peer reviewed). Although the data paper method seems helpful, we have some concerns about its applicability.

First, it is important to understand potential authors of data papers and their behavior. There are three basic groups that collect and share primary biodiversity data. One is natural history museums and collections, most of which are publicly funded organizations. Data papers may be of limited interest to them, because sharing data with the public is a basic responsibility for them, and their focus is already on making data directly available online either at the institutional level or as part of a larger consortium. Another group is that of so-called “citizen scientists.” The members of this group may be interested in sharing their biodiversity data through a data paper, but the mechanics of scientific

publication may be unfamiliar to many of them. The third group is professional researchers, whose research is typically question driven. Their *modi operandi* dictate that most of them are unwilling to publish primary data without analyses. We recently conducted an international survey on biodiversity data sharing in which we found that most researchers are willing to share their primary biodiversity data only after related research papers are published (Huang et al. 2012).

For the third group, it then becomes a question of whether it is appropriate to publish a data set as part of a research paper (e.g., as online supplemental materials) and then again in a data paper. Authors are already expected by researchers and journal editors to share their primary data sets when publishing research papers (Vision 2010), and the utility of these data can be increased by modifying data-archiving policies instead of by publishing duplicate data sets in data papers (see below about the joint data-publishing and -archiving policy). In addition, although the publication of data papers may help simplify and clarify intellectual property rights issues, such as data ownership and citation, the peer-review process for data papers may put an extra burden on the limited reviewer resource in the scientific community, a problem that is described as “the tragedy of the reviewer commons” (Hochberg et al. 2009).

Another issue is whether a data paper should have a minimum size. For example, a data set including 20 occurrence records of a widespread species is probably not worth publishing as a data paper. Who should foot the bill for the publication of

data papers is also debatable. Article charges required by open-access journals (e.g., the Pensoft journal *Zookeys* requires a minimum fee of €300 for papers shorter than 20 printed pages) will prevent some proportion of data sets from being published, especially for citizen scientists or for underfunded workers in parts of the world where primary species distribution data are most needed.

To address professional recognition and data quality control, there are viable alternatives to the data paper. The implementation of a joint data-publishing and -archiving policy by databases and journals (Huang and Qiao 2011) can resolve most problems. First, data repositories should issue a unique data set identifier or DOI (digital object identifier) for each data set (whether it is directly submitted or submitted with a paper), which can easily resolve data ownership and citation issues. The optimal mode would use an interoperable identifier or index across all databases. Second, since it is a basic responsibility of authors to share usable data sets, journals and databases could adopt a strict data-archiving policy and could require authors to prepare detailed metadata descriptions when they submit primary data. Such a policy would also help control data set quality. To achieve this, scientists, journals, and data repositories should agree to certain data standards. A recently reported coordinated initiative of adopting data-publishing policy by freshwater journals is a good example (De Wever et al. 2012).

Although the data paper advocates recommended a GBIF–Pensoft workflow (Chavan and Penev 2011, Penev et al. 2011), only five data papers were published in the Pensoft journal

Zookeys from May 2011 through October 2012. We think that instead of popularizing a new kind of publication, it is more important to improve current peer-review processes and the operating policies and integration of journals and databases. We should applaud any effort for biodiversity data sharing. However, the suggestion that researchers publish data papers gives us pause, for the reasons explained above. The opinions of all stakeholders must be considered and incorporated into policymaking frameworks to determine how best to share primary biodiversity data.

Acknowledgments

We thank Colin Favret for valuable comments on the early drafts of this article. XH and GQ are funded by National Science Foundation of

China grants no. NSFC-30900133 and NSFC-31025024, respectively.

References cited

- Chavan V, Penev L. 2011. The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12 (suppl. 15, art. S2). doi:10.1186/1471-2105-12-S15-S2
- De Wever A, Schmidt-Kloiber A, Gessner MO, Tockner K. 2012. Freshwater journals unite to boost primary biodiversity data publication. *BioScience* 62: 529–530.
- Hampton SE, Tewksbury JJ, Strasser CA. 2012. Ecological data in the Information Age. *Frontiers in Ecology and the Environment* 10: 59.
- Hochberg ME, Chase JM, Gotelli NJ, Hastings A, Naeem S. 2009. The tragedy of the reviewer commons. *Ecology Letters* 12: 2–4.
- Huang X, Qiao G. 2011. Biodiversity databases should gain support from journals. *Trends in Ecology and Evolution* 26: 377–378.
- Huang X, Hawkins BA, Lei F, Miller GL, Favret C, Zhang R, Qiao G. 2012. Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. *Conservation Letters* 5: 399–406.

- Penev L, Mietchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D. 2011. Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Pensoft. (29 October 2012; <http://www.pensoft.net/page.php?P=23>)
- Vision TJ. 2010. Open data and the social contract of scientific publishing. *BioScience* 60: 330–331.

Xiaolei Huang (huangxl@ioz.ac.cn) is an associate professor and Gexia Qiao is a professor at the Institute of Zoology, Chinese Academy of Sciences, in Beijing. XH studies biodiversity patterns and biogeography and is interested in data-sharing issues in the biological sciences. GQ is also director of the National Zoological Museum of China, in Beijing, and principal investigator of projects to digitize animal-specimen data in natural history collections in China. Bradford A. Hawkins is a professor at the University of California, Irvine. He is interested in biogeography, macroecology, and diversity gradients.

doi:10.1525/bio.2013.63.1.2

Graduate Student Leaders Sought to Shape Science Policy

Earn recognition as an Emerging Public Policy Leader and receive a free trip to Washington, DC to meet with policymakers.

Students in the biological sciences with demonstrated leadership in science policy are encouraged to apply.

Application deadline: January 28, 2013

www.aibs.org/public-policy/eppla.html

AIBS PUBLIC POLICY
ADVOCACY · COMMUNICATION · ANALYSIS