

# Willing or unwilling to share primary biodiversity data: results and implications of an international survey

Xiaolei Huang<sup>1,2</sup>, Bradford A. Hawkins<sup>2</sup>, Fumin Lei<sup>1</sup>, Gary L. Miller<sup>3</sup>, Colin Favret<sup>4</sup>, Ruiling Zhang<sup>1</sup>, & Gexia Qiao<sup>1</sup>

<sup>1</sup>Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China

<sup>2</sup>Department of Ecology & Evolutionary Biology, University of California, Irvine, CA 92697, USA

<sup>3</sup>USDA-ARS Systematic Entomology Laboratory, 10300 Baltimore Avenue, Beltsville, MD 20705, USA

<sup>4</sup>Université de Montréal Biodiversity Centre, 4101 rue Sherbrook est, Montréal QC H1×2B2, Canada

## Keywords

Biodiversity; conservation; data archiving; data sharing; database; journal policy; primary biodiversity data.

## Correspondence

Gexia Qiao, Xiaolei Huang, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China. Tel: 0086 10 64807133; fax: 0086 10 64807099. E-mail: qiaogx@ioz.ac.cn, huangxl@ioz.ac.cn

## Received

25 February 2012

## Accepted

25 May 2012

## Editor

Andrew Pullin

doi: 10.1111/j.1755-263X.2012.00259.x

## Abstract

Biodiversity science and conservation increasingly depend on the sharing and integration of large amounts of data, but many researchers resist sharing their primary biodiversity data. We recently conducted an international survey to ascertain the attitudes, experiences, and expectations regarding biodiversity data sharing and archiving of researchers. The results show that whereas most respondents are willing to share article-related biodiversity data, more than 60% of respondents are unwilling to share primary data before publishing. Results indicate an underdeveloped culture of data sharing and several major technological and operational barriers. A major concern for researchers is appropriate benefits from data sharing. Most respondents would accept data archiving policies of journals. Researchers also express concerns about how to easily and efficiently deal with data and data quality in public databases. Expectations for biodiversity databases include standardization of data format, user-friendly data submission tools, formats for different types of data, and coordination among databases. The survey results provide suggestions for improving data sharing and archiving by individual scientists, organizations, journals, and databases.

## Introduction

Biodiversity data such as species distributions are fundamental to biodiversity research, natural resource management, and conservation policy. Public biodiversity databases are of great potential benefit for both science and practice. Many already exist, but the global databases (e.g. The Global Biodiversity Information Facility [GBIF], <http://www.gbif.org/>; The Encyclopedia of Life [EoL], <http://www.eol.org/>) and the various regional or national initiatives often have different objectives. For example, EoL is aiming to publish “an electronic page for each species of organism on Earth” (Wilson 2003), and GBIF has been focusing on organismal occurrence data mainly from archived specimens in natural history collections (Edwards *et al.* 2000).

Sharing primary data is fundamental to the advancement of science as it is currently practiced. Although a stated objective of all existing biodiversity databases is to promote biodiversity awareness and the sharing of primary data, the extent to which this goal will be achieved depends on the willingness and practices of data providers and users. Biodiversity databases have aggregated a great deal of biodiversity data by using natural history collections as a major data resource (e.g., GBIF). However, it is important to remember that individual researchers from different kinds of organizations (e.g., academic, governmental, NGOs) as well as “citizen-scientists” are the people who collect and curate specimens, identify species, publish both results and data, and ultimately decide whether or not to share data. Therefore, the degree and success of biodiversity data sharing

relies on their attitudes and practices. Although there are increasing numbers of appeals, and data sharing has become more common (Parr & Cummings 2005; Reichman *et al.* 2011), many researchers actively or passively resist sharing primary data (Sedransk *et al.* 2010; Tenopir *et al.* 2011).

Recently, a group of key journals in evolution and ecology began implementing a joint data archiving policy to preserve article-related data (<http://datadryad.org/jdap>; Vision 2010; Whitlock 2011). With the aim of promoting biodiversity data sharing and archiving, Huang & Qiao (2011) suggested that adoption of joint data archiving policies by journals and databases would be a sustainable methodology and benefit both sides. However, to evaluate the utility of these procedures, it is important to ascertain the attitudes and expectations of the primary data gatherers and providers, the researchers themselves.

Data reuse is the fundamental goal of data sharing. Using interchangeable standard data structures and formats by biodiversity databases is beneficial to data exploration, comparison, and integration (Madin *et al.* 2007; Goddard *et al.* 2011; Thessen & Patterson 2011). However, biodiversity databases at global, regional, and national scales usually have different objectives and use different data structures. Beyond the role of data provider, individual scientists are also the main data users. More and more researchers would like to use databases for aims such as large-scale diversity patterns and policy making (e.g., Carpenter *et al.* 2009; Tittensor *et al.*, 2010; Drew 2011). Therefore, their expectations for the practicability and usability of biodiversity databases are key to database development. The development of the new discipline of biodiversity informatics as well as technical and infrastructural innovations should take the expectations of researchers seriously.

Most researchers would agree that a much better understanding of the existing data cultures is needed to promote data sharing in the life sciences (Thessen & Patterson 2011). Nevertheless, a survey aiming to study the attitudes, experiences and expectations about biodiversity data sharing and data archiving policies of researchers have yet to be undertaken. In this article, we report the results of a recent international survey on these issues. The results should provide an empirical view of issues in biodiversity data sharing and archiving and stimulate further discussion.

## Methods

A confidential and anonymous online questionnaire was designed using Qualtrics survey software (Qualtrics Labs Inc., Provo, UT, USA). The questionnaire consisted of three sections to evaluate (1) the respondents' demo-

graphics and research background, (2) their attitudes and experiences regarding biodiversity data sharing, and (3) their expectations regarding future data archiving practices. No questions about personally identifiable information were included. The questionnaire closed with an open-ended request (the 24th question) for general comments. The questionnaire is available in Table S1.

The survey was open from 17 July, 2011 to 10 September, 2011. An e-mail invitation including a survey link and background information of the survey was sent to active researchers in biodiversity, biogeography, and conservation. From 17 July to 10 August, we sent the invitation to corresponding authors of publications in 2009 and 2010 in three leading biodiversity and conservation journals, *Biodiversity and Conservation*, *Diversity and Distributions*, and *Journal of Biogeography*, in which articles are usually based on large amounts of primary biodiversity data, i.e., species data generated from collections and observations, such as species occurrences and distributions. From 10 August to 10 September, we sent the invitation letter to communication officers of appropriate scientific societies. The communication officers then helped us spread the invitation letter among their members by using email alerts, newsletters, web news, blog posts, Twitter, and Facebook.

We used a strict criterion to select responses to be included in the analysis. We considered responses with answers for at least three quarters of the questions, including the first and the 23rd questions, as valid. In all, 372 valid responses and 154 comments were evaluated. The full results and the comments are provided as online Supporting Information (Table S1; Text S1) for repurposing and reanalysis. For the sake of confidentiality, personally identifiable information in the comments of several respondents was deleted. The original data deposited in the Dryad repository: <http://dx.doi.org/10.5061/dryad.jr40f>, and have been openly archived on this webpage: <http://www.naturethinker.org/survey-results/>.

We conducted all statistical analyses using SPSS version 15.0 (SPSS, Chicago, IL, USA). Descriptive statistics were reported as percentages and numbers. Chi-square analyses using Monte Carlo exact calculation were conducted on geographical differences in attitudes to data sharing, culture of sharing, and expectations for journal data archiving policies of respondents from different continents.

## Results

### Respondents' background

About one-third (34.4%,  $n = 128$ ) of respondents were from Europe, 25.5% ( $n = 95$ ) from Oceania, 16.7% ( $n = 62$ ) from North America, 11.3% ( $n = 42$ ) from Asia,

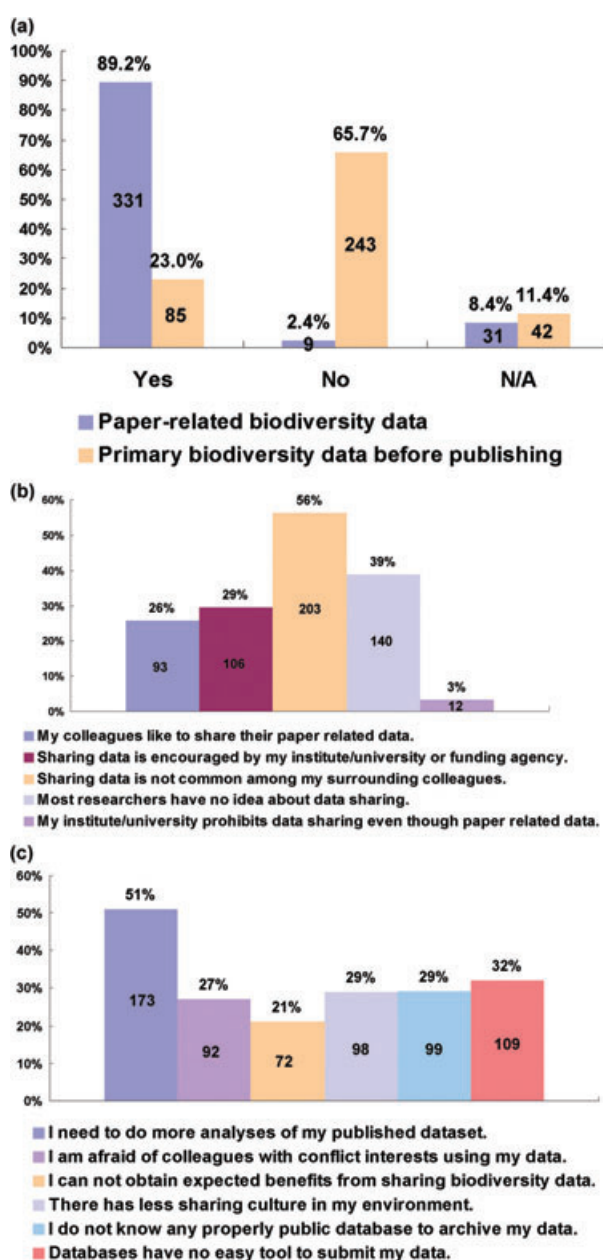
7.5% ( $n = 28$ ) from South America, and 4.6% ( $n = 17$ ) from Africa. As for the ages of respondents, 53.2% ( $n = 198$ ) were between 33- and 50-years old, 7.8% ( $n = 29$ ) between 18 and 27, 20.7% ( $n = 77$ ) between 28 and 32, 12.6% ( $n = 47$ ) between 51 and 60, and 5.6% ( $n = 21$ ) over 60. More than 85% of respondents were active researchers (e.g., Ph.D. students, post-docs, faculty). About 65% ( $n = 241$ ) of respondents were based in academic institutions and universities, 17% ( $n = 63$ ) in governmental organizations, and 17.9% ( $n = 66$ ) in nongovernmental organizations, intergovernmental agencies, and private companies.

A majority (56.6%,  $n = 210$ ) of respondents reported working on specific taxa, and 75.8% ( $n = 282$ ) conducted fieldwork to collect primary biodiversity data. For respondents who did not actively collect primary data via fieldwork (24.2%,  $n = 90$ ) as well as some who did, the published literature (65.1%,  $n = 123$ ), established databases (59.3%,  $n = 112$ ), and scientist colleagues (48.7%,  $n = 92$ ) were the main data sources for their research. For respondents who chose the “Other” option, all answers can be sorted into these three data sources. Over three quarters of respondents (77.8%,  $n = 253$ ) have more than 7 years research experience, and 65% ( $n = 211$ ) have more than 10.

### Scientists' attitudes and practices toward sharing biodiversity data

More than 90% (91.8%,  $n = 338$ ) of respondents agreed the sharing of biodiversity data is very important, 7.6% ( $n = 28$ ) thought it of some importance, and two respondents thought it unimportant. Over 80% (84.3%,  $n = 311$ ) of respondents agreed sharing article-related data is a basic responsibility, whereas 11.1% ( $n = 41$ ) disagreed. A strong majority of respondents would be willing to share article-related data, but almost two-thirds would prefer not to share before publication (Figure 1a). The respondents from different geographical regions were in agreement ( $\chi^2$ ; all  $P > 0.85$ ; Table 1) with regard to these questions. There was no relationship between respondents' age and willingness to data sharing (Table S2). However, a lower willingness to share article-related data and a greater unwillingness to share primary data before publishing in some groups with more working years was indicated to some extent (Table S3).

A majority of respondents reported having experiences of sharing data. Approximately 85% have always (10.1%,  $n = 37$ ), often (22.4%,  $n = 82$ ), or sometimes (52.7%,  $n = 193$ ) shared article-related data. The most frequent data archiving approach was through files supplementary to articles (51.5%,  $n = 169$ ), followed by public databases (38.1%,  $n = 125$ ), institute/university



**Figure 1** Respondents' attitudes toward biodiversity data sharing. (a) Their willingness to share article-related data and primary data before publishing. (b) A weak culture of data sharing within scientific community. (c) Some main obstacles to data sharing.

websites (25%,  $n = 82$ ), or personal websites (12.8%,  $n = 42$ ). Many researchers also shared article-related data with colleagues through e-mail.

The respondents claimed a weak culture of sharing within their scientific community (Figure 1b). African respondents reported the weakest climate of data sharing: they constituted the lowest percentage (5.9%,  $n = 1$ )

**Table 1** Geographic agreement of some attitudes toward biodiversity data sharing

Region	Sharing biodiversity data is very important % (n)	Sharing article-related data is a basic responsibility % (n)	Willing to Share article-related data % (n)	Unwilling to share before publishing % (n)
Africa	100 (17)	94.1 (16)	94.1 (16)	76.5 (13)
Asia	95.2 (40)	90.5 (38)	90.5 (38)	69.0 (29)
Europe	89.7 (113)	77.3 (99)	90.6 (116)	64.6 (82)
North America	93.5 (58)	85.2 (52)	90.2 (55)	67.7 (42)
Oceania	88.3 (83)	88.3 (83)	82.1 (78)	62.8 (59)
South America	100 (27)	85.2 (23)	100 (28)	64.3 (18)
	$\chi^2 = 1.31$ ; $P = 0.94$	$\chi^2 = 2.00$ ; $P = 0.85$	$\chi^2 = 1.87$ ; $P = 0.87$	$\chi^2 = 1.94$ ; $P = 0.86$

Note: Percentages based on total number of respondents in each region.

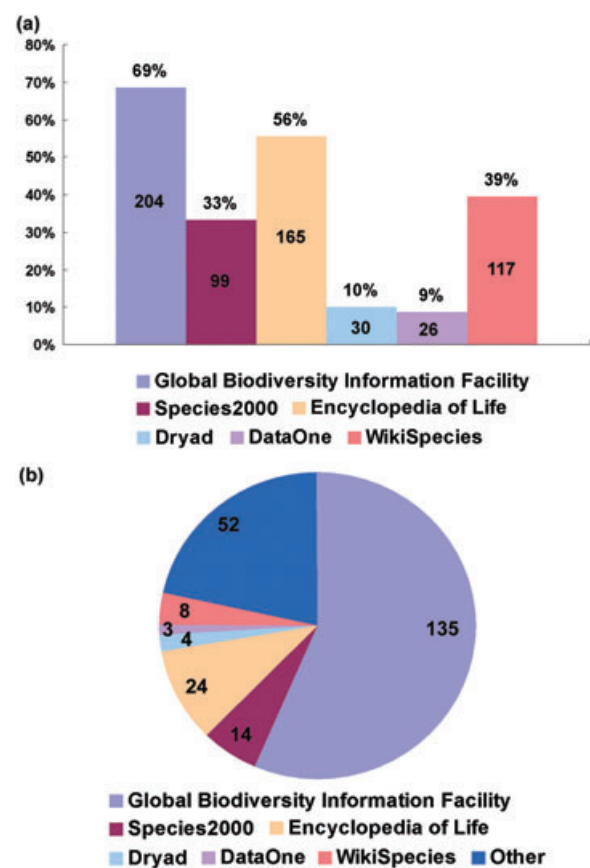
of respondents who thought their colleagues wished to share data ( $\chi^2 = 18.51$ ;  $P = 0.002$ ) and the highest percentage (64.7%,  $n = 11$ ) who thought their colleagues were ignorant about data sharing ( $\chi^2 = 19.65$ ;  $P = 0.002$ ). Europe ranked lowest for encouragement of data sharing by employers and funders (19.4%,  $n = 24$ ;  $\chi^2 = 12.49$ ;  $P = 0.03$ ). However, the respondents whose affiliations and funding agencies encourage data sharing were more willing to share (Table S4).

Approximately half of the respondents preferred keeping primary data private beyond their first publication to conduct other analyses. A similar number of respondents indicated other obstacles to biodiversity data sharing, including conflicts of interests with colleagues (27.1%), lack of benefits (21.2%), weak sharing culture (28.8%), unfamiliarity with public databases (29.1%), and lack of user-friendly data submission tools (32.1%; Figure 1c). Many respondents voiced further concerns in their comments. For example, about 40% of the comments expressed a lack of time and funding to properly format, input, update, and otherwise manage their publicly available data.

When asked what benefits they expected from sharing data, nearly 90% (88.5%,  $n = 322$ ) of respondents indicated the desire to contribute to scientific progress. More than one-third wanted credit (37.1%,  $n = 135$ ) and higher citation rates (33.5%,  $n = 122$ ). In the comments, a number of respondents stated that manuscript coauthorship should be offered to researchers whose shared primary data constitute a significant part of an analysis.

### Scientists' expectations toward archiving biodiversity data

Of the respondents who answered the question about established databases, over half knew of GBIF and EoL, more than one-third knew of WikiSpecies (<http://species.wikimedia.org/>) and Species2000 (<http://www.sp2000.org/>), and relatively few knew of Dryad (<http://datadryad.org/>) and DataOne (<https://www.dataone.org/>; Figure 2a). Over three quarters (76.6%,  $n = 281$ ) of respondents agreed that a small number of unified biodiversity databases using standard data format are needed. More than half (56.3%) thought GBIF best



**Figure 2** (a) Respondents' familiarity with certain established databases. (b) Respondents' choice of the database best exemplifying the ideal for archiving species richness and occurrence data.

org/), and relatively few knew of Dryad (<http://datadryad.org/>) and DataOne (<https://www.dataone.org/>; Figure 2a). Over three quarters (76.6%,  $n = 281$ ) of respondents agreed that a small number of unified biodiversity databases using standard data format are needed. More than half (56.3%) thought GBIF best



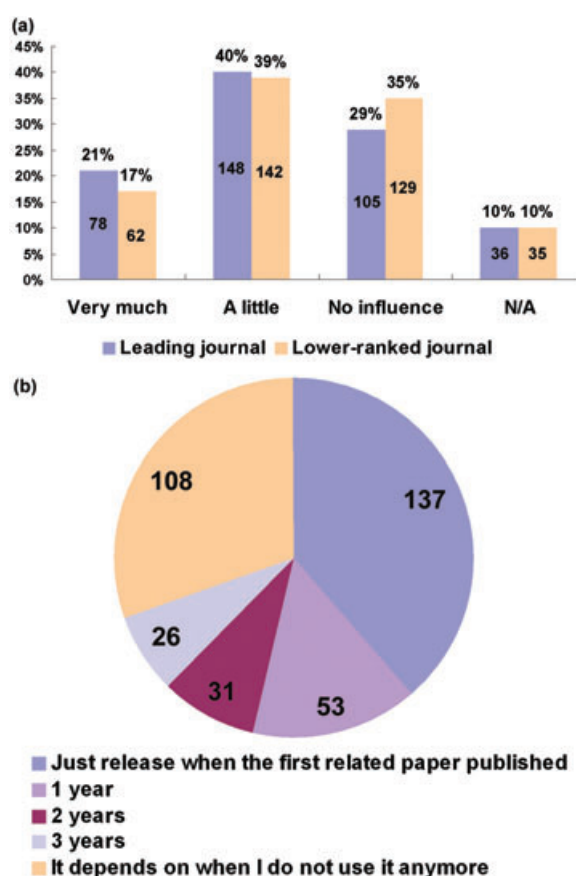
exemplified the ideal for archiving species richness and occurrence data (Figure 2b). About 22% of respondents indicated an "Other" database, unnamed and perhaps not yet built (Figure 2b). Some thought improved data formats were needed, for example for storing plot-based records. Some thought regional and/or national databases would be more effective and accurate than global databases. Another voiced opinion was that databases for different types of data were needed to avoid oversimplifying biodiversity data to a few common metrics.

Respondents were asked whether a journal's data archiving policy would influence their manuscript submission. Approximately one-fifth (21.3%) indicated that such a policy by a leading journal would influence them "very much," whereas such a policy by a lower ranked journal would influence a smaller proportion of respondents "very much" (16.8%). A majority of respondents (at least 68.9% and 73.7% for leading and lower ranked journals) would accept journals' data archiving policies (Figure 3a). When asked how long they would like to keep private their article-related data, 30% (30.4%) of respondents indicated they would prefer to keep their data private until they were completely finished with them. However, more respondents indicated that they would be willing to make their data public as soon as related articles were published (38.6%) or after 1–3 years (Figure 3b).

## Discussion

Individual scientists are both data providers and data users during the research process. It is therefore critical that we understand what they are thinking and doing about sharing and archiving primary biodiversity data. To our knowledge, the results of this survey provide a first empirical view of these issues. The original data files created by this survey are also important for reuse.

Echoing increasing appeals (Parr & Cummings 2005; Reichman *et al.* 2011), almost all respondents understand the importance of sharing data and most are willing to share article-related biodiversity data. However, more than 60% of respondents are unwilling to share their primary data before publishing. Respondents also reported a weak culture of sharing within their research community. There seems to be a contradiction between two views: research data gathered with public funding should be made public, but the data gatherers deserve more benefits and better recognition for doing so. A major obstacle to sharing data revealed by our survey is researchers' need for further analyses of their data. The other main obstacles include conflicts of interest with colleagues, lack of benefits, unfamiliarity with public databases, lack of user-friendly data submission tools, and insufficient time



**Figure 3** Respondents' expectations toward data archiving policy. (a) Influence of journal data archiving policy on manuscript submission. (b) The length of time respondents prefer to keep their article-related data private.

and funding. Some of these are common obstacles that other disciplines also face, as reported by a previous survey (Tenopir *et al.* 2011). In fact, barriers such as data retention for further analyses and conflicts of interest are directly related to the perceived benefits to the data sharers.

The present survey indicates that incentives such as receiving credit, coauthorship, and higher citation rates would encourage data sharing. The sharing of detailed research data may be associated with increased citation rates (Piwowar *et al.* 2007), but some respondents suggested data sharing be given greater weight in the assessment system for the professional productivity of researchers (see also Kueffer *et al.* 2011). As a possibility, statistics on the reuse or citation of publicly available data sets should be treated as are article citations. Recently, the development of a Data Usage Index has been proposed for improving the professional recognition in the biodiversity community (Ingwersen & Chavan 2011). In addition, we think data registry methods, for example an interoperable

dataset identifier, can resolve data ownership or intellectual property rights and citation issues.

Proper archiving methods and policies are necessary conditions to achieve the fundamental goal of sharing data. The survey results indicate a majority of respondents would accept data archiving policies by biodiversity and conservation journals (Huang & Qiao 2011; Whitlock 2011). They would prefer to make primary data public as soon as related articles are published or a short period of time thereafter. Many respondents consider that sharing the primary data supporting a publication is a simple ethical principle in science. There is also concern about the credibility of a article if the authors are reluctant to share their data. Although data archiving is routine for journals in some fields such as genetics, genomics, and biomedicine, many researchers still fail to publicly archive their data after publication (Noor *et al.* 2006; Alsheikh-Ali *et al.* 2011). This has implications for the future practices of joint data archiving policies and working flow by journals and databases in biodiversity science. On the one hand, journals and databases should adopt policies that can ensure appropriate benefits to the authors or data providers. On the other hand, to help authors adopt data archiving practices, journals should insist on more rigorous policies; for example, journals could require authors to guarantee (e.g., sign an agreement) the submission of manuscript-related data before final publication of their articles. Journals and databases would also gain benefits, such as quality control of data sets and articles (Huang & Qiao 2011) as well as increased impact because of higher citation rates (Piwowar *et al.* 2007), from such policies.

Only one-third of respondents reported that sharing data was encouraged by their employers or funding agencies. However, the respondents whose affiliations and funding agencies encourage data sharing were more willing to share (Table S4). This has important implications for future policies and practices of affiliations and funding agencies. They can provide detailed instructions or policies about data management. They can also improve their assessment systems to give improved recognition for data sharing. Beginning January 2011, the U.S. National Science Foundation (NSF) requires all grant proposals to include a data management plan to disseminate and share research results and primary data (NSF 2011). This is a good example of how other organizations can promote data sharing. We think the guiding role of organizations would be especially important in regions with a weak climate of data sharing.

It should be noted that only major public databases with global orientation and without restriction to specific taxa were included in the questionnaire. However, the respondents provided valuable comments beyond this list. The survey results indicate that biodiver-

sity databases should promote themselves and help improve researchers' awareness of data sharing. Researchers expressed concerns about how to deal with their data easily and efficiently. The expectations for biodiversity databases included using standard formats, user-friendly submission tools, proper storage mechanisms for different types of biodiversity data, and collaboration between databases employing interchangeable data structures. Although many researchers agreed that a small number of unified databases are needed to effectively preserve biodiversity data, some also warn of oversimplifying data that is inherently complex. Database developers should address the concerns of data providers and data users to improve the usability and universality of biodiversity databases.

In their comments, respondents also expressed concerns about data quality and reliability in public databases, taxonomic accuracy perhaps being the most important. Biodiversity databases should devise ways to resolve this problem (Page 2008; Patterson *et al.* 2010) and improve data management. The quality of data from journal articles can be partially guaranteed by rigorous data archiving policies. Theoretically, the quality of data from natural history collections and voucher specimens can be improved by continuous update by qualified taxonomic staff. However, insofar as the "taxonomic impediment" is a challenge to taxonomy as well as biodiversity science (Ebach *et al.* 2011). Strict joint data archiving policies by journals and databases and certain data standards would also help control data quality and preserve detailed metadata descriptions. Long-term maintenance of databases is another concern, which suggests that funding agencies and databases should find sustainable methods together (see also Thomas 2009). For example, the proliferation of databases with narrow research agenda or even personal ambitions perhaps should be restricted to avoid resource waste.

We should acknowledge that it is hard for our survey to include all aspects of important issues. Instead of trying to include all, we hope our work can stimulate more discussion and further survey researches. If the biodiversity research community (individual scientists, organizations, journals, databases) were to address the related issues more diligently, major barriers preventing wide biodiversity data sharing may be overcome and a scientific culture of sharing and collaboration would also be cultivated and improved.

## Acknowledgments

We thank the many colleagues who took the time to complete the survey and who provided a wealth of

valuable comments. We also express our gratitude to the societies and colleagues who helped publicize the survey, including but not limited to the Canadian Society for Ecology and Evolution (Sarah Otto, Veronique Connolly); Ecological Society of Australia; Ecological Society of Chile (Paulina Chacón); European Ecological Federation (Ceri Margerison); International Biogeography Society (Karen Faller, Mike Dawson); Israel Society of Ecology & Environmental Sciences (Marcelo Sternberg); New Zealand Ecological Society (Laura Young); Africa Section (Beth Kaplin), Asia Section, North America Section and Oceania Section of Society for Conservation Biology; Spanish Ecological Society (David Nogués-Bravo); and Shen-Horn Yen at National Sun Yat-Sen University (Taiwan). We thank Dr. Ashwini Chhatre, Dr. Juliette Young, Dr. Andrew Pullin and two anonymous reviewers for their valuable comments and constructive suggestions. Thank Xiaolan Lin for her support in design and implementation of the survey. Portions of the analyses were done while the first author was a visiting scholar at USDA, Systematic Entomology Laboratory. XLH is funded by NSFC-30900133, GXQ by NSFC-31025024, 31061160186 and 30830017, and FML by NSFC-30925008.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA; USDA is an equal opportunity provider and employer.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1:** The questionnaire and total results statistics.

**Table S2:** Respondents' ages and their attitudes toward data sharing.

**Table S3:** Respondents' working years in related fields and their attitudes toward data sharing.

**Table S4:** Encouragement and respondents' sharing culture.

**Text S1:** The 154 comments for the open-ended question.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## References

Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, M.H. & Ioannidis, J.P.A. (2011). Public availability of published

research data in high-impact journals. *PLoS One*, **6**, e24357. doi:10.1371/journal.pone.0024357.

Carpenter, S.R., Armbrust, E.V., Arzberger, P.W. *et al.* (2009). Accelerate synthesis in ecology and environmental sciences. *Bioscience*, **59**, 699–701.

Drew, J. (2011). The role of natural history institutions and bioinformatics in conservation biology. *Conserv. Biol.*, **25**, 1250–1252.

Ebach, M.C., Valdecasas, A.G. & Wheeler, Q.D. (2011). Impediments to taxonomy and users of taxonomy: accessibility and impact evaluation. *Cladistics*, **27**, 550–557.

Edwards, J.L., Lane, M.A. & Nielsen, E.S. (2000). Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, **289**, 2312–2314.

Goddard, A., Wilson, N., Cryer, P. & Yamashita, G. (2011). Data hosting infrastructure for primary biodiversity data. *BMC Bioinform.*, **12** (Suppl 15), S5. doi:10.1186/1471-2105-12-S15-S5.

Huang, X.L. & Qiao, G.X. (2011). Biodiversity databases should gain support from journals. *Trends Ecol. Evol.*, **26**, 377–378.

Ingwersen, P. & Chavan, V. (2011). Indicators for the Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure. *BMC Bioinform.*, **12** (Suppl 15), S3. doi:10.1186/1471-2105-12-S15-S3.

Kueffer, C., Niinemets, U., Drenovsky, R.E. *et al.* (2011). Fame, glory and neglect in meta-analyses. *Trends Ecol. Evol.*, **26**, 493–494.

Madin, J., Bowers, S., Schildhauer, M. *et al.* (2007). An ontology for describing and synthesizing ecological observation data. *Ecol. Inform.*, **2**, 279–296.

NSF (National Science Foundation). (2011). NSF 11–1. Grant Proposal Guide. <http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/>. (visited Dec. 26, 2011.)

Noor, M.A.F., Zimmerman, K.J. & Teeter, K.C. (2006). Data sharing: how much doesn't get submitted to GenBank? *PLoS Biology*, **4**, e228. doi:10.1371/journal.pbio.0040228.

Page, R.D.M. (2008). Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Brief Bioinform.*, **9**, 345–354.

Parr, C.S. & Cummings, M.P. (2005). Data sharing in ecology and evolution. *Trends Ecol. Evol.*, **20**, 362–363.

Patterson, D.J., Cooper, J., Kirk, P.M., Pyle, R.L. & Remsen, D.P. (2010). Names are key to the big new biology. *Trends Ecol. Evol.*, **25**, 686–691.

Piwowar, H.A., Day, R.S. & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, **3**, e308. doi:10.1371/journal.pone.0000308

Reichman, O.J., Jones, M.B. & Schildhauer, M.P. (2011). Challenges and opportunities of open data in Ecology. *Science*, **331**, 703–705.

Sedransk, N., Young, L.J., Kelner, K.L. *et al.* (2010). Make research data public?—not always so simple: a dialogue

- for statisticians and science editors. *Stat. Sci.*, **25**, 41-50.
- Tenopir, C., Allard, S., Douglass, K. *et al.* (2011) Data sharing by scientists: practices and perceptions. *PLoS One* **6**, e21101. doi:10.1371/journal.pone.0021101.
- Thessen, A.E. & Patterson, D.J. (2011). Data issues in the life sciences. *Zookeys*, **150**, 15-51.
- Thomas, C. (2009). Biodiversity databases spread, prompting unification call. *Science*, **324**, 1632-1633.
- Tittensor, D.P., Mora, C., Jetz, W. *et al.* (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, **466**, 1098-1101.
- Vision, T.J. (2010). Open data and the social contract of scientific publishing. *Bioscience*, **60**, 330-331.
- Whitlock, M.C. (2011). Data archiving in ecology and evolution: best practices. *Trends Ecol. Evol.*, **26**, 61-65.
- Wilson, E.O. (2003). The encyclopedia of life. *Trends Ecol. Evol.*, **18**, 77-80.