

生物多样性数据共享和发表：进展和建议

黄晓磊* 乔格侠*

(中国科学院动物研究所动物进化与系统学重点实验室, 北京 100101)

摘要: 生物多样性研究、保护实践、自然资源管理及科学决策等越来越依赖于大量数据的共享和整合。虽然关于数据共享的呼吁和实践越来越多, 但很多科学家仍然主动或被动地拒绝共享数据。关于数据共享, 现实中存在一些认知和技术上的障碍, 比如科学家不愿意共享数据, 担心同行竞争, 认为缺少足够的回报, 不熟悉相关数据保存机构, 缺少简便的数据提交工具, 没有足够时间和经费等。解决这些问题及改善共享文化的关键在于使共享者获得适当的回报(比如数据引用)。基于同行评审的数据发表被认为不但能够为生产、管理和共享数据的科学家提供一种激励机制, 并且能够有效地促进数据再利用。因而, 数据发表作为数据共享的方式之一, 近来引起了较多关注, 在生物多样性领域出现了专门发表数据论文的期刊。在采取数据论文的模式上, 数据保存机构和科技期刊采用联合数据政策在促进数据共享方面可能更具可行性。本文总结了数据共享和发表方面的进展, 讨论了数据论文能在何种程度上促进数据共享, 以及数据共享和数据发表的关系等问题, 提出如下建议: (1)个体科学家应努力践行数据共享; (2)使用DOI号解决数据所有权和数据引用的问题; (3)科技期刊和数据保存机构联合采用更加合理和严格的数据保存政策; (4)资助机构和研究单位应当在数据共享中起到更重要的作用。

关键词: 数据发表, 数据论文, 数据期刊, 科学政策, 可重复性, 生态学, 环境保护

Sharing and publishing of biodiversity data: recent trends and future suggestions

Xiaolei Huang*, Gexia Qiao*

Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101

Abstract: Biodiversity research, conservation practices, natural resource management, and scientific decision-making increasingly depend on the sharing and integration of large amounts of primary data. In recent years, there has been an appeal increased sharing of biodiversity data, however, many scientists actively or passively resist sharing data. Some major cultural and technological obstacles exist among scientists, such as keeping data private to conduct other analyses, conflicts of interests with colleagues, lack of benefits, unfamiliarity with public databases, lack of user-friendly data submission tools, and lack of time and funding. One solution to improve the culture of data sharing is to provide benefits to scientists who share data (e.g. data citations). Recently, some organizations and scientists have advocated data publishing under peer review as a reward mechanism for individuals involved in data creation, management and sharing, and as a way to effectively increase the use and reuse of data. New data journals have been launched to fulfill the function of publishing data. In fact, besides the advocate of scholarly publication of data, an improved joint data archiving policy by databases and scientific journals may be more practically feasible to improve data sharing in a broader sense. In this article we review recent progress in data sharing and publishing and discuss to what extent data papers can boost data sharing and how to define 'data sharing' and 'data publishing'. We also provide suggestions for improving data sharing by individual scientists, data repositories, journals, and funding agencies/institutions.

收稿日期: 2013-12-30; 接受日期: 2014-05-12

基金项目: 国家自然科学基金面上项目(31272348)、中国科学院动物进化与系统学重点实验室开放课题(Y229YX5105)、中国科学院战略性先导科技专项(XDA05080703)和国家科技基础条件平台——动物标本资源共享平台项目

* 共同通讯作者 Co-authors for correspondence. E-mail: huangxl@ioz.ac.cn; qiaogx@ioz.ac.cn

Key words: data publishing, data paper, data journal, science policy, science reproducibility, ecology, environmental conservation

科学数据共享正越来越引起广泛的关注。数据共享不仅对于科学发展和大科学(Big Science)研究模式至关重要,也是各国政府进行科学决策的重要基础。近期*Nature*和*Science*杂志都曾发表社论和出版专刊来讨论相关问题(<http://www.nature.com/news/specials/datasharing/index.html>; <http://www.Sciencemag.org/site/special/data>),众多科学家也不断呼吁实施数据共享并给出了各自的建议(Vision, 2010; Whitlock, 2011; Reichman *et al.*, 2011; Hampton *et al.*, 2013)。从政府层面推进数据共享的一个重要例子是,美国白宫科学和技术政策办公室以及总统执行办公室于2012年3月宣布启动“大数据(Big Data)计划”,6家联邦机构首期资助两亿多美元用于科学数据共享的模式、工具和技术的研发;进而在2013年5月发布行政命令,要求联邦资金支持的研究项目所获得的数据向社会公开。

生物多样性和生态学研究往往需要长期的野外和实验室观测,来积累大量的数据。然而,这些散布在个体科学家实验室中的数据缺乏有效的整合。在21世纪的今天,生物多样性的流失依然严重,生物多样性科学研究和保护都面临着新的挑战,并亟需国际社会的关注。众多学科领域(尤其体现在高能物理和基因组学领域)大科学时代的来临以及人类所面临的环境问题,对生物多样性和生态学也提出了新的要求,科学家不但要通过大尺度研究来理解众多全球性问题(如生物多样性丧失、气候变化、粮食安全等),而且要基于科学证据提出相应的对策(Michener & Jones, 2012; Hampton *et al.*, 2013)。因此,如何有效地共享及整合生物多样性数据,是生物多样性领域在信息时代面临的一个重要挑战。这不但是提升生物多样性研究深度和广度以及开展整合性研究的需要,而且是科学制定生物多样性保护政策的重要基础,需要科学家和决策部门的重视并大力推进。

我们总结了生物多样性数据共享和发表方面的主要进展,讨论了数据论文(data paper)和数据期刊(data journal)在数据共享方面的作用以及数据共享与发表的关系等问题,并针对数据共享的不同利益方(科学家个体、数据保存机构、科技期刊、资助

和研究机构等)提出了一些建议。在国际核苷酸序列数据库联盟的3个数据库即GenBank、DNA Data Bank of Japan (DDBJ)、European Nucleotide Archive (ENA)以及学术界的共同推动下,遗传学数据尤其是序列数据(遗传多样性数据)的共享已经形成规范,因此本文所谈及的生物多样性数据更多地侧重于物种多样性和生态系统多样性的相关数据。

1 生物多样性数据共享的重要性

生物多样性数据共享的重要性至少体现在三个方面。首先,相关领域的基础科学研究需要数据共享。比如,生物多样性研究的核心问题之一是构建生命之树(tree of life),而构建巨大的生命之树需要众多“枝叶”数据的积累及整合(Stoltzfus *et al.*, 2012);此外,全球生物多样性的格局研究及物种分布的监测,也需要不同科学家及时共享数据。第二,制定切实可行的生物多样性和环境保护政策,需要对科学证据及原始数据开展系统的分析研究(Pullin & Salafsky, 2010; Haddaway & Pullin, 2013)。这在政策水平相对落后的发展中国家(如中国)显得尤为重要。第三,科学研究的可重复性(science reproducibility)一直都是学术界关注的重要问题(Cassey & Blackburn, 2006; Giles, 2006),而原始数据共享是检验研究结果可重复性最重要的前提;再加上野外研究很多时候难以实现真正的重复,以及目前很多原始数据并未共享,原始数据的共享显得更为重要(Ryan, 2011; Shapiro & Báldi, 2012)。

生物多样性数据共享的实现需要以下几方面的有效配合(图1):首先是数据生产者,主要包括职业科学家和公众科学参与者(非职业研究人员或收集生物多样性数据的爱好者)。其次是生物多样性数据保存机构,这是实现数据共享的重要支撑。它们的运行模式、共享策略、工具等在很大程度上决定了数据共享的效果。第三是资助和研究单位,它们为数据生产和共享提供经费支持,并在相关政策制定和导向方面也对数据共享效果有重要影响。此外,基于生物多样性数据的研究成果大多以论文的形式发表,因此作为论文载体的期刊在数据共享中也有重要作用。

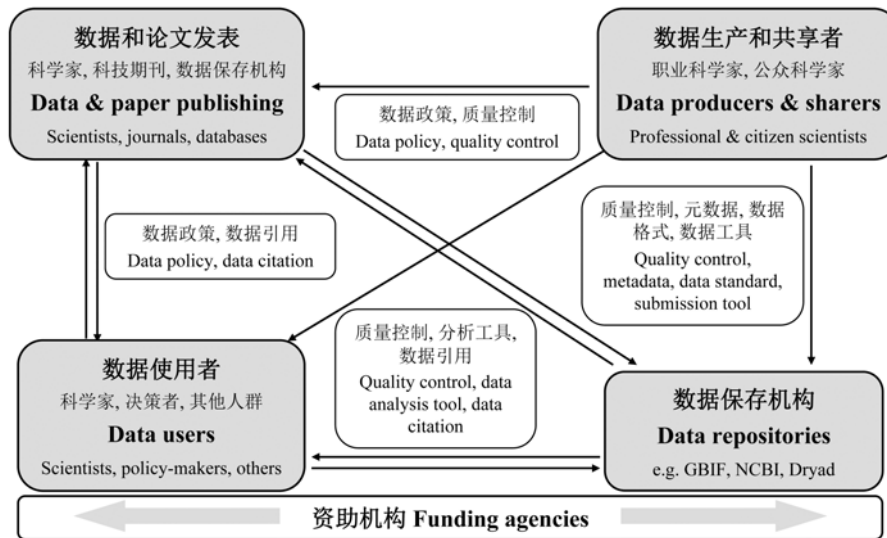


图1 数据共享流程简图, 包括主要参与方和决定共享效果的主要因素。资助机构在数据共享的不同环节都可能起作用。

Fig. 1 A sketch of the data sharing flow, including main participants and determining factors in data sharing. Funding agency spans each stage of the data flow.

2 数据共享现状: 让理想变为现实

2.1 数据共享的诉求

近年来从政府到资助和研究机构, 再到数据保存机构及期刊, 关于数据共享的呼声越来越高, 已经开始制定并采取有效的数据共享政策和措施。相对突出的例子如, 2009年5月, 美国政府启动了国家数据网(www.data.gov); 同年9月, 中国科技部也启动了“中国科技资源共享网”(www.escience.gov.cn)平台, 都旨在促进数据资源的公开。从2011年1月起, 美国国家科学基金会明确要求项目申请人在其申请书中增加“数据管理计划”(Data Management Plan), 明确数据保存和共享的承诺与措施(National Science Foundation, 2011); 美国国立卫生研究院、美国科学院、美国国家研究委员会也有类似的倡议和政策(National Research Council, 2003)。英国生物技术和生物科学研究委员会于2010年6月发布了详细的数据共享政策(BBSRC, 2010); 英国国家环境研究委员会、Wellcome Trust等也签署了联合声明, 承诺促进研究数据共享(van den Eynden *et al.*, 2011)。

科技期刊如*Nature*、*Science*, 以及一些成熟的开放获取期刊如BioMed Central和PLoS系列期刊, 都明确给出了各自数据保存和共享的要求。近期,

生态学和进化生物学领域一些主流期刊开始与美国自然科学基金会支持的数据保存机构Dryad (<http://datadryad.org>)联合, 共同致力于保存已发表论文的相关数据(Whitlock, 2011)。Dryad目前主要侧重于生态学和进化生物学领域。虽然这种模式刚刚起步且需要更多的推广, 但毫无疑问是期刊数据保存政策近期发展的一个重要进程。

一些数据保存机构, 比如采用自上而下运作模式(数据节点式管理, 未直接面向个人用户)的全球生物多样性信息网络(Global Biodiversity Information Facility, GBIF; www.gbif.org)和地球观测数据网络(Data Observation Network for Earth, DataONE; www.dataone.org)也在大力呼吁数据共享。

然而, 数据共享的诉求和现实效果并不一定成正相关。比如, 由于作者和期刊在执行数据共享要求时并不严格, 大量应该共享的论文数据并没有被合理共享(Yesson *et al.*, 2007; Alsheikh-Ali *et al.*, 2011)。Vines等(2013)跟踪调查了1991–2011年间所发表的516篇论文, 这些论文都涉及到植物或动物的形态数据并开展了判别分析, 结果发现这些论文数据的可获得性受文章发表时间的严重影响, 以每年17%的速度快速递减。这项调查也暗示出个体科学家并不能可靠地保管数据, 需要完善的公共数据库和数据保存政策来推进数据共享。实际上, 作为

最主要的数据生产者、分享者和使用者,科学家群体的态度和行为在很大程度上决定了数据共享的效果。

2.2 科学家群体的行为和共享文化

数据共享的道理其实很简单:受到公共资金支持的研究,理应将其所有研究成果和相关数据共享给社会。但是,事实上很多科学家主动或被动地拒绝共享数据(Tenopir *et al.*, 2011; Huang *et al.*, 2012),主要是担心自己的竞争力受到影响,这种观念在科学发展相对落后的地区更常见(Huang *et al.*, 2012)。而科学家及非职业研究人员或爱好者不仅是采集数据、整理数据、发表数据的主要群体,还是数据共享和数据使用者,他们的态度、行为和期望将是数据共享能否有效推进以及塑造数据共享文化的决定力量,故应该在科学决策过程中充分考虑。

了解生物多样性领域科学家群体的共享文化对于更好地推进生物多样性数据共享非常重要(Thessen & Patterson, 2011)。Huang等(2012)针对生物多样性相关领域科学家的调查研究发现,大部分科学家(>90%)都认为数据共享非常重要,并愿意分享已发表研究论文的相关数据,但超过65%的科学家不愿意共享未发表的原始数据。该调查结果揭示出生物多样性领域的科学家面临一些认知和技术上的障碍,包括不愿共享数据以便进行后续分析,担心同行竞争,认为缺少足够的回报,不熟悉相关数据保存机构,缺少简便的数据提交工具,没有足够的时间和经费来管理数据,等等。其中一些共享障碍在Tenopir等(2011)针对不同学科领域科学家的调查研究中也有发现。Huang等(2012)的调查还显示,科学家希望从数据共享中获得相应的回报或利益,比如数据所有权、更多的论文引用、共同作者身份等。很多被调查者指出,数据共享应该在科研评价体系中获得足够的认可,数据集的重新利用或引用应该像论文引用一样对待。科学家对生物多样性数据保存机构的期望包括采用标准的数据格式、提供简易的数据提交工具、提高数据质量、数据库间的交互等。

上述调查结果暗示,数据共享要想获得更多科学家的支持,需要做的事情还很多。比如,资助机构和研究机构应该更多地鼓励数据共享并提供详细的共享指南,数据库和期刊应采用更加合理和严

格的数据保存政策,数据库应更多地考虑数据生产者和使用者的期望,提高数据库的可操作性和数据质量。实际上,很多共享障碍跟科学家能否从共享中获得足够的回报有关。

2.3 中国生物多样性数据共享现状

我国的生物多样性数据共享主要体现在以政府部门为主导的物种数据库及共享平台建设方面。我国从20世纪80年代末开始物种数据库建设。21世纪初,中国科学院生物标本馆网络信息系统建设项目启动(马克平等, 2010),并得到科技部“标本资源的标准化整理、整合与共享平台建设项目”的后续支持,包括动物、植物、微生物等标本资源的国家标本资源共享平台(NSII, www.nsii.org.cn)于近期建成。该平台包括6个子平台,参建单位达137个,基本涵盖我国各类标本资源及主要的标本资源收藏机构,收录1,000多万份各类标本及相关的名录、文献和照片等信息(许哲平等, 2014)。一些公众科学平台近年来也取得很大发展,比如2007年成立的中国自然标本馆(CFH, <http://www.cfh.ac.cn/>)已建立了生物多样性基础信息共享、野外调查、物种鉴定、数据管理等系列功能体系,并开始在生物多样性调查监测方面发挥重要作用。

然而,与数据共享文化相对发达的美国、英国等科学强国相比,我国生物多样性数据共享仍有很多工作要做。我国一直以来依靠自上而下的任务来推进数据共享,长远来看这种模式是不可持续的。如何提高科学家群体的共享意识并促进其共享行为,形成自下而上的、自觉的数据共享模式是未来工作的关键。此外,我国公共生物多样性数据库的数据质量和更新周期也需要提高。

3 新的数据共享策略

要推进数据共享的发展,需要增强生物多样性数据所有利益方(包括科学家、研究和资助机构、期刊等)对于数据共享价值的认识,也需要所有利益方积极采用有效的共享策略。近来提出的一些共享策略大多关注了如何让共享数据的科学家获得相应的收益。

3.1 集中保存论文数据

2011年,美国国家科学基金会支持的数据保存机构Dryad启动,与一些主流的生态学和进化生物学期刊(包括*The American Naturalist*、*Ecology*、

Evolution、*Heredity*、*Biological Journal of the Linnean Society*等)签署合作协议,在期刊的投稿指南中鼓励作者在投稿时将论文数据提交至Dryad保存并共享。*Nature*、*Science*、*PLoS*系列期刊等也都先后与Dryad达成合作意向。

科学家将数据提交到Dryad时,每个数据集会被分配一个DOI (Digital Object Identifier),使得这些数据集能够被引用和跟踪。在学术论文正式发表时,作者可以在论文材料和方法部分链接相应数据集地址,或以参考文献的形式引用。当一个数据集被引用时,Dryad建议使用者同时引用原始论文和数据集。该策略的目的是尽可能地实现论文(目前主要限于生态和进化领域)相关数据的保存和共享,并使得数据集可被引用,从而鼓励科学家更多地共享数据。这一模式已收到一些成效。截至2013年12月25日,Dryad已经保存了来自285个期刊、5,145个数据集的13,439个数据文件,这些文件已经被下载751,658次。

另一个值得提及的数据保存机构是同样于2011年启动的figshare(<http://figshare.com/>),它的优势在于不仅可以保存论文相关数据,还保存研究过程中产生的其他数据、影像、文本等资料。

为了尽可能多地保存多元化的数据,Dryad和figshare对数据集都没有格式方面的要求。这虽然有助于保存多样化数据,然而,就某个学科来说,统一的数据格式对于数据管理和基于多数据集的整合性研究很有必要。我们建议不同类型的数据,比如物种地理分布数据(经纬度)和基于样地的长期生态观察数据,应使用各自统一的数据标准。

3.2 数据发表: 数据论文和数据期刊

2012年2月GBIF提议推进基于同行评审的生物多样性数据发表(data publishing)。这是GBIF和出版商Pensoft在2011年所提议的生物多样性数据论文的延续,提议“用于网络共享的数据集的元数据文件可以采用学术论文的形式发表”(Chavan & Penev, 2011; Penev *et al.*, 2011),此类论文主要是一个包含数据集链接的元数据文件。《生物多样性公约》(Convention on Biological Diversity, CBD)也在其2012年4月的一份旨在促进生物多样性数据共享的文件中推荐了数据论文(Convention on Biological Diversity, 2012)。

数据论文和数据发表为改善科学评价机制和

数据质量控制提供了一种方法。首先,数据论文是正式的学术论文,其引用可以被用于科研评价体系,从而可能会激励科学家更多地共享数据;其次,同行评审可以进一步控制数据集的质量(Costello *et al.*, 2013)。数据发表近来获得了很多关注,并出现了多个专门发表数据论文的期刊。2013年5月,在英国牛津举行的“Now and Future of Data Publishing 2013”会议(<http://nfdp13.jiscinvolve.org/wp/programme/>)上,与会的出版商、数据保存机构、科学家一起讨论了数据论文和数据期刊,大部分与会者认为,基于同行评审的数据发表模式会促进数据共享(Peggy Schaeffer, 个人交流)。2013年9月,出版商Pensoft创办了数据期刊*Biodiversity Data Journal*,用于发表生物多样性数据论文;同年10月,自然出版集团(Nature Publishing Group)也宣布启动数据期刊*Scientific Data*,将于2014年5月正式出版。

数据论文在理论上是可行的,也得到了很多出版商和数据保存机构的支持。但是,科学家对数据论文和数据发表持什么态度呢?Huang等(2013)通过分析数据论文潜在作者,即生物多样性数据生产者(包括博物馆或标本馆、职业科学家、非职业研究者或爱好者等3个群体)的行为倾向,认为数据论文有其局限性。比如,数据论文对公共资金支持的博物馆可能没有太多吸引力;公众科学参与者不熟悉学术出版模式,且没有经费支持论文的发表;而对于职业科学家来说,他们的研究更多是问题驱动,往往不愿意共享尚未经过分析的原始数据(Huang *et al.*, 2012)。

另一个值得讨论的问题是,数据论文是否应该有最小数据量限制? *Biodiversity Data Journal*所发表的论文中,有些仅仅报道了某种昆虫单个个体的分布记录(Thorpe, 2013)。这种发表模式无法有效杜绝将数据集故意拆分为多篇文章发表的行为,对数据共享评价机制的改善也不会有积极的促进作用,笔者认为并不值得提倡。因此,数据论文和数据期刊的推广和发展仍有很多问题需要解决,这种策略能在多大程度上促进生物多样性数据共享,还需要时间的检验。

3.3 联合数据政策

集中保存论文数据的模式对生物多样性数据共享有重要的借鉴意义,但数据共享不仅仅是把数据收集到数据库,还需要通过严格合理的同行评审

来控制数据质量, 以及通过制定较为规范和统一的格式来管理数据, 以便于进行整合分析。为此, Huang和Qiao(2011, 2012)建议生物多样性数据库和科技期刊采用联合数据政策(图2), 即生物多样性数据库保存机构(如GBIF)可以与生物多样性相关领域的期刊一起制定和遵守联合数据政策, 将共享数据至公共数据库作为论文发表的必要前提条件, 要求作者和数据共享者提供详细的元数据描述, 并采用标准化的数据格式; 像Dryad和figshare一样, 每个数据集被分配可供引用和跟踪的DOI。不管是论文数据发表过程中, 还是将数据直接提交至数据库, 期刊和数据保存机构都应该采用严格的同行评审标准来控制数据质量。

联合数据政策对于数据保存机构和期刊是双赢的策略, 在数据数量和质量上均具有可持续性。比如, Huang和Qiao(2011)通过对 *Journal of Biogeography* 2010年所发表的论文进行了统计分析, 发现104篇涉及物种分布数量的论文(占全年论文的一半)贡献了12,782个物种的367,000多条分布记录, 这是一个可观的数据量。在Huang等(2012)所开展的调查研究中, 372名生物多样性研究者中约70%表示接受期刊和数据库的联合数据政策, 约40%表示愿意在论文发表后立刻共享论文原始数据, 31%表示愿意在1-3年内共享数据。2012年6月, 17家有关淡

水生物多样性的期刊开始采用联合数据政策, 要求投稿作者将论文相关数据上传至GBIF共享(De Wever *et al.*, 2012)。这是近期较大规模采用联合数据政策的例子, 也暗示生物多样性数据库和科技期刊采用联合数据政策的策略值得推进。

3.4 数据共享和数据发表的关系

厘清“数据共享”和“数据发表”的概念和内涵, 有助于制定更加合理的数据共享政策。笔者认为, 从概念上来说, 数据共享是指通过某种形式将数据向同行或社会公开, 数据发表指通过发表的形式将数据公开和共享; 从内涵上来说, 数据共享包含数据发表, 数据发表则包含数据论文、数据库、个人网站等发表形式(图3)。在这一概念体系下, 数据共享和数据发表并不冲突, 数据发表是为了促进数据共享, 科学家也没有必要只采用数据论文的单一发表形式。因此, 有些科学家仅以数据论文来代表或定义数据发表并不可取(Chavan *et al.*, 2013; Costello *et al.*, 2013)。

考虑到科学家的研究模式, 大部分数据不可能通过数据论文的形式发表, 而是更多地与研究论文相关联。虽然数据论文有助于数据的传播和再利用(Chavan *et al.*, 2013), 但在论文发表量急剧增长和信息爆炸的时代, 数据使用者通过主要的公共数据库而非数据论文来检索数据更为有效。故除专门的

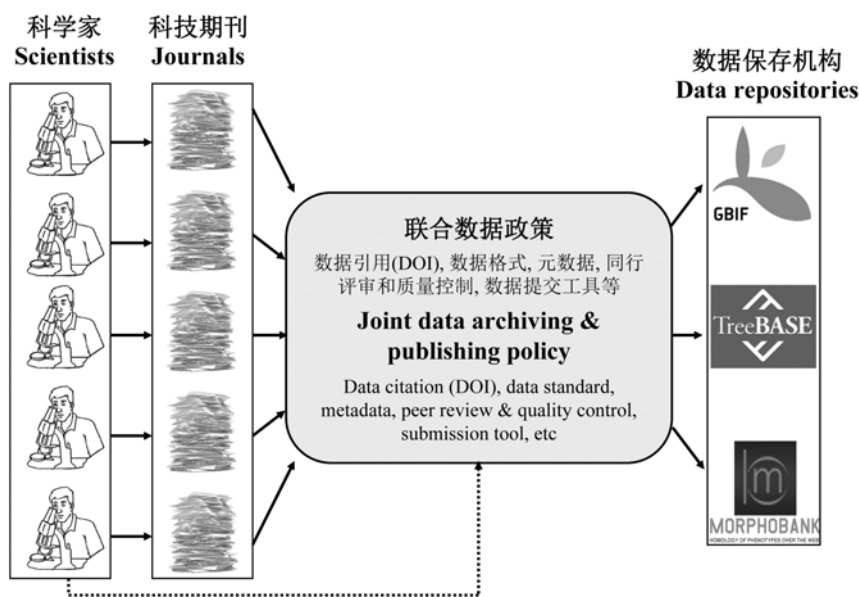


图2 数据保存机构和科技期刊的联合数据政策

Fig. 2 Improved joint data archiving policy for biodiversity databases and scientific journals

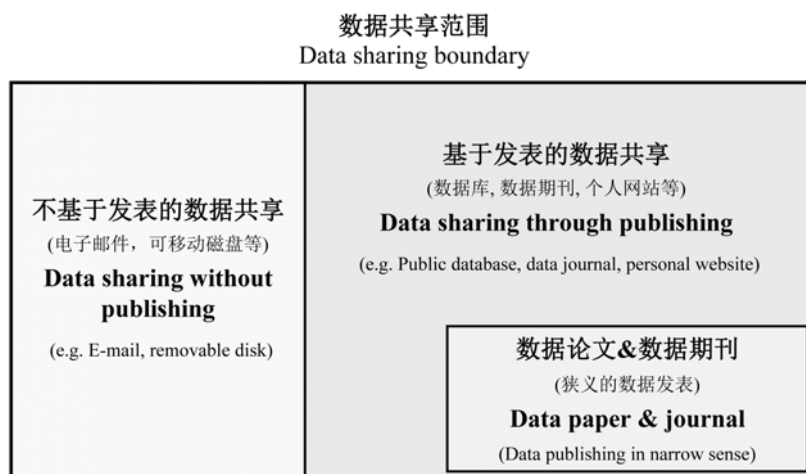


图3 数据共享和数据发表的关系。数据共享包含非基于发表的数据共享(电子邮件、可移动磁盘等)和基于发表的数据共享(数据库、数据论文、个人网站等), 即数据发表。

Fig. 3 The conceptual relationship between 'data sharing' and 'data publishing'. Data sharing includes data sharing without publishing (e.g. E-mail, removable disk) and data sharing through publishing (public database, data journal, personal website), i.e., data publishing.

数据期刊外, 其他期刊和公共数据库也可以采用同样的数据同行评审制度来实现数据共享。共享者可以在向公共数据库提交数据时提交元数据文件, 数据保存机构在对数据集和元数据文件进行严格的同行评审后分配给可供引用和跟踪的DOI号。因此, 数据保存机构完全可以行使数据期刊的角色。

在现有的评价体制下, 科学家看重的是数据的所有权和数据被利用情况, 如果数据集被分配了DOI便很容易地解决了这一问题。Thomson Reuters于2012年10月首次发布了针对数据集评价的“数据引证索引”(Data Citation Index), 开始基于数据库统计数据被引用和再利用情况。

4 生物多样性数据共享建议

推进生物多样性数据有效共享, 以及培育更好的共享文化, 需要科学家、数据保存机构、科技期刊、资助机构和研究单位等所有的利益方共同努力。我们尝试提出以下几个方面建议:

(1) 科学家应关注并努力践行数据共享。从事生物多样性研究的科学家首先应该了解数据共享和数据可用性对于科学发展的重要性, 并熟悉所研究领域的数据格式和管理规范, 以及相关的数据保存机构(一些代表性数据保存机构见表1)和数据提交工具, 在数据收集和管理过程中做好质量控制并准

备元数据描述文件, 发表和共享数据时应选择公认的专业公共数据库(最好能够为数据集提供可引用对象如DOI)。此外, 研究者还可以将数据备份在单位的数据库或个人网页, 比如提供指向数据集的DOI或链接地址以及数据集的引证方式。近来有些学者就数据管理和发表给出了一些切实可行的建议(Costello & Wiczorek, 2013; White *et al.*, 2013), 例如: 如何控制数据质量、确定数据格式、选择数据保存机构、处理敏感数据等。另外, 公众科学(citizen science)在数据收集和共享方面的重要性越来越被关注(Hampton *et al.*, 2013; 张健等, 2013), 职业科学家应该更多地支持和参与公众科学研究。

(2) 建议生物多样性数据保存机构使用DOI来解决数据共享者关注的数据所有权和数据引用等问题。采用标准化的数据格式将有利于数据的可用性, 通过诸如由数据生产者来参与数据维护和更新的措施来控制数据质量, 并充分考虑数据提交者和使用者的期望, 开发更易使用的数据提交工具。

(3) 推进生物多样性数据库与科技期刊采取联合数据政策。二者可以联合实行更加合理和严格的数据保存政策, 这将大幅度促进生物多样性共享数据的数量和质量。鉴于大量理应共享的论文数据并没能被合理共享(Yesson *et al.*, 2007; Alsheikh-Ali *et al.*, 2011), 期刊应该严格执行数据保存政策, 并在

表1 一些生物多样性数据保存机构及其特点(更多数据保存机构及介绍参见Databib (<http://databib.org>))
Table 1 Some popular data repositories archiving biodiversity data. More repositories refer to Databib (<http://databib.org>)

数据保存机构 Data repository	数据类型和说明 Data type and notes	元数据 Metadata	可引用对象 Citable object	开放程度 Access
GenBank http://www.ncbi.nlm.nih.gov/genbank	DNA序列; 标准格式 DNA sequences; standard format	要求 Required	序列号 Accession number	开放 Open
Barcode of Life Database (BOLD) http://www.boldsystems.org	DNA条形码及相关数据; 标准格式 DNA barcodes; standard format	要求 Required	序列号 Accession number	开放 Open
Global Biodiversity Information Facility (GBIF) http://www.gbif.org	物种分布数据; 标准格式 Species occurrence data; standard format	要求 Required	无 No	开放 Open
TreeBASE http://www.treebase.org	系统发育树及相关数据 Phylogenetic trees and related data	建议 Suggested	无 No	开放 Open
Dryad http://datadryad.org	生态进化出版物相关数据; 无标准格式 Ecology & evolution data associated with publications; no standard format	建议 Suggested	数字对象标识 DOI	开放 Open
figshare http://figshare.com	任何类型的数据; 无标准格式 Any type of data; no standard format	可选 Optional	数字对象标识 DOI	根据情况 Variable
中国国家标本资源共享平台 China National Specimen Information Infrastructure http://www.nsii.org.cn	标本数据 Specimen data	可选 Optional	无 No	根据情况 Variable

推进与数据保存机构联合数据政策的过程中起主要作用。

(4) 资助机构和研究单位应当发挥更重要的作用。Huang等(2012)发现受到资助机构或研究单位鼓励及要求的科学家明显更愿意共享数据。可以通过提供有关数据管理的详细政策或指南来促进数据共享, 同时改进科研评价体系, 给予数据共享等公益行为足够的支持。此外, 资助机构应该支持对数据共享起重要作用的公众科学研究项目的发展。

致谢: 本文的部分思想在参加发展中国家科学数据共享国际培训班(2013年11月)、第三届全国生物多样性信息学研讨会(2013年9月)、第一届地球系统科学数据论坛(2013年6月)等会议时形成, 感谢马克平老师等参与讨论。

参考文献

- Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Loannidis JPA (2011) Public availability of published research data in high-impact journals. *PLoS ONE*, **6**, e24357.
- BBSRC (2010) *BBSRC Data Sharing Policy*. <http://www.bbsrc.ac.uk/datasharing>. (accessed 2013-12-20)
- Cassey P, Blackburn TM (2006) Reproducibility and repeatability in ecology. *BioScience*, **56**, 958–959.
- Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, **12**(Suppl. 15), S2. doi:10.1186/1471-2105-12-S15-S2
- Chavan V, Penev L, Hobern D (2013) Cultural change in data publishing is essential. *BioScience*, **63**, 419–420.
- Convention on Biological Diversity (2012) *A Review of Barriers to the Sharing of Biodiversity Data and Information, with Recommendations for Eliminating Them*. <http://www.cbd.int/doc/meetings/cop/cop-11/information/cop-11-inf-08-en.pdf>. (accessed 2013-12-20)
- Costello MJ, Michener WK, Gahegan M, Zhang ZQ, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology and Evolution*, **28**, 454–461.
- Costello MJ, Wieczorek J (2013) Best practice for biodiversity data management and publication. *Biological Conservation*, doi: 10.1016/j.biocon.2013.10.018.
- De Wever A, Schmidt-Kloiber A, Gessner MO, Tockner K (2012) Freshwater journals unite to boost primary biodiversity data publication. *BioScience*, **62**, 529–530.
- Giles J (2006) The trouble with replication. *Nature*, **442**, 344–347.
- Haddaway N, Pullin A (2013) Evidence-based conservation and evidence-informed policy: a response to Adams & Sandbrook. *Oryx*, **47**, 336–338.
- Hampton SE, Strasser CA, 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.
- Huang XL, Hawkins BA, Lei FM, Miller GL, Favret C, Zhang RL, Qiao GX (2012) Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conservation Letters*, **5**, 399–406.
- Huang XL, Hawkins BA, Qiao GX (2013) Biodiversity data sharing: will peer-reviewed data papers work? *BioScience*, **63**, 5–6.
- Huang XL, Qiao GX (2011) Biodiversity databases should gain support from journals. *Trends in Ecology and Evolution*, **26**, 377–378.
- Huang XL, Qiao GX (2012) Biodiversity data sharing is not just about species names: response to Santos and Branco. *Trends in Ecology and Evolution*, **27**, 7–8.

- Ma KP (马克平), Lou ZP (娄治平), Su RH (苏荣辉) (2010) Review and outlook of biodiversity research in Chinese Academy of Sciences. *Bulletin of Chinese Academy of Sciences* (中国科学院院刊), **25**, 634–644. (in Chinese with English abstract)
- Michener WK, Jones MB (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution*, **27**, 85–93.
- National Research Council (2003) *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. The National Academies Press, Washington, DC.
- National Science Foundation (2011) *Proposal Award Policies and Procedures Guide*. <http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001>. (accessed 2013-12-20)
- Penev L, Mietchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D (2011) Pensoft data publishing policies and guidelines for biodiversity data. http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf. (accessed 2013-12-20)
- Pullin AS, Salafsky N (2010) Save the whales? Save the rain-forest? Save the data! *Conservation Biology*, **24**, 915–917.
- Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and opportunities of open data in ecology. *Science*, **331**, 703–705.
- Ryan MJ (2011) Replication in field biology: the case of the frog-eating bat. *Science*, **334**, 1229–1230.
- Shapiro JT, Báldi A (2012) Lost locations and the (ir)repeatability of ecological studies. *Frontiers in Ecology and the Environment*, **10**, 235–236.
- Stoltzfus A, O'Meara B, Whitacre J, Mounce R, Gillespie EL, Kumar S, Rosauer DF, Vos RA (2012) Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes*, **5**, 574.
- Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M (2011) Data sharing by scientists: practices and perceptions. *PLoS ONE*, **6**, e21101.
- Thessen AE, Patterson DJ (2011) Data issues in the life sciences. *Zookeys*, **150**, 15–51.
- Thorpe SE (2013) *Scutellista caerulea* (Fonscolombe, 1832) (Hymenoptera: Pteromalidae), new to New Zealand for the second time! *Biodiversity Data Journal*, **1**, e959.
- van den Eynden V, Corti L, Woollard M, Bishop L, Horton L (2011) *Managing and Sharing Data: Best Practice for Researchers*. www.data-archive.ac.uk/media/2894/managing-sharing.pdf. (accessed 2013-12-20)
- Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore JS, Renaut S, Rennison DJ (2014) The availability of research data declines rapidly with article age. *Current Biology*, **24**, 94–97.
- Vision TJ (2010) Open data and the social contract of scientific publishing. *BioScience*, **60**, 330–331.
- White EP, Baldrige E, Brym ZT, Locey KJ, McGlenn DJ, Supp SR (2013) Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*, **6**(2), 1–10.
- Whitlock MC (2011) Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution*, **26**, 61–65.
- Xu ZP (许哲平), Chen B (陈彬), Wang LS (王利松), Qiao HJ (乔慧捷), Liu FH (刘凤红), Qin HN (覃海宁), Ma KP (马克平) (2014) The development and trends of research on biodiversity informatics. In: *The New Biology Yearbook 2013* (新生物学年鉴) (ed. Poo, MM (蒲慕明)), pp. 290–312. Science Press, Beijing. (in Chinese)
- Yesson C, Brewer PW, Sutton T, Caithness N, Pahwa JS, Burgess M, Gray WA, White RJ, Jones AC, Bisby FA, Culham A (2007) How global is the global biodiversity information facility? *PLoS ONE*, **2**, e1124.
- Zhang J (张健), Chen SB (陈圣宾), Chen B (陈彬), Du YJ (杜彦君), Huang XL (黄晓磊), Pan XB (潘绪斌), Zhang Q (张强) (2013) Citizen science: integrating scientific research, ecological conservation and public participation. *Biodiversity Science* (生物多样性), **21**, 738–749. (in Chinese with English abstract)

(责任编辑: 马克平 责任编辑: 周玉荣)