

# 生物多样性数据共享现状和建议

数据共享对科学发展和大科学研究模式至关重要,也是各国政府部门科学决策的重要基础,因此越来越受到关注。*Nature*和*Science*杂志近期都曾发表社论和专刊讨论科学数据共享问题,一些国家也从政府层面积极推进数据共享。2012年3月,美国白宫宣布启动“大数据计划”,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。

1992年6月,联合国环境与发展大会通过并公开签署了联合国生物多样性公约(Convention on Biological Diversity),致力于保护人类赖以生存的生物多样性,促进生物多样性的可持续性利用,并以公平合理的方式共享生物多样性资源的利益。然而,CBD签署后的20年已经过去,生物多样性的流失依然严重,2010生物多样性目标远未实现,生物多样性科学研究和保护实践都面临新挑战。在信息时代的背景下,生物多样性领域面临的一个重要挑战是如何有效地共享及整合生物多样性数据。

生物多样性数据共享的重要性至少体现在3个方面:1)拓展生物多样性相关领域研究的深度和广度需要数据共享;2)制定切实可行的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分析,这在发展中国家尤为重要;3)原始数据共享是检验研究结果可重复性最重要的前提。

## 1. 生物多样性数据共享现状

2001年3月,联合国环境规划署和多国政府启动了“全球生物多样性信息机构(GBIF)”,旨在促进全球生物多样性数据的共享。虽然GBIF一直以来在整合生物多样性数据和数据共享策略方面做出了积极贡献,但其自上而下的运作模式却不利于鼓励众多个体科学家共享数据。此外,GBIF所涵盖的生物多样性数据类型还远远不够。

近几年数据共享呼吁在政府、资助和研究机构、数据保存机构、期刊等不同层面上越来越多。2009年5月,美国政府启动了国家数据网,同年9月,中国科技部也启动了“中国科技资源共享网”平台,都旨在促进数据资源共享。澳大利亚政府2010年10月启动了Atlas of Living Australia(ALA),近2年在生物多样性数据共享方面取得重要进展。

资助机构和研究机构也开始采取数据共享政策和措施。从2011年1月开始,美国国家科学基金会要求项目申请人在申请书中增加“数据管理计划”,写明数据保存和共享的承诺与措施;美国国立卫生研究院也有类似政策。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;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。

科技期刊在数据共享进程中也起着重要作用。一些主流期刊,如*Nature*、*Science*、*PLoS*和*BioMed Central*系列期刊,在其期刊政策中要求论文作者将论文数据共享。2011年,一些主流的生态学和进化生物学期刊与数据保存机构Dryad签署合作,承诺在投稿政策中鼓励作者将论文相关数据提交至Dryad保存并共享。Dryad为每个数据集分配一个DOI,使得数据集能

够被引用和跟踪,数据集的可引用性将鼓励科学家更多的共享。然而,Dryad对于数据集没有格式要求,这给生物多样性数据管理和整合利用带来了不便。

虽然生物多样性数据共享被越来越多地呼吁,但理想和现实还差得很远。最近有研究发现,对于已发表的生态学论文,其原始数据以每年17%的速度快速丢失。这暗示需要完善的公共数据库和数据保存政策来推进数据共享。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是数据共享能否有效推进和塑造数据共享文化的决定力量。我们近期开展的一项国际调查研究发现,生物多样性领域数据共享文化并不发达,科学家有一些技术和认知上的障碍。实际上很多障碍跟科学家能否从共享获得回报有关。

近来出现一些其他生物多样性数据共享策略。GBIF和出版商Pensoft在2011年提出了生物多样性数据论文,提议网络共享数据集的元数据文件可以用学术论文的形式发表,他们认为基于同行评审的数据发表将刺激科学家更多的共享,并可以控制数据集的质量。然而,数据论文能在多大程度上促进数据共享,还有待时间的检验。理想的生物多样性数据共享模式不但需要将更多数据收集至公共数据库,还需要严格控制数据质量并采用标准化的数据格式,因此有学者提出了数据库和科技期刊采用联合数据政策的建议。联合数据政策可以解决数据论文所关注的科研评价(DOI可引用)和数据质量问题(同行评审),并且能够在更大范围促进数据共享。

## 2. 建议

生物多样性数据共享需要所有利益方的共同努力:1)科学家应该关注数据共享方面的进展,努力践行数据共享;2)公众科学在数据收集和共享方面的重要性值得关注;3)数据保存机构应使用DOI来解决共享者关注的“数据所有权和引用”等问题,采用标准化的数据格式,控制数据质量,并开发易用的数据工具;4)科技期刊和数据保存机构采用更合理和严格的联合数据政策,从数据数量和质量上大幅促进生物多样性数据共享;5)资助机构和研究机构应当在数据共享中起到更重要作用,提供数据管理的详细政策,并改进科研评价体系,给予数据共享等公益行为足够的支持。

虽然中国的生物多样性数据共享近年有较大进步,但共享文化仍不发达。除以上提到的普适性建议,我国生物多样性数据共享还需关注其他问题。比如,增强公益性项目的规划性,摒除重复资助、项目执行不力等问题,公共数据保存结构在数据规模和质量、可用性和易用性方面需要很大提高。

文/黄晓磊,乔格侠

**作者简介** 黄晓磊,中国科学院动物研究所,副研究员。乔格侠,中国科学院动物研究所,研究员,国家动物博物馆,馆长。

本栏目专门刊登就促进科学技术发展提出的意见和建议,欢迎国内外科技工作者投稿。

(编辑 祝叶华)