问渠哪得清如许，为有源头活水来

——《中国科学数据》发刊词



郭华东 主编

中国科学院院士

人类正在进入大数据时代。作为科学生产力认识的革命性创新，大数据为科学研究带来了新的方法论；作为科学研究的新范式，大数据正在催生人们用全新的思维追求科学发现。在全球大数据浪潮涌动、我国将大数据列为国家战略的时代背景下，《中国科学数据》问世了。

1 数据是科学发现的基石和源泉

纵观科学的发展，无论是四百年前开普勒行星运动定律的提出，还是当今新粒子的发现、新医药的研制、数字地球的发展，等等，无不依赖于对实验、观测、调查、测量、模拟等所产生的原始数据及有关派生数据的分析和利用。正因为如此，信息技术一直以来的发展都以提高人类采集、传输、存储、管理与分析数据和信息的能力作为一项核心目标。在本世纪，上述能力的提高所导致的数据爆炸越来越显著，科学研究也日益表现出数据密集型和以数据为中心的特征，推动科学研究进入大数据时代。大数据作为知识经济时代的战略高地，是国家和全球的新型战略资源。围绕大数据引起的竞争不仅将决定国际信息产业格局，还将深刻影响经济发展、国家安全、科技进步和综合竞争力[1]。

数据成为检验科学研究价值的试金石。一方面，许多学科领域的科学发现以数据为基础，以新的数据发现为目标，辅以挖掘工具与分析手段，将数据与重要发现融合。另一方面，数据成为重复科学试验、确保研究成果真实可靠的检验基础。这一全新阶段在生命科学、高能物理、数字地球、全球变化、深空探测等领域的积极实践，无不彰显大数据与大科学紧密相连、密不可分的时代特色。科学数据走出深闺，成为科学研究的亮点和法宝，使推动产生新的科学发现成为可能。科学大数据作为大数据的分支体系，已成为继实验、理论和计算模式之后的数据密集型科学范式的典型代表，正在从模型驱动向数据驱动模式进行转化，带来了科研方法论的创新[2]。科学大数据已经并将继续在上

述大科学领域中发挥重要作用，未来必将在大科学领域为科学发现做出更加重大的贡献[3]。

2 数据开放共享是科技进步的动力和社会需求

近现代科学的快速发展很大程度上得益于科学的开放式探索。自十七世纪第一批科学期刊创建以来逐渐确立形成的研究结果公开发表与开放交流机制，使得科学具备强大的自我修正与承继向前能力。然而，受限于以前的媒介及传播技术，以期刊论文为典型代表的发表研究成果的文献中，作为研究结果证据的数据不得不被精简和极度省略。随着近几十年来信息技术的飞速发展，数据和信息存储、传播方面原先所面临的容量、时间和空间藩篱已被打破，存储和传播的成本一再下降，翔实的科学证据的公开在如今已具备技术条件。同时，它们的公开也变得更加必要和迫切。除了提高研究结果的可检验性和公信力以保障科学的自我修正能力之外，科学数据的开放共享还为更多的科学研究提供了丰富的基础材料，扩展科学研究的范围，扩张科学研究的视角，产生更多的科学知识，以及使得它们可以为科研以外的生产活动开发利用，产生难以预料的社会和经济价值。为此，科学数据开放共享不仅为学术界自身呼吁，而且为国家政府、资助机构、出版机构、公众等利益相关者关注[4–5]。

3 数据开放共享正在成为科学政策的重要组成

目前，已有不少国际性组织（如 OECD[6]、ICSU[7]、GEO[8] 等）和国家发布了关于科学数据开放共享的政策性文件。例如，美国政府先后颁布了 A-130 通告[9] 和《开放数据政策——将信息作为资产管理》文件[10]，美国国家科学基金会[11]、国立卫生研究院[12] 等科研资助机构也分别发布了相应的科学数据开放共享政策。在我国，科学数据开放共享也已上升到国家战略高度。国务院 2015 年 8 月发布的《促进大数据发展的行动纲要》[13] 中，明确提出了“积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享”的重要目标。很多科研教育机构也制定了本机构的数据管理和共享政策以适应上述国家层面的有关政策要求。例如，纽约大学的研究数据保留和获取政策[14]，剑桥大学的研究数据管理政策框架[15] 等。出版界也日益认识到科学数据开放共享的必要性，一些期刊出版团体（如 Science[16]、Nature[17]、BMC[18]、PLOS[19] 等）已经在行动，革新了期刊出版的有关政策。以上各层面的政策都在促使数据开放共享成为未来科学研究的一项基本公约[5,20]。

科学数据开放共享在具体实践层面，目前呈现出多模式探索之貌：既有大型国际合作项目或组织主导建立的开放数据库或系统（如世界数据系统 WDS〔 www.icsu-wds.org 〕、全球综合地球观测系统 GEOSS〔 www.geoportal.org 〕、全球蛋白质数据库 wwPDB〔 www.wwpdb.org 〕等），也有政府投资建立的国家级科学数据中心或系统（如美国国家航空航天局为其资助产生的科学数据建立的数据中心群〔 data.nasa.gov 〕[21]、英国自然环境研究理事会支持建立的数据中心[22]、我国国家科技基础条件平台中的各科学数据共享平台 / 中心 / 网〔 www.escience.gov.cn 〕等），还有科研教育机构建立的存储和共享本机构科学数据的存储库（如牛津大学的 ORA-Data〔 researchdata.ox.ac.uk 〕、哈佛大学的 Dataverse〔 dataverse.harvard.edu 〕等），出版界所进行的要求作者提交和公开论文支撑数据的实践，以及一些通用型数据存储库或数据存储与共享平台（如 Dryad〔 www.datadryad.org 〕、Figshare〔 figshare.com 〕、ScienceDB〔 www.sciencedb.cn 〕） 等。尽管科学数据开放共享在政策和实践层面都有令人欣喜的进展，但科学数据开放共享的成效总体来说还远未达到预期。

4 数据出版是推动与加速数据开放共享的有效战略

现代科学的职业化进程加速了科学的进步，其间确立起来的以文献为中心的科学家科研贡献和学术声望评价机制起到了重要作用。如今，虽然科学数据开放共享广为科研人员及各种利益相关者所认同，但科学家科研贡献和学术声望评价机制基本还未发生改变，这被认为是阻碍科学数据开放共享实践的一个根本问题[23]。此外，缺乏能够应用于大数据研究实践成果和学术著作快速出版的开放工具和平台，也是一重要挑战。目前亟需可用于不同学科、不同制度下的数据出版和数据共享的集成化平台，以及多数据集成化出版[24]。为此，科学数据开放共享领域的研究者们提出了数据出版的概念，其核心内容是使数据达到可引用和永久可访问的状态，为科研评价体系尽快将数据纳入其中，提供便利，因而数据出版被认为有望从机制上改变目前数据开放共享所面临的根本问题[25]。目前，数据出版的基本流程参考了学术期刊的出版流程[26]。从出版形式来看，目前主要有三种数据出版模式：数据附属于出版物的出版、出版物附属于数据的出版、以及独立的不依赖于出版物的数据出版[27]。其中，出版物附属于数据的出版是最新出现的一种数据出版模式，它与现有的科研评价体系相适应，试图通过当前的利益机制协调，推动科学数据开放共享。

科学数据出版提供了一种新的视角。它借助出版媒介发布数据，并引导读者便捷地发现、获取、理解、再分析利用与引用数据。实际上，一些期刊出版团体已经在行动——如革新了期刊出版的有关政策，提出了作者应公开论文支撑数据的要求等。而独立的数据出版是其中较新的一类探索，其科学地沿用了科研共同体的传统知识成果发布体系，试图通过有效的利益机制协调，使数据更具发现性、引用性、解释性和重用性。

中国科学院计算机网络信息中心和依托在该中心的国科联国际科技数据委员会中国委员会及秘书处长期以来关注科学数据开放共享问题，他们注意到了这种最新的数据出版模式，创立了我国首个专门的数据期刊《中国科学数据》。本刊希望借助于以数据论文带动数据出版的这种实践，使数据达到可发现、可解释、可引用、可重用、可评价及永久可访问等，切实推动我国科学数据开放共享的进程，亦为国际数据科学和开放发展作出贡献。可喜的是，本刊同时也是我国网络连续型出版物的首批试点刊物之一，愿能在业界起到抛砖引玉之作用。

数据出版是科研共同体的防腐剂，通过出版数据，我们将清楚地看到公共经费支持项目所捕获“大数据”的真实情况；数据出版是科研共同体的试金石，通过数据论文与数据集的开放获取，科学数据的价值外延将逐步扩展；数据出版是科学前沿的探测器，通过推广数据出版，我们也将更进一步感知国内外不同领域科学数据共享的创新进展。“问渠哪得清如许，为有源头活水来。”为唤醒沉睡的数据，《中国科学数据》希望为科研共同体输送源头活水，承载数据，问路未来。

引文格式： 郭华东 . 问渠哪得清如许，为有源头活水来——《中国科学数据》发刊词 [J/OL]. 中国科学数据 , 2016, 1(1). http://www.csdata.org/paperView?id=55. DOI: 10.11922/csdata.0.2016.0014.

参考文献

[1] Guo H D, Wang L Z, Liang D. Big Earth Data from space: a new engine for Earth science[J]. Science Bulletin, 2016, 61(7): 505 ～ 513.

[2] Guo H D, Wang L Z, Chen F, et al. Scientific big data and Digital Earth[J]. Chinese Science Bulletin, 2014, 59(35): 5066 ～ 5073.

[3] 郭华东 . 大数据、大科学、大发现 —— 大数据与科学发现国际研讨会综述 [J]. 中国科学院院刊 , 2014, 29(4): 500 ～ 506.

[4] 英国皇家学会 . 科学： 开放的事业 [M]. 何巍 , 王仲成 , 李振兴 , 等 , 译 . 上海 : 上海交通大学出版社 , 2015.

[5] 顾立平 . 科学数据开放获取的政策研究 [M]. 北京 : 科学技术文献出版社 , 2016.

[6] OECD. OECD principles and guidelines for access to research data from public funding[EB/OL]. [2016-05-27]. http://www.oecd.org/science/sci-tech/38500813.pdf.

[7] WDS. Data sharing principles[EB/OL]. [2016-05-27]. http://www.icsu-wds.org/services/data-sharing-principles.

[8] GEO-VI. Implementation guidelines for the GEOSS data sharing principles[EB/OL]. [2016-05-27]. http://www.earthobservations.org/documents/geo_vi/07_Implementation%20Guidelines%20for%20the%20GEOSS%20Data%20Sharing%20Principles%20Rev2.pdf.

[9] The White House. CIRCULAR NO. A-130[EB/OL]. [2016-05-27]. http://www.whitehouse.gov/omb/circulars_a130.

[10] Executive Office of the President of the United States. Open data policy-managing information as an asset[EB/OL]. [2016-05-27]. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

[11] The National Science Foundation. Proposal and Award Policies and Procedures Guide[EB/OL]. [2016-05-27]. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aagprint.pdf.

[12] National Institutes of Health. NIH data sharing policy and implementation guidance[EB/OL]. [2016-05-27]. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.

[13] 国务院 . 促进大数据发展行动纲要 [EB/OL]. [2016-05-27]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[14] New York University. Policy of on Retention of and Access to Research Data[EB/OL]. [2016-05-27]. http://www.nyu.edu/content/dam/nyu/research/documents/OSP/PolicyonResearchData030110.pdf.

[15] University of Cambridge. Research Data Management Policy Framework[EB/OL]. [2016-05-27]. http://www.data.cam.ac.uk/university-policy.

[16] American Association for the Advancement of Science. Science: editorial policies[EB/OL]. [2016-05-27]. http://www.sciencemag.org/authors/science-editorial-policies.

[17] Springer Nature. Availability of data, material and methods[EB/OL]. [2016-05-27]. http://www.nature.com/authors/policies/availability.html.

[18] BioMed Central. Open Data[EB/OL]. [2016-05-27]. http://www.biomedcentral.com/about/policies/open-data.

[19] PLOS. Data Availability[EB/OL]. [2016-05-27]. http://journals.plos.org/plosone/s/data-availability.

[20] 侯艳飞 , 胡良霖 . 数据政策的发展 [M]// CODATA 中国全国委员会编著 . 大数据时代的科研活动 . 北京 : 科学出版社 , 2014: 199 ～ 209.

[21] NASA Administrator. Data from NASA's Missions, Research, and Activities[EB/OL]. [2016-05-27]. http://www.nasa.gov/open/data.html.

[22] NERC. Data Centres[EB/OL]. [2016-05-27]. http://www.nerc.ac.uk/research/sites/data.

[23] Fecher B, Friesike S, Hebing M, et al. A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing (February 2015) [EB/OL]. [2016-05-27]. https://arxiv.org/ftp/arxiv/papers/1503/1503.00481.pdf.

[24] 郭华东 , 陈润生 , 徐志伟 , 等 . 自然科学与人文科学大数据 —— 第六届中德前沿探索圆桌会议综述 [J]. 中国科学院院刊 , 2016, 31(6): 707 ～ 716.

[25] 吴立宗 , 王亮绪 , 南卓铜 , 等 . 科学数据出版现状及其体系框架 [J]. 遥感技术与应用 , 2013, 28(3): 383 ～ 390.

[26] 张小强 , 李欣 . 数据出版理论与实践关键问题 [J]. 中国科技期刊研究 , 2015, 26(8): 813 ～ 821.

[27] 张晓林 , 沈志宏 , 刘峰 . 科学数据与文献的互操作 [M]// CODATA 中国全国委员会编著 . 大数据时代的科研活动 . 北京 : 科学出版社 , 2014: 149 ～ 158.