科研数据开放的国际实践及对我国的启示

阿儒涵[1,2]　吴　丛[1]　李晓轩[1,2*]

1 中国科学院科技战略咨询研究院 北京 100190
2 中国科学院大学 公共政策与管理学院 北京 100049

摘要　科研数据的开放共享被认为是下一个十年科研活动最显著的特征。在既有的生态观测和依托大型科研仪器设备所产生的“大科学”数据开放共享的基础之上，欧盟、美国等也在积极布局推进个体科研人员在科研过程中产生的“小科学”数据开放共享。小科学数据的开放共享有望将割裂的科研个体联接为科研创新网络，从而引发科研组织模式、创新范式的新变革。文章从资助机构推动、数据期刊建设、科研数据开放平台实践3个方面梳理了欧美国家在小科学数据开放共享中的实践，并提出对我国科研数据开放共享的启示。

关键词　科研数据，开放，国际实践，启示

DOI　10.16418/j.issn.1000-3045.20191216002

2019年初，爱思唯尔出版公司在其发布的《科研的未来：下一个十年的驱动因素与场景》[1]报告中，预测下一个十年科学研究的发展趋势，提出技术进步、开放科学以及东方（中国）崛起是影响未来全球科研发展的三大因素。其中，开放科学及作为其基础的技术进步两大因素都与本文要讨论的小科学数据开放共享密切相关。该报告认为，以信息技术发展为基础的科研数据开放共享将成为下一个十年科研活动最显著的特征，有望引发科研组织模式与科研创新模式的重大变革。科研人员将不再局限于自己的“一亩三分地”，而是能够快速地跨越地域、专业的界限，通过科研数据的交流加速科研活动的进程。“研究机构将投入更多的资金用于数据分析、加强数据的重复利用”，“只有附带原始数据及代码的研究才有可能得到发表”[1]。

与以往侧重于生态观测数据、依托大型仪器设备所产生的实验数据等的科学数据（science data）开放共享（在此简称为“大科学数据开放共享”）

*通讯作者

资助项目：国家自然科学基金面上项目（71974184），中国科学院青年创新促进会（Y9J0411601）
修改稿收到日期：2019年12月26日

的概念不同，科研数据（research data）开放共享强调在科研活动过程中产生的过程性数据以及论文等科研成果所依赖的支撑性数据的开放共享，既包括大科学数据开放共享，也包括单个科研人员的小科学数据开放共享。欧盟对科研数据的定义是蕴含在科研成果中的研究数据。经济合作与发展组织（OECD）则认为科研数据是指作为研究基础的事实性记录，包括数据、文档、图像、声音等①。鉴于以侧重于生态观测数据、依托大型仪器设备所产生的实验数据等的大科学数据开放已较为成熟并得到认同，本文中科研数据开放的概念侧重于科研人员在个性化的研究过程中设计、收集、加工而产生的数据开放共享这一新的动向，即小科学数据的开放共享。

小科学数据开放共享作为科研组织模式变革的新动向，许多国家已积极布局和推动，力图抢占科研数据开放的制高点。欧盟、美国、加拿大等在科研数据开放共享中的出发点各有不同。例如，欧盟在开放科学的框架下从政策、平台、治理机制等方面建立了科学领域数据的开放共享，强调“科学”与“效率”。特别是英国积极推动科研数据的开放共享，在组织方式、方法和平台等方面都有很好的探索[2]。美国、加拿大则是在开放政府的概念下，开展以政府决策透明化为目标的政府数据的开放。两国政府资助的科研活动所获取和产生的数据也在此框架下实现开放，强调“民主”与“公平”。虽然各国的出发点有所不同，然而殊途同归，在推动科研数据开放共享的政策、方法和路径上存在诸多共同之处。本文重点从资助机构的推动、学术期刊的新发展以及小科学数据开放平台实践等3个主要方面介绍国际上在小科学数据开放共享方面的实践，并提出对我国科研数据开放共享的启示。

1 政府资助机构是推动科研数据开放政策的执行者

近年来，部分资助机构对研究过程中产出的数据提出了开放共享的要求。据欧盟统计，目前全球26个主要国家的149家资助机构中有27.6%的资助机构要求接受资助的科研人员和团队对科研数据进行开放，14.6%的资助机构鼓励研究数据的开放（图1）。

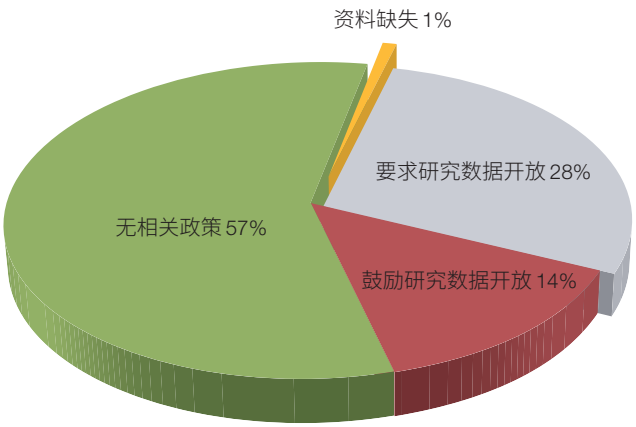


图1　全球26个主要国家的149家资助机构对研究数据开放的政策

资料来源：欧盟官方网站①

英国在科研数据的开放方面一直走在前列。英国科学界把海量数据的收集和处理看作是第二次科学开放革命，认为其规模和影响将超越学术期刊引发的第一次科学开放革命[3]。目前，英国有13家资助机构要求受资助的科研人员对科研数据进行开放。2004年，英国数字存储中心（DCC）正式上线，对受政府基金资助的科学研究过程产生的数据进行监管、存档并对外开放。2011年，原英国研究理事会（RCUK）颁布了《RCUK数据政策共同准则》（*RCUK Common Principle on Data Policy*），提出科研数据是研究过程的重要产出，科研数据的开放和集成有助于产生新的研究问题，推动研究发展[4]。2016年RCUK、英国高

① European Commission. Facts and Figures for open research data. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en/.

等教育拨款委员会、英国大学联盟以及英国惠康基金会（Welcome Trust）共同颁布了《公开研究数据的协定》（*Concordat on Open Research Data*），该协定要求接受其资助的科研人员对科研数据进行开放[5]。

在美国，科研数据的开放被作为美国信息自由以及开放政府的重要组成部分。早在1967年美国就颁布了《信息自由法》，用于规范美国联邦政府各机构政府信息的公开。2009年，美国政府层面的公共数据平台data.gov上线，科学研究数据作为政府公共数据中的重要组成部分也被要求在data.gov上实现开放。对于联邦政府资助的科研项目所产生的研究数据，美国预算管理办公室（OMB）制定了“A-89通告”“A-130通告”来规范科研数据的采集、提交与公开等。在资助机构层面，美国国家科学基金会（NSF）在其项目管理指南中强调，申请书需包含“数据管理计划”（Data Management Plan），其中需要对数据类型、数据标准、数据获取的共享政策以及数据存档的保存计划等做出说明，未包含数据管理计划的项目申请将不予受理②。

美国国立卫生研究院（NIH）作为美国生命健康领域的重要政府资助机构也致力于积极推进科研数据的开放。NIH将其科研数据的共享数据库（NIH Data Sharing Repositories）设置于美国国家医学图书馆③。该数据库收录了NIH支持的所有可再利用的数据，其中大部分数据来源于NIH及其资助的科研人员，也有少部分数据来源于某些封闭研究网络的科研人员及生物医药与信息分析的门户网站资源。该共享数据库总共包含82个分库，涵盖了生物医药学中不同领域、实验室、计划等内容。例如，表观基因组学（Epigenetics）、癌症纳米技术实验室（Cancer Nanotechnology Laboratory）、全球蛋白质资源（The Universal Protein Resource）等。所有分库都遵循开放的政策，为美国生物医药学相关的研究人员提供了庞大的基础数据支撑。在数据库的使用方面，基于不同数据的特点，每个数据库设有不同的数据公开保护期（protection period）。例如，表观基因组学数据库的保护期为9个月。在保护期内，研究人员需与数据拥有者提前沟通，在征求其同意后获得数据使用权。保护期结束后，研究人员则可随意下载使用该数据库中的数据资源[6]。

德国以德国科学基金会（DFG）为代表，推动科研数据的开放。DFG在2010年颁布的《研究数据操作指南》（*DFG Guidelines on the Handling of Research Data*）中提出“有质量保证的科研数据是产生科学知识的基础，科研数据的共享与再利用对科学研究甚至人类社会都具有巨大的意义”。并基于此提出了“数据留存且可获取、数据保存和共享要得到科学共同体的认可，数据要符合标准、构建数据开放共享的基础设施”等操作指南。并先后启动了“研究数据信息基础设施”（Information Infrastructure for Research Data）、“实践中的研究数据”（Research Data in Practice）等项目推动科研数据的开放[7]。2018年，DFG在其最新的指南中指出，DFG将以“科研数据”作为重点项目，制定不同学科的政策和规章制度，以适应不同学科的发展脚步；加强信息素质教育和信息管理专项培训，为科研人员的科研数据共享提供专业培训；将研究数据库和信息系统有机结合，在充分发展、优化现阶段资助项目的基础上加强对外合作和互通。

2 数据期刊的出现成为学术期刊发展的新方向

学术期刊对论文实验数据开放的要求在一定程度上推动了科研数据特别是实验科学科研数据的开放。一项对318种生命科学类期刊科研数据开放要求的统

② NSF. Guidance on Data Management Plans. https://www.nsf.gov/eng/general/dmp.jsp.

③ NIH. NIH Data Sharing Repositories. https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html.

计显示，11.9% 的期刊将实验数据的开放作为出版条件；9.1% 的期刊提出了实验数据开放的要求，但并未作为出版条件；23.3% 的期刊鼓励作者共享实验数据但并未做强制要求；而 31.8% 的期刊并未提到实验数据开放的问题。从期刊的影响力与论文实验数据开放的关系来看，期刊的影响因子与对期刊论文实验数据开放的强调程度呈正相关的关系[8]。

数据期刊的出现则在科研数据开放的激励与回报机制中迈出了重要的一步。科研数据的生产者在原有的期刊体系下只能作为数据的提供者与其他科研人员共同发表文章，而且很难获得第一作者或通讯作者等有分量的署名权。数据期刊的问世，标志着对科研数据生产者工作的高度认可与激励。2012 年 7月，Wiley 出版集团相继推出数据期刊*Geoscience Data Journal*（*GDJ*）、*Biodiversity Data Journal*。前者侧重发表地球科研数据论文，涵盖天气、气候、海洋学、大气化学、地质学等多个主题。发表的论文与存放在经认可的数据中心的数据集以及数字对象标识符（DOI）相关联。发表在*GDJ*的文章需要对其数据来源、处理过程、使用的软件等进行详细描述。自然出版集团（NPG）于 2014 年 5 月正式推出在线出版的开放获取杂志 *Scientific Data*，旨在发布具有科学价值的数据集、加强对科研数据的共享与再利用。*Scientific Data* 主要刊出一种新兴的被称为“数据描述符（data descriptors）”的内容，以帮助科研人员重复利用数据，也解决了收集数据的科研人员难以获得平等评价的问题。*Scientific Data* 发表的数据描述符遵从可追溯（findable）、可获取（accessible）、可通用（interoperable）、可重复利用（reusable）的FAIR数据准则④。

3 科研数据开放平台建设积累了初步经验

国际社会在大科学数据开放平台的建设方面已经积累了丰富的经验，建立了较完善的数据标准、存储及应用机制，为跨地区的大科学合作提供了良好的数据基础与平台。然而，在小科学数据开放平台的建设方面，仍处于起步阶段，主要停留在单一机构层面的开放共享。以美国著名的医学研究机构约翰·霍普金斯大学为例，部分学院在院系层面建立了科研数据的开放共享信息系统，科研人员可将科研活动中产生的数据按照一定的标准在学院范围内进行开放共享。这种院系层面的开放共享，虽然学科相对单一、数据体量有限，但也在小科学数据开放共享方面形成了积极的实践。

2018 年，欧盟启动建设的“欧洲开放科学云计划”（EOSC）则是目前国家层面最重要的小科学数据开放平台。EOSC 从构想到实现实质性推进历时 3 年。2015 年 5 月，在数字单一市场战略（digital single market strategy）下，欧盟委员会宣布将建设一个用于存储与处理科研数据的可信赖的平台，以支持欧盟科学在全球的领先地位，即 EOSC。EOSC 与传统科研数据平台最大的区别在于，其终端用户及数据提供者为个体科研人员。通过 EOSC 的建设，加强研究的证据性，使科研活动中产生的数据可追踪、可重复利用。更重要的是，通过 EOSC 的建设更好地利用信息技术的发展，实现数据层面的跨学科、跨地域的交叉融合[9]。

EOSC 的核心思想体现在 FAIR DATA 政策的提出与实施。FAIR DATA 相对于科研数据的开放更进一步，是指数据“可查找（findable）”“可获取（accessible）”“可交互操作（interoperable）”和“可重复利用（reusable）”（表 1）[10]。FAIR DATA 提出的背景在于，当前全球科研活动中产生的数据已经远远超过了人工可读取并处理的范围，信息技术的发展恰好为科研数据的机器处理提供了

④ *Scientific Data*. Principles. https://www.nature.com/sdata/about/principles.

良好的技术支持。小科学数据开放的目的不仅限于将数据公开到网络平台上，更重要的是实现数据的可机器处理和重复利用。借助信息技术、人工智能等实现基于数据的交叉融合与创新，有效弥补个体科研的局限性，引发科研组织模式与创新范式的变革。

表 1　FARI DATA 的标准[10]

要求	标准
可查找	元数据具有唯一且持久的标识符
	数据由丰富的元数据来描述
	元数据是清晰的且明确包含其描述的标识符
可获取	元数据可由标识符通过标准化的通信协议检索
	协议是免费开放和通用的
	该协议允许在必要时进行身份验证和授权
	即使数据不可再访问，元数据也可以访问
可交互操作	元数据使用正式的、可访问的、共享的和广泛适用的语言来表达知识
	元数据使用遵循FAIR原则的词汇
	元数据包括对其他元数据的合格引用
可重复利用	元数据由多个准确且相关的属性进行了丰富的描述
	元数据使用清晰、可获取的数据使用许可
	元数据标明数据来源相关联
	元数据符合领域相关共同体的标准

在 EOSC 特别是 FAIR DATA 的推进过程中，比硬件平台更重要的是数据标准及激励回馈机制的设计与实施。数据标准的建立是 FAIR DATA 实施的基础，涉及不同学科、主题数据在数据类型、格式、操作质量标准等一系列标准的设计。FAIR DATA 的实施将改变科研的传统——将科研活动从割裂的个体、小团队上升到网络化、分布式的层面。科研评价体系、资助机构对 FAIR DATA 的认可和采纳也将成为推进 FAIR DATA 实施的关键因素。

4 我国科研数据开放共享的进展及国际经验的启示

4.1 我国科研数据开放共享的进展

自 20 世纪 80 年代起，我国通过参与国际科技组织的科学数据开放共享，逐步开展自下而上的科学数据的开放共享。1984 年 6 月，中国科学院代表中国作为国家会员加入国际科联（International Council of Scientific Unions）科学技术数据委员会（Committee on Data for Science and Technology）。进入 21 世纪，我国开始自上而下地实施科学数据的开放共享。2000 年，科学技术部和国家自然科学基金委员会联合提出“实施国有科学数据公益性共享，并于 10 年内建成我国科学数据共享服务体系”的总体思路。这两个阶段科研数据的开放共享，主要还是侧重于大科学数据。

2018 年，国务院办公厅颁布《国务院办公厅关于印发科学数据管理办法的通知》（国发 2018〔17〕号），对政府预算资金资助的各级科技计划（专项、基金等）项目所形成的科学数据做出强制汇交的要求。据此要求，2019 年科学技术部重点研发计划项目指南中要求“在科研活动过程中的模型、算法、数据、数据产品、评估方法、研究结果、参数必须公开发表或在线免费共享”[11]。至此，我国对科研数据开放的要求明确涉及小科学数据的开放，在科研数据开放共享历程中迈出了重要一步。

4.2 国际经验对我国的启示

正如爱思唯尔出版公司在其《科研的未来：下一个十年的驱动因素与场景》报告中所提出的，以信息技术发展为基础的科研数据开放将引发科研组织模式与创新范式的重大变革。对此，我国必须高度重视，积极出台切实推进科研数据开放共享理念落实、落地的政策和措施；同时，应关注和借鉴国际经验。基于以上欧美国家在科研数据开放共享中的实践，得出对我国科研数据开放共享的 3 点启示。

（1）未来已来，科研数据开放共享的趋势势不可挡。科研数据的开放共享以数据为媒介，将单个科研人员的创新思想联结成创新网络，基于既有科研数据的二次创新为跨学科、跨地域的思想融会贯通提供了良好的平台，缩短了创新的进程，成为未来科研组织模式发展的新趋势。欧美国家在科研数据开放方面先行探索实践的经验值得我们研究吸纳，包括组织推动方式、数据期刊以及技术平台等多个方面。

（2）科研数据开放共享为我国科研管理中面临的评价、诚信等问题提供了新的解决思路。在论文、专利等产出之外，将科研数据纳入评价指标体系能够更准确地实现对不同层次、不同类型科研人员的有效评价。同时，数据的开放共享可使更广泛的科学共同体参与到科研诚信的监督中来，从而更好地解决因实验、数据造假等出现的科研诚信问题。

（3）科研数据开放共享仍面临诸多问题与挑战，需要我们在积极寻求国际合作的同时，走出自己的路。科研数据的开放有助于打破原有以科研人员个体为单元的科研组织模式，建立基于科研数据开放的新型科技创新模式。然而，在这一过程中，面临着政策、技术、管理、激励、伦理道德、科研诚信管理等诸多方面的问题。虽然欧美国家在科研数据的开放共享方面做出了积极探索，但是现有实践大都停留在收集、获取和汇集科研数据的阶段，尚未全面进入基于数据的融合、创新阶段，未来还有相当长的路要走。这既需要我们走出自己的道路，也需要积极参与国际合作。

致谢 欧盟联合研究中心Koen Jonkers和国家自然科学基金委员会龚旭研参与本文讨论。

参考文献

1 Mulligan A, Herbert R. Report: Research futures: drivers and scenarios for the next decade. [2019-02-14]. https://www.elsevier.com/connect/elsevier-research-futures-report.

2 胡明晖, 孙粒. 英国科学资助机构开放数据政策及其对我国启示. 中国科学基金, 2018, 32(5): 539-544.

3 Finch J, Bell S, Bellingan L, et al. Accessibility, sustainability, excellence: How to expand access to research publications. International Microbiology, 2013, 16(2): 125-132.

4 Ryan B. Research Data Management EPSRC's Expectations. [2019-12-03]. http://damaro.oucs.ox.ac.uk/docs/EPSRC%20RDM%20(Oxford%2028-Jun-2013).pdf.

5 Higher Education Funding Council for England, Research Council UK, Universities UK, Wellcome Trust. Concordat on Open Research Data. [2016-07-28]. https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf.

6 NIH. Open data access overview. [2009-11-01]. https://www.drugabuse.gov/funding/funding-opportunities/nih-common-fund/epigenomics-data-access-policies.

7 DFG. Management and infrastructure of research data: the national perspective and the DFG approach. [2018-07-04]. http://www.fz-juelich.de/SharedDocs/Downloads/UE/EN/18-07-11_OpenData_Stefan-Winkler-Nees.pdf?__blob=publicationFile.

8 Vasilevsky N, Minnier J, Haendel M, et al. [2017-04-25]. Reproducible and reusable research: Are data sharing policies meeting the mark? https://peerj.com/articles/3208/.

9 European Commission. Promoting an EOSC in practice. [2018-11-01]. https://op.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1/language-en.

10 European Commission. Turning FAIR into reality. [2018-11-01]. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

11 科技部. 科技部关于发布国家重点研发计划“干细胞及转化研究”等重点专项2019年度项目申报指南的通知. [2019-01-21]. https://service.most.gov.cn/2015tztg_all/20190125/2844.html.

International Practice of Open Research Data and Its Enlightenment to China

Aruhan[1,2] WU Cong[1] LI Xiaoxuan[1,2*]

（1 Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;

2 School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China）

Abstract Open research data is considered to be the most significant feature of scientific research activities in the next decade. On the basis of existing ecological observations and the “opening of large scientific data” generated by large scientific research equipment, the European Union and the United States are also actively promoting the opening of “small science” research data generated by individual researchers in the process of scientific research. The opening of small scientific research data is expected to connect the fragmented scientific research individuals into a scientific research innovation network, thereby triggering new reforms in the scientific research organization model and innovation paradigm. This research firstly sorts out the practices of European and American countries in open research data from three aspects: the promotion of funding institutions, the construction of data journals, and the practice of open platforms for scientific research data. Then, three enlightenments for open research data in China are proposed.

Keywords research data, open, international practices, enlightenments



阿儒涵 中国科学院科技战略咨询研究院副研究员，中国科学院科技战略咨询研究院第三方评估研究支撑中心主任助理，硕士研究生导师，中国科学院青年创新促进会会员。主要研究领域为政府科研经费配置与管理。承担来自财政部、中国科学院、国家自然科学基金委员会的多项决策支撑项目，研究成果多次支撑国家科研经费管理改革政策的制定。承担国家自然科学基金青年科学基金项目、面上项目等多项研究项目，在国内重要期刊发表多篇学术论文。E-mail: arh@casisd.cn

Aruhan Associate Researcher of Institutes of Science and Development, Chinese Academy of Sciences (CASISD), director assistant of the Third Party Evaluation and Research Center, CASISD, member of the Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS). Her research interest focuses on the allocation and management of government scientific research funds. Dr. Aruhan undertook a number of decision support projects from the Ministry of Finance of China, CAS, and the National Natural Science Foundation of China (NSFC). The research results have repeatedly supported the reform policy making of national research funding management. Dr. Aruhan also undertook a number of research projects such as the young scientists fund project, general project from NSFC, etc., and published nearly 20 academic papers in important domestic journals.

E-mail: arh@casisd.cn

* Corresponding author



李晓轩　中国科学院管理创新与评估研究中心主任，中国科学院科技战略咨询研究院研究员，博士研究生导师。现任北京市政协委员。主要研究领域为科研管理，涉及科技评价、科技人力资源管理、科研经费管理等研究方向。长期从事政府科技管理与政策方面的决策研究工作，近年来主要承担来自中国科学院、国家发展和改革委员会、财政部、科学技术部、国家自然科学基金委员会等重要科研课题。撰写了大量研究报告和研究论文。E-mail: xiaoxuan@casisd.cn

LI Xiaoxuan　Director of the Evaluation Center of Chinese Academy of Sciences (CAS), professor and doctoral supervisor at Institutes of Science and Development, CAS. Member of Beijing Committee of the Chinese People's Political Consultative Conference. His research area focuses on scientific research management, including research evaluation, human resource management, and research funding management. In recent years, he has mainly undertaken important scientific research projects from CAS, the National Development and Reform Commission, the Ministry of Finance, the Ministry of Science and Technology of China, National Natural Science Foundation of China, and so on. He has published more than 100 research reports and papers in domestic and international journals in management innovation and evaluation area. E-mail: xiaoxuan@casisd.cn

■责任编辑：岳凌生