

国家级基金项目数据强制汇交的 必要性与可行性分析

陈亚明 李刚

(南京大学信息管理学院, 南京 210023)

摘要: 科研数据是科学创新的核心生产要素, 是国家重要的战略资源。本文分析国家自然科学基金委员会和全国哲学社会科学工作办公室对两大国家级基金项目建立数据强制汇交制的必要性和可行性, 认为相关机构施行科研数据强制汇交制度, 将是提高我国科研数据要素使用效率的一种操作性较强的有效手段, 会对我国科研数据管理的建设起到示范性作用, 进而成为我国数据基础制度建设的突破口。

关键词: 科研数据; 数据管理; 数据强制汇交; 国家自然科学基金; 国家社会科学基金

中图分类号: G203 **DOI:** 10.3772/j.issn.1673-2286.2022.10.006

引文格式: 陈亚明, 李刚. 国家级基金项目数据强制汇交的必要性及可行性分析[J]. 数字图书馆论坛, 2022(10): 28-31.

科学研究过程会产生大量的数据, 对其精炼汇总构成科研数据。科研数据既是当下科学研究的产出要素, 也是未来科学研究的生产要素。依法收集、组织、利用的科研数据应成为我国数据科学研究的基础性资源。国家自然科学基金委员会(以下简称“基金委”)和全国哲学社会科学工作办公室(以下简称“社科办”)是我国最重要的科研资助机构, 其资助的项目总体上代表了我国科研的最高水平。国家自然科学基金和国家社会科学基金产生的数据是科研数据中的精华, 是国家科技创新重要的数字资源, 对其充分使用显得尤为重要。然而, 目前相关机构虽然出台了项目数据管理政策, 但是并未要求结项时同步提交项目科研数据。以高校为例, 虽然高校建有高标准的信息中心, 图书馆的数字化工作也具备了一定基础, 然而科研数据的汇交和管理还没有引起相应的重视。据调查, 只有北京大学、南京大学、大连理工大学、武汉大学、西北工业大学等高校制定了科研数据相关的管理办法。国家自然科学基金和国家社会科学基金两大国家级基金实施项目数据强制性汇交对推动我国数据基础制度建设、科研诚信工作, 释放科研数据红利具有重要意义。

1 国家级基金项目数据强制汇交的必要性

1.1 现代公共财政纪律的要求

“数据洪流”是现代科学家面临的普遍问题。随着数据收集与处理技术的快速发展, 科学家们认为现代学术研究正进入“第四范式”^[1]。数据已然成为这一时代最重要的资源。如果说产业数据是现代商业的“化石燃料”, 那么科研数据便是现代科学创新的“发动机引擎”, 同样具有基础性资源的地位。正如煤的价值在于燃烧, 科研数据的价值在于使用, 不过数据的使用具有零边际成本特征, 使用的人越多越能释放其价值。目前, 我国大量国家财政投入的科研数据并没有公开发表, 而是储存在个人或团队手中, 无法充分发挥科研数据的使用价值。这既是巨大的学术浪费, 在一定程度上制约了科学技术的创新和发展, 实际上也是国有资产流失的问题。我国一直非常重视科研经费的投入, 对相关机构官网调研发现, 近五年, 国家自然科学基金和国家社会科学基金的一般公共预算拨款除2020年外, 都在

逐年增长。基金委和社科办资助的项目成果是中央财政投入形成的国家核心科技资源, 对其加以利用是政府的责任。然而, 项目成果不应仅指论文、专著, 还应包括项目数据, 项目生命周期产生的各类数据、实验数据和调查数据是其主体。如果能对项目数据进行强制性汇交, 将两大国家级基金的优质数据汇聚整理, 提供给科学界和社会各方依法合规利用, 使其成为我国数据基础制度建设的样板工程和示范工程, 将从根本上改变目前数据基础制度建设喊得很响却无法落地的被动局面。

1.2 科研诚信与学术透明度建设的需要

在学术研究中, 寻求证据解释现象、论证观点等都离不开数据作为支撑, 因此数据质量的优劣决定了学术成果的品质。学术论文是对研究工作及其结论的描述与论证, 但并不能完全披露出研究过程的全部数据集, 这就给数据造假提供了可乘之机, 数据造假情况时有发生^[2]。2018年中共中央办公厅、国务院办公厅印发了《关于进一步加强科研诚信建设的若干意见》, 明确要求科技计划管理部门、项目管理专业机构要加强对科技计划成果质量、效益、影响的评估, 建立科研过程可追溯制度、科研成果检查管理制度。从操作层面看, 完善科研项目的数据汇交制度将使得学术过程的追溯、学术成果的鉴定分析成为可能, 有益于增加学术透明度。作为学术界项目管理的权威机构, 基金委和社科办建立科研项目数据强制汇交制度, 将有利于学者加强自我约束, 杜绝数据造假, 形成优良的科研生态环境。

1.3 数据开放共享复用有利于科研创新

科研数据的开放共享有利于学术质量的提升。一方面, 现代科学和技术的重大进步不再是由个人或单一的“科学小圈子”实现, 而是多个学科、多个团队协同努力的结果。数据共享使学者能够轻松找到其他领域的数据, 从而促进跨学科交流和合作。另一方面, 数据共享可以避免重复劳动, 加速科学发展。公开的、易于访问的学术数据集, 既避免了重复收集数据, 提高了数据的使用价值, 又有利于不同专业的研究人员从新的角度分析数据, 提出新的见解, 有助于融合创新的实现。然而, 我国大多数出版物没有对数据公开做出要求, 数据的供给远远低于学者对数据的需求。以信息行为领

域为例, 元分析是研究行为科学的重要方法, 由于我国相关研究不同步发表实验数据, 或者公开论文中公布的数据不系统、不完整, 导致我国研究者在做相关主题的元分析和元评价时不得不采用西方学术文献中的数据。这样, 用西方的理论、西方的数据, 研究西方社会科学界提出的问题, 对破解中国发展难题难免有隔靴搔痒之感。长此以往, 不利于我国独立的学术话语体系建设。数据汇交是数据分享与利用的基础, 基金委和社科办对数据汇交进行强制规定, 对我国学术界的“数据使用习惯”将起到示范性作用, 有利于带动其他相关科研单位、学术期刊与出版社对数据管理与利用的重视, 促进全社会重视科研数据共建共享工作。

2 国家级基金项目启动数据强制汇交工作的可行性

2.1 数据汇交制度的实施已经具备良好的政策环境

我国政府早就认识到科研数据的独特价值。早在1996年, 第八届全国人民代表大会常务委员会第十九次会议通过的《中华人民共和国促进科技成果转化法》就对职务科技成果中的数据所有权做出了规定, “科技成果完成人或者课题负责人不得将职务科技成果相关数据占为己有”。2001年, 科技部印发的《国家科技计划管理暂行规定》指出, 所有国家科技计划都应建立兼容的数据库, 共享信息和数据资源。2002年2月20日, 科技部、教育部、中国科学院、中国工程院及基金委联合提出《关于进一步增强原始性创新能力的意见》, 指出数据资料是研究工作最重要的基础条件之一, 要求建立国家科技计划项目数据资料汇交制度, 对科研数据资料实行分时段、分级、分类的共享。2003年, 科技部印发《关于加强国家科技计划成果管理的暂行规定》, 进一步要求对项目的相关科研数据进行收集整理, 建立档案并作为项目验收通过的基本条件。2006年, 科技部、财政部印发的《国家科技支撑计划管理暂行办法》, 要求项目组织单位和课题承担单位按时上报项目和课题有关数据和成果。2007年修订通过的《中华人民共和国科学技术进步法》对科学技术数据相关的信息系统与资源库的建设做出了要求。2016年, 国务院出台的《政务信息资源共享管理暂行办法》提出政务信息资源“以共享为原则, 不共享为

例外”的原则。2018年，国务院办公厅印发《科学数据管理办法》，该办法继承了《政务信息资源共享管理暂行办法》关于数据分享的原则，并对各级科技计划（专项、基金等）管理部门建立科学数据汇交制度提出要求。随后，一些省市相继出台了科研数据管理政策。截至2021年底，网络调研发现江苏、山东、上海、安徽、重庆、甘肃、广西、海南、黑龙江、内蒙古、吉林、陕西、湖北、河北、宁夏、四川、天津、新疆、云南等19个省（自治区、直辖市）都在《科研数据管理办法》政策精神的基础上发布了科研数据管理的区域性规则，均从国家财政和社会两个资金来源渠道明确了科研数据汇交制度和汇交方式^[3]。由此可见，我国密集出台的相关法律要求和管理办法已经为基金委和社科办启动数据强制汇交工作提供了法律和政策依据。

2.2 科研数据管理实践提供了值得借鉴的模式

随着相关政策与规定的出台，我国在科研数据的管理制度建设、数据标准的研究制定、数据存储与分享建设等工作方面持续推进，取得了可喜的成绩，可以为基金委和社科办建立数据汇交制度提供参考。

在数据管理政策方面，从科研机构到高校都在加快政策制定和实施^[4]。例如，中国科学院于2019年制定了《中国科学院科研数据管理与开放共享办法（试行）》，其中明确建立“先制定数据管理计划再立项”的项目数据汇交机制和论文关联数据的汇交与管理机制，同时还出台了数据中心建设、项目科研数据汇交、汇交数据质量管理和共享等方面的一系列工作指南^[5]。

在数据标准的研究制定方面，数据规范的实施促进了数据的管理与复用。根据中国知网标准数据库，有多达700多条包括元数据和数据的质量控制在内的国家与行业的数据标准，涉及科研项目整个生命周期数据管理的方方面面^[6]。此外，《科学数据标准化白皮书》编制工作已于2022年3月启动，这将进一步助推全国科研数据工作的规范化发展^[7]。

在数据存储与分享建设方面，科研数据的存储与分享离不开科研数据中心、档案馆等数据存储单位的支撑。我国已将科研数据中心纳入重要的战略科技力量和重要的基础设施予以支持，形成一大批具有影响力的科研数据中心。目前已在不同领域形成20个国家科研数据中心^[8]。各大科研机构也建立了一批类型多样

的科研数据中心，既有组学原始数据数据库、地理空间数据云、北京大学开放研究数据平台等知名的特定学科数据库，也有像ScienceDB数据银行这样的综合数据库。截至2022年9月，ScienceDB数据银行包含超过60门学科的数据，647万个数据集，已经成为得到施普林格·自然、爱思唯尔等多家国际出版商认证的通用型科研数据存储库^[9]。

2.3 两大国家级基金项目数据的产权归属清晰

数据的产权问题一直是数据管理中的热点，其能否健全完善制约了数据基础制度的发展。基金委和社科办下的科研项目数据的归属权明确，对其使用和管理具有较强的可操作性，这为实施基金项目强制汇交制度提供了相应的基础。国家级基金项目只有强制汇交才能提升研究人员对数据质量的更高要求，才能保证汇交完整数据的意愿，然而部分研究人员对项目中的数据汇交有着抵触情绪，这是没有必要的。一方面，这是对基金项目数据的所有权缺乏认识。实际上根据相关规定，我国以财政资金资助为主的科研项目，其研发成果的知识产权属于受资助单位^[10]。作为国家资助的项目，基金委和社科办有权对项目全程进行监督管理，可在上级相关部门的协调下实行项目数据的强制汇交。另一方面，强制汇交数据制度在确保国家的数据资源安全的同时完全可以实现项目人员的权益保障。在实际操作中，基金委和社科办可以通过积极开展科研数据强制性汇交的试点工作，将数据汇交作为项目申报执行过程中不可或缺的一个环节，如要求项目人在申报书中必须提交数据管理和汇交计划，按期提交项目执行过程中产生的科研数据，将数据的提交情况与质量纳入项目的考核。同时，参考国际上的做法让项目人员在提交的数据管理计划中对其数据可否共享、是否需要延期共享、敏感性等问题作出详细说明^[11]，充分尊重项目人的劳动成果与数据的优先使用权。对于商业目的的科研数据使用，应该保障数据生产者应有的利益分配。

3 结语

2022年6月22日，中央全面深化改革委员会第二十六次会议审议通过《关于构建数据基础制度更好发挥数据要素作用的意见》，进一步把完善数据基础制度，推

动数据要素发展, 释放数据要素价值提到了事关国家发展和安全大局的新高度。然而, 数据作为新兴崛起的生产要素, 在涉及数字的产权、监管、流通、安全等问题中还有大量障碍有待破除。科研项目中产生的科研数据是处理过的数据, 与复杂的商业数据相比, 相对容易管理。而且, 作为生产数据的一种, 受资助的科研数据的数据权属相对明确, 国家在对其使用管理时没有产权方面的障碍。这意味着科研数据的管理与其他类型的数据管理相比更有可操作性, 基金委和社科办率先实施基金项目数据强制性汇交制度是构建我国数据基础制度的阻力最小、条件最优、时机最佳的路径选择, 也必将对其他类型数据的管理和使用起到示范引领效果, 从而有益于我国数据基础制度体系建设。

参考文献

- [1] BELL G, HEY T, SZALAY A. Beyond the Data Deluge [J]. *Science*, 2009, 323 (5919): 1297-1298.
- [2] 即将被Science撤稿, “明星”论文数据造假, 百万科研经费打水漂 [EB/OL]. [2022-03-29]. <https://baijiahao.baidu.com/s?id=1728563616205085351&wfr=spider&for=pc>.
- [3] 高瑜蔚, 石蕾, 朱艳华, 等. 《科学数据管理办法》实施细则比较研究——以正式发布的11份细则为例 [J]. *中国科技资源导刊*, 2019, 51 (3): 1-10, 17.
- [4] 王芳, 卜昊昊. 科学数据管理政策发展比较研究 [J/OL]. *中国图书馆学报*: 1-23 [2022-09-16]. <http://kns.cnki.net/kcms/detail/11.2746.G2.20220629.0853.002.html>.
- [5] 中国科技信息杂志社. 中科院科学数据中心: 把我国关键数据“牢牢掌握在手中” [EB/OL]. [2022-08-29]. <https://baijiahao.baidu.com/s?id=1734470998060579335&wfr=spider&for=pc>.
- [6] ZHANG L, DOWNS R R, LI J H, et al. A review of open research data policies and practices in China [J/OL]. *Data Science Journal*, 2021, 20 (3) [2022-10-09]. DOI:10.5334/dsj-2021-003.
- [7] 《科学数据标准化白皮书》研讨会顺利召开 [EB/OL]. [2022-09-15]. http://www.cnigc.cn/xwzx/zxd/202009/t20200915_4558259.html.
- [8] 国家科技资源共享服务平台优化调整名单出炉 [EB/OL]. [2022-09-11]. <https://kjt.hebei.gov.cn/www/xwzx/15/ssyw/kjyw85/182160/index.html>.
- [9] 国内首创 这家银行存的全是科学数据 [EB/OL]. [2022-08-28]. <https://baijiahao.baidu.com/s?id=1690088181755023799&wfr=spider&for=pc>.
- [10] 郑友德, 欧广远. 国家自然科学基金资助项目形成的知识产权归属问题研究 [J]. *河南省政法管理干部学院学报*, 2004 (4): 61-65.
- [11] 克莉丝汀 L.伯格曼. 大数据、小数据、无数据: 网络世界的数据学术 [M]. 孟小峰, 张祎, 赵尔平, 译. 北京: 机械工业出版社, 2017: 81.

作者简介

陈亚明, 男, 1988年生, 博士研究生, 研究方向: 智库研究与评价、数据管理。

李刚, 男, 1969年生, 博士, 教授, 通信作者, 研究方向: 智库研究与评价、图书馆学档案学理论与方法, E-mail: ligang@nju.edu.cn。

Necessity and Feasibility Analysis of the Implementation of Mandatory Data Submission for National-level Funds

CHEN YaMing LI Gang

(School of Information Management, Nanjing University, Nanjing 210023, P. R. China)

Abstract: Research data is the core production factor of innovation and an important strategic resource of a country. This paper analyzes the necessity and feasibility of establishing data submission rule by the National Natural Science Foundation of China and the National Office of Philosophy and Social Sciences for two national fund projects, and argues that the rule of mandatory data submission implemented by NSFC and NOPSS should be an effective and operable means to improve the value of research data, and will play a model role in the construction of research data management, then become a breakthrough point in the construction of the basic systems for data in China.

Keywords: Research Data; Data Management; Mandatory Data Submission; National Natural Science Foundation of China; National Social Science Foundation

(收稿日期: 2022-09-21)