

美国联邦政府科学数据管理政策及实践

王 炼

(中国科学技术交流中心, 北京 100045)

摘 要: 数据开放是美国政府长期执行的重要国家战略, 是联邦政府科研管理和鼓励开放创新的重要内容, 也是美国促进科技发展、提升创新能力和创新效率的重要手段。多年来美国建立了国会立法、联邦政策和部门规章三层面的制度框架, 将数据开放共享的理念贯穿整个科研项目管理周期, 并与时俱进、不断更新, 保证了联邦资金效率和影响最大化。

关键词: 美国; 数据开放; 科学数据管理

中图分类号: G311 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2018.07.008

科学数据是从事科技活动所必需的重要信息资源。随着现代科学技术的迅猛发展, 学科交叉融合加快, 海量科学数据呈井喷式增长。有效地收集、管理、开放和共享科学数据, 能够避免重复研究、降低科研成本、提高国内和国际研究人员的参与度, 从而提升研究效率, 推动创新创业和科技经济发展, 其重要性和经济影响已经被越来越多的国家所公认。

美国在开放科学数据方面走在世界前列, 早在 20 世纪 80 年代, 里根政府就曾提出收集国家科学基金会支持的研究项目所产生的所有数据并将其商业化, 但该举措当时仅涉及一个部门, 成效有限。奥巴马政府认为开放数据不仅有助于确保政府的公开、透明和负责, 且有助于促进创新创业、科学发展并带来其他公共利益, 因此包括科学数据在内的数据开放成为美国政府近年来力推的一项重要举措, 各联邦部门都积极制定扩大公共资助研究成果开放获取的方案。

1 美国科学数据管理的立法及政策体系

自 20 世纪中叶起, 联邦政府就开始设计和建设数据资源开放的相关法律框架和政策体系, 为科学数据开放共享奠定坚实的制度基础。

1.1 数据开放的法律基础

1966 年美国颁布实施的《信息自由法》(Freedom of Information Act, FOIA) 开创了联邦政府信息公开化的先河, 成为联邦政府促进政府数据和信息资源公开的范例。该法规规定, 各联邦政府部门每年 3 月 1 日前必须向国会提交该法案的年度执行报告, 接受国会的监督和质询, 任何人都有权向行政机关申请查阅和复制政府信息^[1]。《信息自由法》先后经过 4 次修订, 成为美国信息和数据资源公开的基本制度框架。此外, 2002 年的《电子政务法》和《联邦信息安全管理法》、1996 年的《信息技术管理改革法》、1994 年的《政府管理改革法》、1993 年的《政府绩效法》、1974 年的《隐私法》以及 1950 年的《联邦档案法》等国家法, 均涉及公共资源的管理和使用规定, 为科学数据管理体系的建立提供法律基础。

1.2 科学数据开放的联邦政策体系

美国早期的联邦科技报告体系是政府科学数据资源开放共享的雏形。二战后, 美国开始建立联邦政府资助的科研项目报告体系, 随后逐步完善, 形成完整的国家科研项目资料库, 并有条件地向公众开放。

2001 年, 联邦政府通过“总统管理议程”,

作者简介: 王炼 (1982—), 女, 一级翻译, 主要研究方向为基础科学、交通。

收稿日期: 2018-07-03

促进科技资源开放共享^[2]。白宫管理与预算办公室（OMB）和科技政策办公室（OSTP）在国家法律框架基础上，制定相应的规章制度，要求联邦研发经费的申请和审核信息、联邦经费资助产生的科学数据、购置的仪器设备等，都须本着服务国家利益的目的，在最大程度上实现公开、开放和共享。

奥巴马执政期间，高度重视科技创新对经济复苏和社会进步的支撑作用，进一步完善相关政策和法规。2013年，白宫科技政策办公室发布《提高公众获取联邦资助科研成果的备忘录》（以下简称《备忘录》），要求研发支出在1亿美元以上的联邦机构制定开放共享计划，将公共财政资助的科研成果、科研论文和数据在公开发表后不晚于12个月内向社会免费公开，同时要求接受联邦政府资助的研究人员加强对电子数据的说明和管理^[3]。2014年3月和11月，该备忘录进行两次完善更新，敦促联邦各科研管理机构加大力度对科学数据资源进行开放共享。该政策的目标是长期保存公共资助直接产生的未分类研究成果，便于公众检索和分析，从而使联邦研究资金的影响和作用最大化。

2 联邦政府数据资源管理的特点

2.1 注重开放共享

美国政府认为开放共享联邦资助的科研成果和信息数据，可减少重复投入、节约社会资源、提高企业等私营部门的创新创业。此外，经同行评议的出版物和科学数据的开放，能够扩大联邦投资的社会影响，还可以通过市场反馈提高联邦投入的目标性和针对性。因此实施数据开放政策可以加速创新突破，增强经济效益和创造就业。

2.2 强调数字化和可读性

2013年5月9日，时任总统奥巴马签署为推进政府信息公开和数字化共享的行政令，要求联邦政府进一步促进信息数据资源以机器可读的方式向社会开放^[4]。同日联邦政府数据开放政策发布，提出联邦部门数据开放的五大原则：（1）公共性，即政府数据在安全、保密及隐私法律规定限制之外，应默认向社会开放。（2）易获取性，即政府数据应便于主流数据库软件与信息系统进行自动索引、查询、下载等操作，并应采取版权格式，使用开源许可证书以供重复使用。（3）充分说明性，

即政府数据应配有充分的元数据，对其相关属性与背景情况进行说明，并明确数据的优缺点、应用限制等信息。（4）完整性，即政府数据在符合隐私法规的条件下，应以原始数据格式向社会开放，并尽可能达到最高的精度和最细的粒度。（5）及时性，即政府数据应在可能的最短时间内向社会开放以实现使用价值最大化，同时应公开相关联系人信息^[5]。

2.3 狠抓源头控制

根据2013年白宫科技政策办公室《备忘录》的要求，联邦科技部门在资助对外（Extramural）项目过程中实施全生命周期数据管理和共享计划。项目单位在提交申请书时，必须同时提交项目（包括国际合作项目）可能涉及的科学数据管理和共享计划，计划内容贯穿项目始终，包括对可能产生数据或成果的预测、数据呈现的形式、公开和保密数据的分类、未来数据共享路径等。对于数据管理计划的评议是项目立项评审的重要内容之一。在项目执行过程中，项目单位必须严格按照计划完成数据的汇交与共享，并在项目年报中进行报告，接受监督。项目结束时，主管部门将对项目单位的数据管理计划执行情况进行评估验收。因此，美国科研项目的数据管理计划从源头上保证了公共财政资助的科学数据的汇交，明确了各方的责任和义务，为科学数据开放夯实了基础。

2.4 推动跨部门协作

2013年白宫科技政策办公室《备忘录》明确了联邦部门在数据开放中的主体责任和任务，把数据管理作为各部门联邦资产管理的一部分，加强对数据资源的收集、分类、储存、维护和开放。联邦政府各机构通过统一的数据开放平台 data.gov 向社会发布开放数据库，目前该平台上共有284 964个开放数据库。此外白宫还大力推进公私合作的数据创新生态系统建设，在开源程序平台 GitHub 上启动在线开放数据工程（Open Data Project），共享各种面向政府开放数据库的创新解决方案。如今，联邦政府资助的研究成果和科学数据开放已成为各部门普遍遵循的规则，大大促进了公众和社会对联邦科技资源的获取、利用和再创新。

2.5 倡导国际合作与跨国共享

为推动科学数据开放的国际合作，2016年12

月, 联邦政府开放数据部际协调小组与国家科学技术委员会共同制定发布《通过国际科技合作促进联邦资助科学数据开放共享的原则》, 明确了美国联邦科技部门在跨国科学数据交流合作中应遵循的原则, 提出国际科技合作活动的合作伙伴应在项目启动时建立数据管理计划, 各伙伴方应确保科学数据技术和法律上的互操作性, 以及联邦政府和国际科技合作伙伴应实施共同的数据开放共享政策和数据标准, 以促进开放数据政策在国际科研合作中的应用^[6]。

2.6 注重保密

作为为政府信息的一部分, 美国联邦政府还注重科学数据的保密工作。《信息自由法》规定, 公众有权向联邦政府机关索取任何材料, 但明确指出 9 类材料可以作为例外不予公开, 包括加密信息 (Classified Information)、其他法律授权不予公开的信息、机密信息 (Confidential Information) 以及个人隐私等。白宫管理与预算办公室根据《信息自由法》的规定, 将科学数据定义为通过电子方式存储的、非保密的、被科学界所共同接受的、为验证科学成果所必须的相关数据, 不包括初步分析、论文草稿、未来研究计划、同行评议信息、学术通信、实物样本等, 也不包括商业秘密及可导致个人身份识别的相关信息^[7]。项目申请者应根据联邦法律法规确定联邦资助项目的成果数据能否公开, 如果不能则应在项目申请书的数据管理计划中阐明原因, 由资助机构项目办公室进行审核。

3 联邦部门管理科学数据的具体实践

按照 FOIA 的规定, 各联邦行政机构都要指定一名高级官员作为首席信息自由法官员 (Chief FOIA Officer) 和信息自由法公共联络员 (FOIA Public Liaisons), 负责本机构对《信息自由法》规定国家的执行。包括国立卫生研究院、国家科学基金会、国家航空航天局 (NASA)、能源部、农业部在内的联邦主要科研资助部门, 都根据《信息自由法》以及白宫科技政策办公室备忘录的要求制定了本部门科学数据管理政策, 并由首席信息自由法官员负责监督执行。大多数联邦机构将数据管理作为项目申请、评审和验收环节必不可少的一部分, 在项目申请和管理指南中做出具体规定, 而国家航

空航天局和国立卫生研究院则专门发布计划或设立机构对科学数据进行管理。

3.1 国家航空航天局

2014 年, 国家航空航天局发布《加强科研成果的可获取性计划》, 系统制定了科学数据管理与同行评议论文发表的相关政策, 并将适用范围从国家航空航天局的各项航天任务扩大到所有其资助或部分资助的内部及外部科研项目, 以扩大科研成果及数据的可获取性, 加强科学数据管理及保存的规范性, 保证科研成果的可验证性与可重复性^[8]。

3.1.1 基本原则

国家航空航天局科学数据开放政策的宗旨是使其资助产生的科学数据能够被最广泛的公众免费获取, 从而加速科学发现, 使科学数据得到更多的衍生利用, 提高政府研发投入的效率和价值。但不同领域的数据保存和共享在管理上具有一定灵活性, 但是保护美国国家利益, 使知识产权、商业秘密及其他产权数据继续得到有效保护。

3.1.2 具体要求

自 2016 年 1 月起, 国家航空航天局所有雇员或获得国家航空航天局资助的机构和个人, 在申请研发项目时应包括数据管理计划 (DMP), 并包含以下内容: (1) 研究项目中产生数据的类型。(2) 数据与元数据的格式和内容标准。(3) 获取及共享这些数据的政策及方式, 包括对于隐私、保密、安全、知识产权及其他权利必要的保护措施。(4) 数据再利用、再传播、衍生使用的相关政策。(5) 支撑科学发表的相关数据的可获取计划。(6) 数据存储及保护的计划, 鼓励利用现有数据库或公共资源, 以及数据保存的时间范围。(7) 数据管理的相关成本应在项目预算中体现。

申请方提交的数据管理计划经国家航空航天局项目管理办公室审批后正式生效。项目管理办公室可针对具体项目, 对数据管理计划提出更加详细的要求, 如对于元数据内容、数据互操作性、永久性数据标识符、数据评估检查等的要求。所有项目申请必须同时提交数据管理计划, 但在某些情况下, 可以声明不予公开数据, 如涉及具体学生的教育项目数据、具有产权信息的数据、可导致个人身份识别的数据、出口管制数据、非保密但敏感性数据、受管制非保密数据、保密数据 (如与国家安全相关

的数据)、小企业创新计划与小企业技术转移计划项目数据等。

在数据不公开时,申请方还需制订专门计划,说明发表的科学论文结果应如何验证。

国家航空航天局项目办公室在整个项目管理周期中严格执行科学数据管理政策,通过定期报告等多种途径确保受资助方遵守其数据管理计划。如若违反数据管理计划要求,国家航空航天局将终止后续经费的拨付。

3.1.3 组织保障

国家航空航天局的数据管理政策由首席信息官负责牵头实施,并建立一个由来自四大技术部门共同参与的指导委员会,在各部门的研究计划中贯彻科学数据管理政策,针对具体项目制定数据管理计划相关要求,并加强部门间和跨部门协调。

3.2 国立卫生研究院

国立卫生研究院共有 73 个生物、医药、健康科研信息数据库,由其下属的国家医学图书馆(National Library of Medicine, NLM)统筹管理,并由各专业研究所实际运营。国家医学图书馆还汇总了生物医药领域其他机构的数据库及国立卫生研究院非科研信息类数据库,总共超过 300 个。此外,国家医学图书馆还支持 and 开展生物医学信息学和健康信息技术的研究、开发和培训,并负责协调由 6 500 个成员组成的全国医学图书馆网络,为美国各社区提供医学健康信息。

国立卫生研究院科研信息数据库可分为两类:一是开放数据库(Open Access),任何人都可以通过互联网自由访问并获取数据;二是受限数据库(Controlled-Access),仅限研究人员在符合相关规定的前提下用于特定项目,但受限数据库中,也包括开放访问的部分。例如在基因型和表现型数据库(Database of Genotypes and Phenotypes, dbGaP)中,研究步骤、描述信息等内容属于开放访问部分,而基因型编码等实质数据属于受限内容。

国立卫生研究院下设科技政策办公室,专职负责生物医学及健康相关的科技政策,管理科学数据的共享。长期以来,国立卫生研究院在科技政策办公室牵头或参与下,发布了众多科学数据管理与共享政策,其中具有代表性的政策包括以下两个。

(1) 国立卫生研究院数据共享政策(NIH Data Sharing Policy)。该政策发布于 2003 年,是对以往国立卫生研究院数据共享管理办法的归纳、总结和延伸,对于加速国立卫生研究院科研成果转化具有重要意义。该政策规定了以保护实验对象、专利和涉密数据为原则的具体数据管理措施,包括设立密级标准、确保数据站点的保密安全性、禁止数据二次传播、确保数据用于研究目的、加强各机构审查委员会的功能等,并就违反协议条款的处罚措施做出了规定^[9]。

(2) 基因数据共享政策(Genomic Data Sharing Policy, GDS)。2014 年的基因数据共享政策是在 2007 年“全基因组联合研究政策”基础上所做的更新与升级。该政策详细规定了生物学者共享数据的义务和责任,明确要求研究者在基因型和表现型数据库等国立卫生研究院资助的数据库中进行数据共享;在数据获取方面,提出尊重隐私和专利以及遵守联邦和地方政府法律法规,并强调充分发挥各机构审查委员会的审查作用^[10]。

综合来看,国立卫生研究院的数据共享政策框架结构完整、可操作性强,既有原则性的规定,又配套了具体实施办法,并在遵守联邦和地方法规的基础上充分发挥机构审查委员会的职能,赋予政策实施单位一定灵活性。

4 看法与启示

数据开放是美国政府长期执行的重要国家战略,是联邦政府科研管理和鼓励创新的重要内容,也是美国促进科技发展、提升创新能力和创新效率的重要手段,其数据共享体系和政策中有几点值得关注:

(1) 从国会立法基础、联邦政策法令及部门规章制度 3 个层面搭建数据开放共享的制度框架,加强部际协调和归口管理,保证了联邦政府资助的科研项目成果数据的公开、开放和共享。

(2) 从科研项目着手,将数据开放的理念融入到项目管理的全过程,实现全生命周期监管,确保科学数据资源开放共享的来源。

(3) 注重数据开放政策的整体规划,在发布行政命令的同时通盘考虑政策执行的各个具体环节,包括数据开放的平台和基础设施建设、科学数

据的评价指标体系以及对科学家开放数据的技能培训等,从而保证数据开放政策能够真正落实到位,有效执行。

(4) 与时俱进,根据新时代的特点完善法律政策体系,强调数据的电子化和可读性,促进数据开放共享的实效和时效。

值得注意的是,美国联邦政府的数据开放政策并不是对所有科学数据实行开放,其开放部分主要集中在与社会发展和公众福利密切相关的领域,例如气候观测、地理信息、农业种植、卫生健康等,对加密信息、其他法律授权不予公开的信息、机密信息以及个人隐私等可以不予公开,且近年来联邦政府有借“国家安全”之名扩大科学数据保护范围的趋势。■

参考文献:

- [1] US Congress. Freedom of Information Act (5 U.S.C. Section 552) [Z/OL]. [2018-06-30]. <https://www.justice.gov/oip/blog/foia-update-freedom-information-act-5-usc-sect-552-amended-public-law-no-104-231-110-stat>.
- [2] White House. President's management agenda[EB/OL]. [2018-06-29]. <https://www.whitehouse.gov/wp-content/uploads/2018/03/Presidents-Management-Agenda.pdf>.
- [3] Office of Science and Technology Policy. Expanding public access to the results of federally funded research[EB/OL]. [2018-06-30]. <https://www.whitehouse.gov/wp-content/uploads/2018/03/Presidents-Management-Agenda.pdf>.
- [4] White House. Making open and machine readable the new default for government[EB/OL]. [2018-06-29]. <https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->
- [5] Executive Office of the President Office of Management and Budget. Open Data Policy—managing Information as an asset[EB/OL]. [2018-06-30]. <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.
- [6] National Science and Technology Council. Principles for promoting access to federal government-supported scientific data and research findings through international scientific cooperation[EB/OL]. [2018-06-30]. http://www.nesdisia.noaa.gov/docs/iwgodsp_principles_0.pdf
- [7] White House Office of Management and Budget. OMB Circular A110[Z/OL]. [2018-07-01]. https://obamawhitehouse.archives.gov/omb/circulars_a110/.
- [8] National Aeronautics and Space Administration. NASA plan for increasing access to the results of scientific research[EB/OL]. [2018-06-30]. https://www.nasa.gov/sites/default/files/atoms/files/206985_2015_nasa_plan-for-web.pdf.
- [9] National Institute of Health. NIH data sharing policy[EB/OL]. [2018-06-29]. https://grants.nih.gov/grants/policy/data_sharing/.
- [10] National Institute of Health. Genomic data sharing policy[EB/OL]. [2018-06-30]. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html>.

Policies and Practices on Scientific Data Management by US Federal Government

WANG Lian

(China Science and Technology Exchange Center, Beijing 100045)

Abstract: Data sharing has long been a national strategy in the US, and an integral component of research management as well as a powerful tool to drive innovation by the Federal government. Over the years, the US has put in place a sound mechanism for data sharing consisting of legislation, policy and regulation. The federal government practices data sharing through the lifespan of project management and keeps it up to date to maximize the efficiency of federal funding.

Key words: US; data sharing; scientific data management