

英国科学数据管理政策研究

张红亮

(河南护理职业学院 安阳 455000)

〔摘要〕 阐述英国非政府机构、政府部门、大学机构 3 类组织的科学数据管理相关政策, 详细介绍非政府机构数据管理计划、开放与限制、元数据与数据保存、管理资金使用与监测等方面, 对政府部门及大学机构数据政策内容及特点进行分析总结, 以期为国内相关研究提供借鉴与参考。

〔关键词〕 科学数据; 数据政策; 英国

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.07.004

Study on UK Scientific Data Management Policy ZHANG Hongliang, Henan Vocational College of Nursing, Anyang 455000, China

〔Abstract〕 The paper elaborates the policies related to scientific data management of non-governmental organizations, government departments and university institutions in the UK, introduces the data management plans, opening and limitation, metadata and data preservation, management fund use and monitoring, and other aspects of non-governmental organizations in detail, analyzes and summarizes the content and characteristics of data policies of government departments and university institutions, so as to provide references for relevant domestic study.

〔Keywords〕 scientific data; data policy; UK

1 引言

近年来科学数据及开放数据、开放科学日益受到重视。科研人员利用公共资源开展研究所产生的数据应是公共物品, 在所有权得到确认的情况下, 所有科学数据原则上应开放可用。制定适宜政策以保障科学数据为社会发展服务具有重要意义。科学数据政策制定涉及科研群体工作需求及一系列社会问题, 是一项重要而复杂的工作。我国于 2018 年发布了《科学数据管理办法》, 但配套实践落地政策较少。本文对英国科学数据管理政策进行综述研

究, 以期提供参考、借鉴。

2 《开放科学数据协议》简介

英国对科学数据管理较重视, 将数据及数据服务视为国家基础设施重要组成部分。英国众多机构的科学数据政策都与 2016 年发布的《开放科学数据协议》^[1] 定义的模式密切相关。该协议得到英国研究理事会、英格兰高等教育基金委员会、英国大学联合会、维康基金会等重要科研管理机构或科研资助机构广泛认可。协议内容包括定义、引言和 10 个核心原则, 其中规定了科学数据的保存管理、数据访问、存取限制、技能与培训、管理成本、审查等方面问题。该协议经多家机构评论、修订并吸收其他文件有益成分而逐渐成型^[2-3], 随科研政策发

〔收稿日期〕 2020-01-07

〔作者简介〕 张红亮, 硕士, 馆员, 发表论文 9 篇。

展和资助情况变化而修改,以保证其良好适应性。

3 非政府机构政策

3.1 英国研究理事会

英国研究理事会 (Research Councils UK, RCUK) 为非政府部门公共机构,拥有独立政策制定、经费使用和管理权。共包括 7 个不同领域理事会成员:艺术与人文科学研究理事会 (Arts and Humanities Research Council, AHRC), 生物技术与生物科学研究理事会 (Biotechnology and Biological Sciences Research Council, BBSRC), 工程与自然科学研究理事会 (The Engineering and Physical Sciences Research Council, EPSRC), 经济与社会科学研究理事会 (Economic and Social Research Council, ESRC), 医学研究理事会 (Medical Research Council, MRC), 自然环境研究理事会 (The Natural Environment Research Council, NERC) 以及科学与技术设施理事会 (Science and Technology Facilities Council, STFC), 在英国科研工作中发挥重要作用。RCUK 现已并入英国国家科研与创新署 (UK Research and Innovation, UKRI)^[4]。此外维康基金会等社会慈善机构为英国科研事业提供资金扶助。

3.2 政策基础

3.2.1 概述 除《开放科学数据协议》外, RCUK 2011 年发布的《数据政策通用原则》^[5] 也对各专业理事会政策影响较大。两者互有异同, 其中《数据政策通用原则》更强调数据可发现的重要性, 《开放科学数据协议》在技能培训、长期数据管理方面关注更多。《数据政策通用原则》为各理事会政策制定提供总体框架及贯穿整个研究基础的一致方法^[5]。上述两个文件构是各理事会主要的数据政策基础。

3.2.2 数据范围 多数理事会数据政策覆盖其资助的科研项目。BBSRC 数据政策侧重于有较强科学数据共享需求的领域, 包括大量实验领域; 长期性、累积性地生产低容量数据领域; 会使用系统方法生成模型领域。对有强烈科学需求且成本效益明

显的其他领域, BBSRC 也鼓励数据共享, 但不排除有某些限制性说明的情况^[6]。ESRC、NERC、MRC 等政策基本覆盖科研人员在理事会支持下产生的大部分或全部数据^[7-9]。NERC 在政策和指南文件中表示允许所有研究的环境数据被存取、组合、再生成, 提供详细数据清单帮助研究人员评估不同类型数据价值^[10]; 同时其政策和工作程序尚在不断完善以明确具有长期价值数据, 希望合作制定相关标准。STFC 政策中区分原始、派生、公开数据^[11]。需要注意的是各理事会政策在科研出版物数据方面较不明确, 这可能缘于该类型数据管理常深度涉及版权、保密与涉密、权益保护等复杂问题。

3.3 数据管理计划

3.3.1 DPM 相关政策内容 除 EPSRC 外^[12] 其他研究理事会及维康基金会政策^[13] 均要求申请者提交数据管理计划 (Data Management Plan, DMP), 为项目申请必备部分, DMP 质量将影响高质量科研提案评估。多数理事会建议 DMP 制定应在英国数据监管中心 (Digital Curation Center, DCC) 指导下进行。MRC 建议 DMP 应关注同行科学家、专业数据管理人员这两个群体并提供模板和指导。对人口和病患方面的项目, MRC 对 DMP 数据共享提出 20 多项额外要求。数字产出、数字技术是有计划的科研成果重要组成部分, AHRC 要求提供包括技术方法、技术支持、技术经验在内的技术计划, 以解决保存、可持续性、利用等方面关键问题^[14]。NERC 表示 DMP 大纲作为资助申请初始要求的一部分有 1 篇幅即可, 但申请成功后必须在 NERC 数据中心指导下制定更详细的计划。维康基金会政策要求 DMP 解决存取共享数据的时间、地点、方式, 对数据保存的安排, 所需资源等方面问题。维康基金会及多数理事会要求在 DMP 包括关于数据格式和标准的信息。但对于数据采集标准化、定期校对、人工检查等^[3] 数据质量控制问题, 在上述机构的 DMP 要求或政策中未找到明确规定。

3.3.2 政策特点 各机构政策对 DMP 内容要求并不统一, 对数据保存的类型、位置、方式、时间、原因、质量控制、审查等方面的操作标准不一。分

析其原因主要包括以下两方面：一是各机构所属学科领域对数据对象管理的客观需求不同；二是各机构在最初制定 DMP 甚至制定政策时的切入视角不同，包括基于数据生命周期视角、基于数据管理所涉及各利益相关者视角或二者兼顾。

3.4 元数据与数据保存

3.4.1 元数据 各理事会及维康基金会均肯定元数据的重要性。例如 BBSRC 数据政策要求为了防止滥用、误读或混淆，要为二级用户提供关于数据来源或数据处理的必要上下文细节信息或文档；BBSRC 规定，如果某些元数据标准适用于特定领域或特定类型数据，则应提倡推行这些元数据标准；EPSRC 希望研究机构要确保在数据生成后的 12 个月内，在互联网上发布结构化元数据并可自由访问；NERC 提供数据目录服务，所有元数据都需遵循基于 GEMINI2 文件的统一元数据标准。各理事会均要求科研人员各自遵循相关数据中心建议，准备有关科研背景与科研设计恰当的详细元数据和文档，确定数据采集方式、时间、地点及采集人，调整处理相关工作程序等问题。

3.4.2 数据保存 (1) 保存时间。各理事会有不同规定，维康基金会在关于最佳科研实践的指南中规定数据保存时间为最少 10 年，临床、公共卫生数据保存时间更长。MRC 对临床数据保存时间规定为 20 年。BBSRC 表示在研究项目完成后希望科研人员确保数据保存 10 年，尽可能利用现有基础设施、期刊网站、开放存取仓储保存数据。EPSRC 规定最短保存期限为 10 年，要求科研机构在整个数据生命周期中进行有效管理，明确责任及资金分配。AHRC 规定最少 3 年且保存的时间和成本应与数字产出价值和重要性成正比。STFC 希望原始数据应尽可能保留更长时间，至少为项目结束后 10 年，而无法重新获得的观测数据应永久保留。(2) 保存位置。NERC 要求其资助的研究人员需要在 NERC 相关数据中心存储数据。STFC 希望由科研机构、大学、实验室的机构知识库或独立运营的特定主题数据库管理数据，前提是其能够最大化提升数据集的科学价值。如果不采用已有数据存储库，

则必须在 DMP 中对数据与服务的寿命及稳定性设置更多保障性规定。

3.5 数据访问与限制

3.5.1 数据访问 各理事会政策关于数据访问基本没有硬性要求。NERC 表示希望通过其数据中心访问数据。BBSRC 首选访问路径是已建立的数据库或机构库，在没有合适第 3 方参与机制的领域，则会根据需求直接共享数据。EPSRC 政策规定科研机构应负责提供数据访问服务。

3.5.2 排他专属期 科学数据是有益于社会进步的有价值资源，应尽量不设访问限制。但各理事会政策对在研项目数据访问设置排他专属期，以降低对正在进行研究工作的不利影响。BBSRC 强调数据应在不迟于主要成果发布之前提供访问，但也约定了特定领域的专属期限，如结晶学 12 个月（未公开的原始数据为 3 年），建议没有约定期的可以将 3 年作为指导性期限。如果将科研成果商业化也应考虑数据排他专属期问题。EPSRC 政策设置了一个有限的特许访问时段允许科研人员开展研究并发布成果，同时建议可设置不同长度限制期。MRC 认为根据数据的性质、价值以及生成和利用情况，对初级研究数据应该较为合理地规定有限度、确定性的排他专属期。不能为了过早、即时共享与分析数据影响当前研究工作及整个生命周期的数据资源采集工作。NERC 规定，从开始收集数据计算专属期限不得超过两年，但如果可以证明的、适当的科研需求则可延长。维康基金认为数据生成者有权享有合理专属期限，同时鼓励数据在发布出版前进行分享。RCUK 政策与维康基金大致相同，同时希望数据发布时间应不迟于由最终数据集形成的出版物被接收时间。NERC 认为通常 2 年的限制期已足够，特殊情况下也可延长时间。

3.5.3 其他限制性规定 基于《开放科学数据协议》和《数据政策通用原则》，各理事会制定相应限制性规定或程序，以避免不适当的数据发布妨碍研究工作。EPSRC 和 BBSRC 更加关注商业限制，认为与合作方达成数据管理共享协议较为重要。MRC 和维康基金更强调通过适当的道德、法律和机

构监管程序平衡免费与开放获取、访问管理与限制之间的关系,避免不适当的信息披露。NERC 表示在其数据中心存储数据不会影响知识产权,但权利所有者必须授予 NERC 非排他性许可,从而实现以重复利用为目的的数据管理。维康基金表示为了保护知识产权、发展服务公益的技术可适当延迟、限制共享数据。

3.6 资金使用与监测

3.6.1 资金使用 BBSRC 表示数据管理所需资金可在拨款申请中作为全部成本的一部分提出。维康基金认为 DMP 中确定的成本应作为所有申请的必备内容予以体现。MRC 认为资金问题可能涉及数据的管理、存储、分析,以及支持数据访问工作的人员、设备、基础设施、工具等方面。NERC 要求其资助的科研人员在拨款申请中标明数据管理活动具体项目成本,包括数据中心维护成本。EPSRC 认为科研机构应从收到科研资金开始对资金做出适当规定以便合理使用,同时建议大学或科研机构委托第3方提供适当的资金服务,或通过多家科研机构合作开发共享资金服务。

3.6.2 监测 BBSRC 政策表示对 DMP 遵守情况进行监测,可能要求被资助者提供更多遵守 DMP 的证明信息。NERC 表示经监测如未达要求可能会停止支付资金,甚至取消资助资格。EPSRC 表示根据具体情况进行监测。RCUK 则通过资助审查流程和资助终结报告进行监测调控。ESRC 表示如果数据在资助工作结束后3个月内未存储将扣留最后拨款。尽管各理事会及维康基金均对开放存取(Open Access, OA)出版物设置监测措施,但对遵守数据政策的监测未做规定。

4 政府部门政策

4.1 概述

英国部分政府部门和机构有重要科研计划,掌握着科研人员感兴趣的数据。2012年颁布的《开放数据白皮书》^[15]是英国宏观政策层面的指导性文件。该白皮书规定了访问公共部门信息的3项原

则:应在适当和可能情况下免费提供数据;公共部门数据应以公平价格提供;在确保数据价值时应该提供。之后多数政府部门基于该白皮书各自制定开放数据策略,但仅针对各部门工作过程中产生的行政信息数据,几乎不涉及组织或委托开展研究中生产的数据。

4.2 数据开放存取政策

4.2.1 概述 英国内政部和退休保障部等部门共同参与行政数据研究网络项目^[16],该项目旨在为社会、经济领域科研人员提供访问相关行政数据的安全环境,由 ESRC 资助,资助期至2018年7月底。2015年英国政府经济与社会研究工作组发布的《2015-2020年政府社会研究战略》^[17]未涉及科研过程中产生数据的访问政策问题,但提及需要发展数据科学技能。

4.2.2 英国内政部 其开放数据策略表明有意在英国旗舰数据平台 data.gov.uk 上发布更多研究数据,通过英国数据存档库(UK Data Archive, UK-DA)在严格限制条件下促进数据访问和利用^[18]。

4.2.3 国际发展署 提供科研数据资料开放存取的详细政策和实施指南^[19],该政策要求科研人员提交 DMP,在12个月内将衍生数据存入开放获取存储库,对申请访问者提供原始数据,元数据必须提交国际发展部存储库且在出版物中提供数据链接。

4.2.4 英国气象局 制定针对3类数据(开放数据、受管理数据、内部数据)的开放政策^[20]:开放数据须根据与公共气象专门工作组商定的标准进行评估,可依据“开放政府许可”免费获取;受管理数据可开放访问,但可能会受到涉及费用的管理规定限制。气象局在政策中同时提出数据访问方面的挑战:高数据量、快速的技术变革、资金和对强劲收入流的需求、与背景数据有关的限制性许可条款。

4.2.5 英国国家健康研究所(National Institute for Health Research, NIHR) 英国政府支持的大型科研机构,负责指导和协调英国国家医疗服务体系(National Health Service, NHS)的转化研究项目。NIHR 要求所有科研人员提交关于

访问基础科研材料及数据的方法说明,对该政策的实施情况进行年度评估,但不要求必须公开数据。NIHR 制定数据共享指南,包含数据共享问题案例,提出可能会以多种方式、原因限制数据访问^[21]。

5 大学机构政策

5.1 概述

目前英国并非所有大学都制定有科学数据政策。根据英国大学联合会 (Universities UK, UUK) 官网数据,目前该组织的大学成员有 136 个^[22],而根据 DCC 发布的清单只有 81 家单位制定有数据政策。但不排除有大学已制定政策但未被该清单收录的情况^[23]。

5.2 内容

DCC 认为数据政策包含 13 个方面内容:术语定义、制度性支持规定、DMP、覆盖范围、所有权说明、外部条件、元数据与文件标准、保留期限、道德问题、访问、开放存取、成本、审查,基于此 DCC 分析了 57 所大学的政策^[23]。结果显示 57 所大学政策中具备要素平均数约为 7。具备要素最多的是伦敦卫生与热带医学院政策,包含 12 个要素。有制度性支持规定相关要素的有 45 所大学,另有 7 所大学仅在政策中提及相关要素而未做专门说明。44 所大学对制定 DMP 有规定要求,另有 11 所将 DMP 列为可选项。对道德问题设置要求的包括 41 所大学,另有 11 所大学政策提及相关立法。40 所大学对开放存取进行规定,另有 6 所未做明确规定。57 所大学中的 21 所对政策覆盖范围未做具体规定。在各大学政策中较少涉及的要素有:数据所有权说明 16 所大学涉及;成本说明 20 所大学涉及;元数据与文件标准说明 22 所大学涉及;最短留存期限说明 23 所大学涉及。

5.3 特点

5.3.1 概述 英国大学机构数据政策具有多样性且存在相互借鉴关系。如米都塞克斯大学、诺桑比亚大学、约克大学等高校政策均与记录管理、安全

性、数据保护等问题政策存在关联。帝国理工学院、普利茅斯大学、苏塞克斯大学等高校政策均涉及知识产权保护问题。

5.3.2 数据保存 对于保存期限,约克大学等高校规定默认最低保存时限为 10 年;玛丽女王大学对医学临床研究数据默认最低时限是 20 年;诺桑比亚大学规定按照研究项目数据保存的风险等级设定时限,低风险 1 年、中风险 3 年、高风险 7 年。多所高校政策明确提及在项目结束时要对保存数据进行选择。玛丽女王大学要求针对“影响对已出版及公开可获取的最终科研成果进行理解的必要信息的最终形式”,要进行有选择的保存;伯明翰大学要求对支撑科研问题答案证据的数据要进行有选择的保存。

5.3.3 成本 伯明翰大学和苏塞克斯大学等高校政策明确提及研究人员需尽量获取资助者的财政支持。帝国理工学院政策规定基于最有效经济手段确定成本。

5.3.4 审查 多所大学确认将在数年内对政策进行审查,但只有少数大学确定具体时间,制定有实施计划的则更少。

5.3.5 政策构建 虽然各机构已普遍认识到科学数据管理是一项具有共同责任的工作,但体现在发展政策中的广泛参与、共同协作层次较低。部分机构政策构建主要基于对科研人员、院系部门、图书馆、信息服务单位、科研办公室等少数政策建设主体角色和责任的分配,其政策制定参与者范围受到限制。

6 结语

因立法体系不同,不同国家政策制定的目的与适用范围及政策执行、运行机制也不同,加之科学数据资源类型间存在较明显差异,其管理政策差别较大。除经费、覆盖范围、原则、计划或服务等内容外,学术共同体或学术机构对某些具体环节(如元数据标准的细节程度、数据访问权限控制程度、数据安全级别、数据审查及版权确认等)的重视程度受其机构特点、学科专业领域等因素影响

会有所差别,呈现不同特点,国内研究在借鉴其经验时应有所选择。从近年来数据政策发展来看,英国政府已将数据及数据服务视为国家基础设施重要组成部分,我国在这方面也应付诸更多实践。

参考文献

- 1 RUCK. Concordat on Open Research Data [EB/OL]. [2019 - 12 - 26]. <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>.
- 2 王丹丹. 英国《开放科研数据协议》的解读与启示[J]. 数字图书馆论坛, 2018 (8): 25 - 30.
- 3 OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding [EB/OL]. [2019 - 02 - 11]. <http://www.oecd.org/sti/inno/38500813.pdf>.
- 4 UKRI. About Us [EB/OL]. [2019 - 07 - 17]. <https://www.ukri.org/about-us/>.
- 5 RCUK. Common Principles on Data Policy [EB/OL]. [2018 - 12 - 26]. <https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>.
- 6 BBSRC. BBSRC Data Sharing Policy [EB/OL]. [2019 - 01 - 15]. <https://bbsrc.ukri.org/documents/data-sharing-policy-pdf/>.
- 7 ESRC. Submitting Datasets [EB/OL]. [2019 - 01 - 15]. <https://esrc.ukri.org/funding/guidance-for-grant-holders/submitting-datasets/>.
- 8 NERC. NERC Data Policy [EB/OL]. [2019 - 01 - 15]. <https://nerc.ukri.org/research/sites/data/policy/data-policy/>.
- 9 MRC. Data Sharing Policy [EB/OL]. [2019 - 01 - 15]. <https://www.mrc.ac.uk/documents/pdf/mrc-data-sharing-policy/>.
- 10 NERC. NERC Data Policy - Guidance Notes [EB/OL]. [2019 - 01 - 17]. <https://nerc.ukri.org/research/sites/facilities/marine/guidance/>.
- 11 STFC. STFC Scientific Data Policy [EB/OL]. [2019 - 01 - 15]. <https://stfc.ukri.org/stfc/cache/file/D0D76309-252B-4EEF-A7BFAF6271B8EC11.pdf>.
- 12 EPSRC. Clarifications of EPSRC Expectations on Research Data Management [EB/OL]. [2019 - 01 - 15]. <https://epsrc.ukri.org/about/standards/researchdata/expectations/>.
- 13 Wellcome Trust. Policy on Data, Software and Materials Management and Sharing [EB/OL]. [2019 - 01 - 20]. <https://wellcome.ac.uk/funding/guidance/developing-outputs-management-plan>.
- 14 AHRC. Research Funding Guide [EB/OL]. [2019 - 01 - 15]. <https://ahrc.ukri.org/funding/research/researchfunding-guide/>.
- 15 HM Government. Open Data White Paper [EB/OL]. [2019 - 02 - 18]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf.
- 16 Office for National Statistics. Administrative Data Research Network [EB/OL]. [2019 - 02 - 22]. <https://adm.ac.uk/about/>.
- 17 UK Government. Government Social Research Profession GSR Strategy 2015 - 2020 [EB/OL]. [2019 - 02 - 22]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/458667/GSR_Strategy_tagged_030915.pdf.
- 18 UK Government Digital Service Team. Home - Office - Open - Data - Strategy [EB/OL]. [2019 - 02 - 22]. <https://ckan.publishing.service.gov.uk/>.
- 19 UK Government. DFID Research Open and Enhanced Access Policy: implementation guide [EB/OL]. [2019 - 02 - 22]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181177/DFID-Research-Open-and-Enhanced-Access-Implementation-Guide.pdf.
- 20 UK Government. Open Data Policy [EB/OL]. [2019 - 02 - 27]. <https://www.metoffice.gov.uk/about-us/legal/open-data-policy>.
- 21 NIHR. Data Sharing [EB/OL]. [2019 - 03 - 02]. <https://www.journalslibrary.nihr.ac.uk/information-for-authors/data-sharing/>.
- 22 UUK. About Us [EB/OL]. [2019 - 03 - 19]. <https://www.universitiesuk.ac.uk/about/Pages/what-we-do.aspx>.
- 23 DCC. UK Institutional Data Policies [EB/OL]. [2019 - 04 - 30]. <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>.