

英国高校科研数据保存政策调查与分析*

王 琼 曹 冉

摘 要 科研数据保存政策是科研数据共享和再利用的前提条件和基础,日益受到学术界的广泛重视,切实可行的保存政策能保障科研数据的有效管理。本文采用内容分析方法,从保存内容、元数据、格式、位置、期限与销毁、机构与责任、保存权益、安全与隐私八个方面,对英国 28 所高校的科研数据保存政策进行调研,发现其科研数据保存政策不同保存阶段内容完备程度不同,具有学科差异的数据保存规定仍需进一步细化和完善。我国在制定科研数据保存政策时,应科学设定保存内容和期限,统一规范元数据标准和格式,搭建科研数据保存环境,严格明确数据管理中的责任与权利,完善版权保护和隐私安全政策体系。参考文献 51。

关键词 科研数据 数据保存 数据管理 政策研究 英国

分类号 G253

Investigation and Analysis of the Data Preservation Policy in British Universities

WANG Qiong & CAO Ran

ABSTRACT

As the prerequisite and basis of the scientific data sharing and reuse, the policy of scientific data preservation has been gaining its wide-range attention in academic circles. Practical policy is a strong guarantee of the effective management of the scientific research data. Methods of content analysis and Internet survey are adopted to carry out the research on the scientific data policies in 28 universities, UK Data Curation Center, United Information System of the United States and some other related websites. In the present study, the scientific data policy on the creating, preservation and reuse of the whole data use cycle was investigated from theory and practice respectively. Theoretically, it is discovered that the study of data preservation policy abroad has mainly been focused on the macro management and the specific process of scientific data preservation, such as the structure, application scope and effect of the preservation policy of national information centers, while domestic scholars mainly concentrate on the content analysis of foreign preservation policy and provide suggestions for the policy-making, such as study of Australian universities research data preservation policy, and the status quo of foreign research. Practically, it is the data preservation project carried by foreign specialized institutions, such as the permanent scientific data access

* 本文系国家社会科学基金项目“基于众包的图书馆数字馆藏建设模式及保障机制研究”(编号:12BTQ009)的研究成果之一。(This article is an outcome of the project “A Study on the Construction Pattern and the Guarantee Mechanism System of the Crowdsourcing Digital Library Collection” (No.12BTQ009) supposed by National Social Science Foundation of China.)

通信作者:曹冉,E-mail:18401614749@163.com,ORCID:0000-0002-7162-8191 (Correspondence should be addressed to CAO Ran,Email:18401614749@163.com,ORCID:0000-0002-7162-8191)

project implemented by the European Union, and the establishment of domestic scientific research data platform and institutional repository guided by the scientific data preservation policy.

To better guide the data management practice in universities and actively explore the scientific data preservation policy, this paper has chosen 28 universities' scientific data preservation policies as the research subject. Eight aspects ranging from preservation content, metadata, format, location, duration and destruction, institutions and responsibility, preservation benefit, to security and privacy are the focus. It is clear that some countries have conducted a significant attempt in data preservation policy-making which takes a guiding role in the whole practice period, from data creation, preservation to reuse. However, it is also found that there are some practical defects in previous research data preservation policies in UK: the incomplete content in different preservation stages, data preservation rules with different subjects are inconsistent. Therefore, when formulating the scientific research data preservation policy, it is proposed to set up a standard for the preservation content and duration, unified and standardized guideline to manage the metadata and format. Actively building a scientific data preservation environment, strictly and clearly defining responsibilities and rights of data management, and improving the protection of copyright and privacy and security policy system can guarantee scientific data preservation practice.

The innovation of this paper lies in its systematic proposal on scientific research data policy concerning the different periods of practical preservation. However, it mainly focuses on the policy guiding data creation, data preservation and data reuse, and other policies concerning data privacy and data security have to be dealt with in future research. 51refs.

KEY WORDS

Research data. Data preservation. Data management. Policy study. U. K.

随着科技的发展,科研数据规模不断扩大,复杂性不断提高。科研数据的应用跨越了学科的藩篱,范围逐渐扩大。同时,对科研数据的保存提出了更高的要求。布里斯托大学的一项能力计划评估结果表明^[1],科研数据的保存数量每年以 127% 的速度增长。美国国家科学委员会在《21 世纪数字资源长期保存的报告》中声明^[2]: 保存完善的科研数据,可以一种前所未有的精度和复杂程度为科学研究展现新的研究现象,并在信息集成过程中提供新的研究方法和创新点。与此同时,切实可行的科研数据保存政策,作为科研数据共享和再利用的前提条件和基础工作,日益受到学术界的广泛重视,以确保研究结果能够长期和精确获取,并为科学利用科研数据创造条件。

1 相关研究及实践

为了解目前高校关于保存政策的开展状况,笔者从理论和实践两个方面对相关研究进行梳理。

1.1 科研数据保存政策理论研究

通过文献调研发现,国外学者主要从国家宏观管理层面和科研数据保存过程两方面进行分析。在国家宏观管理层面,ShilobaGbaje^[3]等为了了解尼日利亚国家信息中心在数字保存方面所应用的政策,以及政策的效用、结构和应用范围,通过案例分析以及问卷和访谈等方式对保存政策的应用情况进行调研,认为保存政策的构建和应用存在管理机构重视程度不足,资金

支持不够,保存意识缺失,保存技能、设备和人员不足等问题,应从国家层面制定数字保存政策和实施策略,为科研数据保存实践提供明确的指导。Santos^[4]等回顾了文献信息的数字保存政策,包括信息技术发展、数字信息的出现、开放标准和元数据的使用、相关成本、安全性、长期保存和规划等方面,提出应加强数字信息的可靠性和长期访问性。Sheldon^[5]通过调研2008至2013年间档案馆、图书馆和博物馆的保存政策,提取政策内容和标准,认为档案馆和图书馆是国家级和省级政策发布的主要机构;随着国家对科研数据保存价值的认识日益加深,出台的保存政策数量平稳增长,特别是近两年出现急剧增长的态势。北美、欧洲和澳大利亚等地区相继制定了科研数据保护政策,并且不同时间、不同地域和机构发布的政策内容也各不相同。在科研数据保存过程方面,Moore^[6]针对实施科研数据管理政策过程中所需的验证评估标准的保存性元数据和相关数据集,提出了科研数据保存理论,并证明保护环境是数据保存的底层技术,是数据保存所必需的。Schottlander^[7]从生态系统的角度讨论科研数据保存,认为科研数据包括获取、管理和保存三个方面,具有繁殖力、脆弱性和复杂性三个特性,并认为以上各元素之间具有复杂多变的生态相关性。

国内学者的研究主要集中在对国外保存政策内容的分析和为我国建立相关政策提供建议两方面。杨云秀^[8]等以英国10所高校的数据政策为研究对象,介绍英国高校在科研数据存储、数据质量保障以及数据传播方面的政策和规范,拟定了科研数据管理政策的核心问题和要点。周晓燕^[9]对澳大利亚22所高校科研数据的前期处理、数据备份、保存期限、保存位置及变更等方面进行分析,介绍了澳大利亚高校科研数据的保存规定。司莉等^[10]通过对国外资助机构和高校制定的科研数据保存政策进行研究,分析科研数据所有权、保存负责人、最低保存期限、保存地点、保存格式等内容,建议我国在制定保存政策时应明确各方权益,建立完整

的权益保护体系,并以提高科研数据利用率为主要目标。丁培^[11]从科研人员需求、科研工作持续性的要求,以及科研项目资助机构的要求等方面,阐述了制定科研数据保存和科研数据管理政策的重要性,介绍了美国、英国和澳大利亚三国科研数据管理政策的现状。完颜邓邓^[12]从科研数据存储与共享平台建设角度,介绍了英美澳三国的数据存储指南,包括如何选择数据存储设施、数据存储的格式要求、元数据要求等注意事项。

1.2 科研数据保存实践研究

在政策研究的基础上,国外高校通过科研数据保存实践活动来促进政策的发展和完善,同时提高社会对科研数据保存价值的认识。2008年,欧盟开展科研信息的永久访问项目^[13](Permanent Access to the Records of Science in Europe),致力于对科研基本数据、过程分析直至研究结果的保存,为科研数据的获取和利用提供建议和指导。英国联合信息系统委员会^[14]认为,任何科研数据的长期有效获取和潜在价值都是在保存政策和原则的基础上展开的,为了让英国高等教育机构更好地了解 and 制定保存政策,2008年委员会通过网络调研,提出了保存政策的框架模型,包括政策声明、内容链接、保存对象、范围界定、相关责任、实施指南以及术语等。此外,相关学科领域也积极开展保存实践。如2005年,第109届美国国会在能源政策法案中创建的国家地质和地球物理资料保存方案项目^[15](National Geological and Geophysical Data Preservation Program),通过保存美国的地质历史,确保未来的工作能够具有安全、可用可信的信息能源。埃塞克斯大学与英国经济和社会研究部门联合信息系统委员会共同开展英国数据获取项目^[16](UK Data Archive),通过加强对科研数据的保存和管理,提高社会科学和人文学科的教学水平,促进研究资源的广泛共享,提高效用价值。Sheldon^[5]等人2013年通过网络调研文化遗产机构的保存政策,发现从2008至

2013 年间,48%的档案馆、45%的图书馆以及7%的博物馆都依据实践工作发布了数字保存政策或指南,保存政策的数量呈现增长的趋势。

与此同时,国内科研数据保存方面的实践也取得了一定的成果。由武汉大学图书馆主持的中国高等教育文献保障系统(CALIS)第三期的科研数据管理项目^[17-18],课题组结合项目实践,已基本建立了数据提交、数据组织、数据保存、数据共享、数据使用等规范,在建立科学数据平台建设流程、方法及科学数据管理的各类标准规范等方面取得一定进展。该项目已在武汉大学进行试点,以期将高校中分散的科研数据集中起来,进行长期保存,并实现数据的共享和再利用。2010年,中国科学院开展了科学数据网格示范项目^[19],基于丰富的科研数据资源和海量存储环境,利用网格技术整合科研数据,提供数据服务平台,有效地促进了科研数据的保存。此外,胡永生等^[20]对武汉大学、武汉理工大学等11所高等院校的教师和研究生进行调研,结果显示随着科学研究的长期开展和科研数据数量的增多,管理难度逐步加大,对科研数据的保存政策、保存工具的使用以及保存人员都提出较高的要求。谢春枝等^[21]对国内外学科数据管理现状进行梳理和分析,认为目前我国政府主管部门开始关注科研数据管理问题,但各类基金项目主管部门未对科研数据保存和共享提出明确要求,高校内部对科研数据管理也没有成形的制度,高校图书馆目前所建立的机构知识库作为科研数据保存的载体,以保存机构科研文献产出为主。

从以上调研可知,国外高校注重从国家层面对科研数据保存政策进行顶层设计,在政府以及相关信息机构的支持下进行探索和实践,制定了符合实际需求的数据保存政策,并通过实践不断对政策进行改进和优化。我国的科研数据保存政策目前还处于探索阶段,更注重学

习国外保存政策并在实践基础上对具体政策内容进行设计,还缺乏系统性和完整性。本文在前人研究和国内外高校实践的基础上,对英国高校科研数据的保存政策进行深入研究,基于数据产生、描述、保存、利用等整个生命周期的相关环节,以及数据生产方、使用方、保存机构三方责、权、利等方面进行全面调查和分析,为我国制定相应政策提供借鉴和启示。

2 英国科研数据保存政策调查

为全面反映英国科研数据保存政策的特点,本文以英国数据监护中心(Digital Curation-Center)所列举的高校机构的科研数据政策^[22]为研究对象,通过对各个院校的网站进行逐一检索和查找,最终获取英国28所大学^①的科研数据保存政策及指南,从保存内容、元数据、格式、位置、期限与销毁、机构与责任、保存权益、安全与隐私8个方面进行调查。政策制定时间为2009年1月至2015年11月,调研时间为2016年2月至3月。

2.1 科研数据保存内容

(1) 科研数据内容的价值。为保证保存数据的有效性,英国高校均对保存内容进行了界定。对于科研数据内容的价值,一是从科研数据保存的完整性上考虑,如邓迪大学^[23]认为,不论内容和格式如何,所有的科研数据均具有保存价值;二是从科研数据发挥的实际作用上考虑,如雷丁大学^[24]认为,并不是所有研究完成后的数据都需要被保留,应该保留能够直接证明研究成果以及具有长期保存价值的数据和数据集。

(2) 科研数据内容的生产途径。多数院校根据研究工作的需要进行科研数据的保存。较

① 28所大学包括:爱丁堡、北安普顿、伯明翰、赫特福德、华威、南安普顿、东伦敦、埃克塞斯、玛丽皇后学院、谢菲尔德、利兹、牛津、伦敦艺术、兰开斯特、牛津布鲁克斯、杜伦、埃克斯特、伦敦国王学院、曼彻斯特、诺丁汉、利物浦、巴斯、伦敦、谢菲尔德海兰姆、邓迪、雷丁、布里斯托、格拉斯哥大学。

有代表性的是赫特福德大学^[25],具体要求保存的内容包括:①观察数据,如传感器数据、调查数据、样本、神经影像;②实验数据,如基因序列、层析谱、环形磁场数据;③模拟数据,如大气模型、经济模型;④衍生或编译的数据,如文本数据挖掘、3D模型、编译数据库;⑤参考数据或者标准,如基因序列数据库、化学结构、空间数据门户等。雷丁大学^[24]规定数据具体的产生途径包括但不限于:实验或者模拟结果、统计和测量结果、计算机模型和软件、观察数据、调查结果、访谈记录以及运行脚本(包括代码)、来自相机和科学设备的图片、从二手资料编译的数据库、文本或语料库、实验书籍、物理对象等。

(3) 科研数据内容的范围。除保存科学研究产生的数据外,学校管理过程中产生的数据也属于保存范围。如赫特福德大学^[25]要求保存以下数据:①在学校管理、注册以及工作中产生的各种形式的数据;②学校的研究活动以及与合作机构共同开展活动所产生的数据;③经董事会批准的、学校拥有的附属公司所产生的数据;④除政策规定单独保存的数据外,其他需要按章程保存的数据。此外,巴斯大学^[26]也对不适用政策规定的数据做出明确规定,包括非数字型数据,受商业资助的对使用有限制的数据,以及已经由独立政策管理的涉及隐私的NHS病人数据和临床数据。

2.2 科研数据保存期限及销毁

(1) 科研数据的保存期限

英国高校对于科研数据的保存期限从3年到长期不等,有的根据具体的数据内容,保存10年以上或者永久保存。杜伦大学^[27]和牛津大学^[28]规定科研数据最短的保存期限是出版或公开发表之后3年;北安普顿大学^[29]等10所大学规定所有的科研数据在研究项目完成之后至少保存10年;伦敦国王学院^[30]要求不论是外部资助还是内部资助的项目,凡是符合学校保存政策的数据都将永久保存。此外,部分院校针对学科特性设定特殊的保存时间。如伯明翰大

学^[31]规定,医学、社会学、环境以及遗传等学科的重要数据可以参照英国研究理事会的相关规定保存20年或者更久;北安普顿大学^[29]要求,在科研数据保存期限满5年之后,由研究者和数据管理人员按学科对科研数据进行审核;巴斯大学^[26]还针对研究项目的网页提出存档期限,为项目结束后至少6个月。

(2) 科研数据的销毁和处理

英国高校对于未保存、失去研究价值或者违反法律和道德规定的数据的销毁和处理也有明确规定,一般由数据管理部门根据数据管理计划及学校的相关原则实施,并在机构知识库中进行登记。如雷丁大学^[24]要求审查保存期满的数据,并决定是否继续保留或者销毁,销毁时应注意个人隐私和敏感数据的处置。杜伦大学^[27]除了以上规定之外,还要求参考当前最佳的销毁方式进行销毁;如果数据管理计划未对销毁问题进行说明,则要求数据管理部门或相关教师重新审查数据,确立最适当的销毁方法,并在学校的知识库中进行登记,内容包括销毁或删除的原因、销毁时间和方式等。

2.3 科研数据保存描述

科研数据的描述主要是通过元数据进行,元数据与科研数据本身同样重要。正如张民^[32]的研究成果显示,科研数据的长期保存,一是维护数据本身,二是构建元数据,确保数据在相当长时间内可访问和重复利用,这就需要维护数据的语义信息、溯源信息、身份信息以及关系和权利信息等。据调查,有11所院校都明确规定在提交科研数据的同时,必须提交元数据、相关声明或协议,可见元数据在数据保存政策中占有重要地位。但是研究同时发现,多数高校对元数据的细节规定较少,内容不够详细,只有部分院校对元数据的标准和更新做出规定,如曼彻斯特大学^[33]和兰卡斯特大学^[34]要求科研数据的存储必须附有元数据,且元数据的设定应满足最低的数据标准,内容主要包括科研数据的创建和搜集方式、利用和传播途径、获取规定

和要求等。谢菲尔德海兰姆大学^[35]提出要定期更新元数据。对于非数字型的科研数据,英国高校也要求对其信息进行数据加工和整理,如雷丁大学^[24]要求建立标签、索引和分类,以便能够及时准确地查找到该数据,非数字型数据的存储应附有说明性文件。此外,多数高校还为研究者提供相关的培训和指导,规范元数据的内容,以便更好地保存和利用科研数据。

2.4 科研数据保存格式

(1) 保存格式的特性

英国目前主要针对数字型科研数据进行保存,保存格式对数据获取和有效利用起到关键性作用。本文通过分析高校的科研数据政策,发现科研数据的保存格式考虑以下几个特性:①兼容性;②可识别性;③持续可读性;④开放性;⑤符合资助者和学校存储要求;⑥便于交流、理解和处理。伯明翰大学^[31]规定,科研数据必须以可读的形式存储,除非有具体法律和规则要求,一般情况下以数据原始格式存储。南安普顿大学^[36]要求科研数据必须以数字或者计算机可读的形式保存,任何类型的数据,必须以持久可读的格式安全存储。

(2) 保存格式的细节规定

部分院校就保存格式的具体细节做出规定。如布里斯托大学的科研数据服务中心^[37]要求:①所有将要出版的数据都应存放在同一个文件夹中,在提交出版数据时应选择整个文件夹,保证数据的完整性,确保数据文件的命名合理、清晰;②组织数据时,除非有特殊原因,一般不能使用压缩工具,如 zip 或者 rar;③最上层的文件夹必须包含一个“readme.txt”文件;④文件或者文件夹命名中不能包含以下符号:> : " / \ | ? * ;⑤任何文件的路径不能超过 260 个字符,文件格式选择易于开放的格式,如.csv 而不是 Microsoft Excel。伦敦大学^[38]在科研数据管理指南中明确规定,应尽可能选择开放性、易于整理、非专属的格式存储,并要求:①文本型数据都采用 UTF-8 的编码格式;②图片采用

TIFF、JP2 或 JPG 格式;③音频文件以 WAV 格式存储;④长期保存的文件应该考虑到文件的有效性和长久性,如办公格式的文件采用 PDF、CSV,而不是 DOCX、XLSX 和 XML。

(3) 非数字型数据保存格式

对于非数字型科研数据的保存格式,明确规定的院校较少,主要通过以下两种方式解决。一种是将非数字型转化为数字型保存,不能数字化的则按原格式保存。如南安普顿大学^[36]规定,对于非数字化以及不便进行数字化但是有重要意义的科研数据应该安全存储,并建立标签、索引和分类,在学校知识库中进行备注。另一种对非数字型科研数据不进行保存,如伦敦艺术大学^[39]和牛津布鲁克斯大学^[40]。

2.5 科研数据保存位置

(1) 利用机构知识库进行保存

机构知识库以机构为主线,全面组织和集合机构及其成员创作的数字化产品,便于操作,易于利用,受到高度关注。如利兹大学^[41]规定学校提供和获取的所有科研数据应存储在一个适当的国内、国外或学校机构知识库中,除非在研究管理计划中有特殊规定。埃克塞特大学^[42]规定,在出版商允许的条件下,除有商业和保密问题之外,学校所有研究者产生的科研数据都应在 6—24 个月的封闭期之后,存储在国内外或学校的机构知识库中。此外,多数院校为了更好地利用科研数据,除特殊规定的例外,要求科研数据应该存储在允许数据再利用和获取的机构知识库中。

(2) 利用数据服务中心进行保存

为更好地保存和管理科研数据,一些大型的数据服务中心(Data Service)建立起来。相对机构知识库,数据服务中心的规模更大,同时可提供数据存储、获取、使用和管理等多方面服务。如英国数据服务中心^[43](UK Data Service)作为综合性的数据资源服务机构,可为研究者、教师和政策制定者提供高质量的社会经济学科的数据。数据服务中心的建立为高校科研数据

保存提供了更好的保存环境和平台。在调研的对象中,兰卡斯特等9所大学在政策中明确要求将科研数据存储在本地或国际的数据服务中心。

(3) 利用云服务进行保存

随着数据数量的急剧增长,科研数据的保存要求越来越大的存储空间,而云服务由于具有灵活性和敏捷性的云架构,为数据的存储利用带来超高效率。伦敦大学^[44]除提供机构知识库和硬盘设备保存之外,还提供基于第三方云存储的保存方式,但由于科研数据存在隐私问题等因素,该校的保存政策并不建议使用此方式。由此可以预见,经过不断发展和改进,云存储必定会成为今后科研数据保存的主要方式。

(4) 利用服务器等存储设备进行保存

对于没有自身机构知识库的院校,则要求存储在学校的服务中,如利物浦大学^[45]即是如此。部分院校对存储设备保存也有明确说明,如利用笔记本或者移动设备,必须考虑到存储设备的数据安全问题。如谢菲尔德海兰姆大学^[35]规定:如果科研数据需要被暂时存储在PC机等便携式存储装置中,研究者必须确保该存储过程是安全的,且符合学校对电子数据的加密政策。雷丁大学^[24]的研究人员还强烈建议,科研数据的存储不能完全依靠计算机的硬盘设备或非网络设备,如可移动驱动或者USB设备。若使用以上方式,数据应该经常进行备份以便安全保存。

2.6 科研数据保存机构与责任

科研数据保存机构主要有科研数据管理者(Principal Investigator, PI)、研究者、学校机构以及科研数据使用者。

(1) 管理者对数据管理过程负责

英国多数高校设有专门的数据管理部门,负责科研数据的管理和保存。在所调研的院校中,雷丁等18所院校明确规定了PI的责任,即负责科研数据管理政策的制定和实施,确保研

究人员有章可循。如诺丁汉大学^[46]规定,PI负责根据法律、道德和商业要求,审核确定保存的数据内容,确保科研数据在研究完成和结果出版之后的最短时间内被获取和使用。伦敦艺术大学^[39]要求科研项目完成后,PI与研究管理资助部门及技术服务部门共同审核科研数据管理计划,除了审核保存内容,还包括保存期限是否合理,元数据是否符合标准,以及获取和传播方式是否便捷等。

(2) 科研数据保存的服务和保障

学校在科研数据保存和管理活动中起着支持和保障的作用,其责任主要包括:①为科研数据的存储、备份和保留等制定明确的行动指南和实践指导;②为研究者提供科研数据保存和管理的培训和建议;③提供必要的保存设施和服务,如机构知识库保存或云服务。如伦敦国王学院^[30]由研究管理委员会和信息管理组共同代表学校负责政策的实施,代言学校与其他领域进行交流与合作。主要工作包括:管理科研数据的注册信息;提供相关建议和培训;制定科研数据有效保存的工作流程;对IT设施、数字监护和保存装置进行维护等。

(3) 研究者保障数据质量

研究者要与数据管理者合作完成相应的管理计划。如伦敦国王学院^[30]规定,研究者必须确保科研数据的真实性、可靠性、保密性和获取性,在项目进行期间遵守科研数据保存的相关原则,保证数据的安全性,并确保研究者授权访问是科研数据获取的唯一渠道。牛津大学^[28]除上述规定外,还要求研究者形成清晰的过程性文件,包括科研数据的搜集、保存、销毁、获取和再利用,以及合作项目中的协议和责任等。此外,研究者还应与数据管理人员制定数据保障计划,即在研究者离开学校或退休之后,对科研数据的保存地点、保存方式和使用权限等进行约定。

(4) 明确使用者权益

部分院校明确规定科研数据使用者的责任,规范其使用目的和行为,确保数据的使用符

合政策和相关规则的要求。如赫特福德大学^[25]要求使用者必须做到:①不泄露自己的账户信息和密码;②不能为个人利益篡改数据;③不能为商业目的获取数据;④不能在作者未授权或未获得PI同意的情况下将数据泄露给第三方;⑤不能泄露有关个人隐私的数据;⑥精确客观地展示和使用数据。

2.7 科研数据保存权益

(1) 所有权

英国高校对数据所有权的规定分为两种情况。一种是数据的所有权归学校所有,如南安普顿大学^[36]规定,学校是所有与研究数据搜集、创建、产生相关的法律权益的拥有者,学校有权订立关于研究数据出售、转移、获取和使用的相关协议。赫特福德^[25]和谢菲尔德海兰姆大学^[35]明确规定,除非研究授权条款或合同提供了例外情况,研究项目产生的科研数据是该学校的财产。另一种是归研究者个人所有,如利物浦大学^[45]规定,所有产生于学校的科研数据由研究者所有,但如果科研人员离开该机构则归学校所有。曼彻斯特大学^[33]还对所有权存在歧义的情况进行规定,如果相关数据的所有权存在歧义,如存在多个合作者,PI有责任提前确认数据的所有权、知识产权,促使达成协议并遵守学校有关知识产权政策及第三方协议。

(2) 知识产权

英国院校对于知识产权的规定只进行了简单说明。如南安普顿大学^[36]规定,研究数据在搜集、创建的过程中涉及知识产权时,知识产权的第一所有者将按照学校的知识产权规则对研究成果负责;如果研究涉及外部资助或与其他机构以及外部团体合作时,知识产权的归属和使用权应该在研究项目开始之前进行明确规定。曼彻斯特大学^[33]规定所有相关的数据以及知识产权都应服从学校的知识产权政策。伯明翰大学^[31]提出,政策的制定者要确保数据创建者的知识产权不被侵犯。此外,由于知识产权涉及学校和研究者的多方利益,多数院校都单

独制定知识产权政策,本文在此不展开研究。

2.8 科研数据保存安全与隐私

(1) 数据安全政策

英国高校对数据安全主要从重要性和具体措施两方面进行说明。如曼彻斯特大学^[33]认为数据的保存必须严格按照规定进行,并应对其他研究者及时开放,但为了数据安全,需严格服从保密协议和数据保护法,尊重所有权和知识产权以及第三方协议。邓迪大学^[23]规定,任何主体在获取研究数据时,若涉及到隐私、保密和其他法律利益,应严格遵守相关政策,如隐私政策、知识产权政策等。爱丁堡大学^[47]提出了具体的数据安全保护措施:①系统方面,要不断更新电脑上的防病毒软件,隐私数据最好存储在未联网的计算机上或考虑使用加密文件;②物理安全方面,要考虑计算机容易被盗或恶意破坏和修改数据,可以使用外部硬盘驱动器并存放在安全的地方,除非必须,否则数据不应该存储在PC或者闪存驱动器中;③数据结构方面,通过将数据转换成代码进行加密,可以有效防止数据被篡改。

(2) 隐私保护方法

主要有三种隐私保护方法:匿名化、获取密码和权利说明。如爱丁堡大学^[47]规定:当数据涉及人文对象,在共享时要同时考虑法律和道德问题,避免个人数据及敏感信息泄露。采用匿名化技术,将数据聚集为数字型数据,把视频进行录音编辑,匿名使用定性数据,以保护隐私数据。雷丁大学^[24]规定,个人和敏感数据的搜集不需要与数据同步共享,研究人员使用适当的知情同意、匿名化以及控制访问的策略进行隐私保护。如果搜集到的机密或者敏感数据可能会以某种形式共享,除上述方式外,还可通过非公开协议授权获取。曼彻斯特大学^[33]规定,当需要出版的支撑研究成果的科研数据有不能开放的合理理由时,可以设定必要的限制和保密协议。

3 英国高校科研数据保存政策分析

3.1 成熟的政策环境为保存政策的制定和实施创造良好的条件

英国良好的制度基础和实施环境为保存政策的制定创造了良好的条件。如英国2000年出台的《信息自由法》对数据使用就有相关规定,2003年英国对公共部门的数据共享制定了数据共享指南^[48],从法律法规层面为保存政策的制定奠定了良好的基础,有利于保存政策的形成和发布。此外,英国高校机构重视科研数据和信息的利用与保护,从多角度制定不同功能的数据使用政策,如伦敦国王学院的保存政策包含数据获取政策、数据保护政策、信息自由政策、信息安全政策、机构数据管理政策、研究出版政策、信息引用政策等,从应用层面为保存政策的实施创造了良好的执行环境。

3.2 保存政策贯穿科研数据从创建、保存到利用的整个实践周期

通过英国保存政策的内容要素可以看出,科研数据的范围界定、数据格式的选取和规范、描述性元数据及其标准的确定、保存期限与销毁、保存位置的规定等都是基于科研数据保存的具体流程而制定的;而科研数据保存机构与责任的规定明确了保存实践中各方的权责,使保存实践工作有序化;科研数据保存的权益保障政策避免了数据生产者、使用者和保存机构责权利不明的问题;科研数据的保存安全与隐私政策保证了科研数据的有效性和真实性。总体而言,政策的制定既包含科研数据保存的实践流程,又保障了各方的权益,贯穿于整个科研数据从创建、保存到利用的实践周期,具有系统性和完整性,有利于规范科研数据的保存和再利用行为,创造更多的社会价值。

3.3 科研数据生命周期中各环节出台的政策完备程度不同

英国数据服务中心^[49](UK Data Archive)将

科研数据生命周期划分为六个阶段:数据创建、数据处理、数据分析、数据保存、数据共享以及数据再利用,而数据保存又包括格式转换、数据迁移、备份和存储、元数据和文档创建以及数据归档等过程。英国科研数据保存政策虽具有系统性和完整性,但从科研数据生命周期的划分来看,不同院校对生命周期各阶段的重视程度也不同,所出台的政策完备程度也不尽相同。谢菲尔德海兰姆大学^[35]将保存政策包含在科研数据管理政策之中,其内容多分散在数据共享、再利用以及责任与权益的规定中,且关于数据创建、处理和分析的内容也较少。还有的院校单独制定保存政策,但各个阶段政策的详略也存在很大差异,如伦敦国王学院^[30]主要介绍了科研数据管理机构及其相应责任,爱丁堡大学^[47]主要规定了数据安全与隐私。由此可以看出,有些院校的保存政策内容并不全面,没有全面考虑数据生命周期和保存过程的各个阶段和环节,容易造成数据损坏,不利于数据再利用。

3.4 学科科研数据的差异性促使保存政策进一步细化和完善

调研发现,英国高校保存政策虽然涵盖的内容较广,但大多都未考虑到学科之间科研数据的差异性,尤其是在数据格式和元数据描述方面。科研数据具有类型多样、格式各异、关联复杂等特性,对于元数据的结构设计和标准的制定应考虑到不同学科的差异。针对不同学科项目的专有数据,如果不能在描述性数据或者文本数据等通用型数据之内表述其特征,则应建立相关学科的保存规定或指南,以实现数据的有效保存和链接。在元数据描述方面,多数高校仅对元数据的描述内容进行要求,而对元数据的创建标准、创建时间、元数据类型以及互操作等问题均未明确说明。使用不同标准的元数据会给研究者带来数据转换等不必要的麻烦,有悖于科研数据政策促进共享的宗旨。在数据保存格式方面,多数院校仅强调使用兼容性、持续可读性且有利于永久性保存的格式进

行存储,但并未明确具体数据保存格式,这势必会造成数据共享和利用方面的困难。因此,学科科研数据的差异性要求保存政策进一步细化和完善。

4 借鉴与启示

我国目前对于科研数据保存政策的研究尚不深入,还未形成基于从数据创建、保存到利用的完整生命周期的管理和保存政策。本文在了解英国科研数据保存政策的基础上,提出如下建议。

4.1 科学设定保存内容和保存期限

英国高校从科研数据内容的价值、创建方式和应用范围等方面制定相关政策,既包括宏观概念,也包含具体的数据内容,划分方式值得我国借鉴。我国在保存政策的制定中,首先应考虑到学科特点、学科发展以及交叉学科之间的相互关系等因素,认识到不同研究数据的保存价值,对科研数据进行选择性保存,让有限的资金和保存空间得到最大程度的利用。其次,科学设定科研数据的保存期限,一般性科研数据保存一定时间之后,应及时进行审核或更新,对于没有任何研究价值的科研数据或者僵尸数据,则按规定进行销毁和处理;对于特殊学科的研究数据,如临床医学、生物学等具有宝贵研究价值或不易获取的学科数据,可根据需求保存20年或者更久。科学设定保存内容和保存期限可以有效保存科研数据,合理运用有限的保存空间,实现保存效益的最大化。

4.2 统一规范元数据标准和数据格式

英国高校的保存政策虽然值得我国借鉴,但是在某些方面仍然缺乏顶层设计。如前文有关科研数据描述和数据保存格式的规定,存在各个院校各自为政的局面。因此,我国在借鉴英国的保存政策时,应在系统全面的调研基础上,由权威部门对元数据描述标准及格式进行

顶层设计,除传统的DC元数据之外,还应该根据不同学科的特点,创建各学科的专业词表,制定元数据标准和使用规范,以便日后进行数据共享和再利用,同时也有利于数据的关联与集成、描述与可视化,确保科研数据的长期保存和获取。

4.3 积极搭建机构知识库数据保存环境

中国科学院文献情报中心对“机构知识库”的定义是:“机构知识库是研究机构实施知识管理的工具,是机构有效管理其知识资产的工具,也是机构知识能力建设的重要机制。”^[50]该定义将机构知识库从收集和使用层面上升到管理和制度层面,范围从机构的“知识产品”扩展到“知识资产”。随着机构知识库影响力的不断提升,越来越多的资助者、出版商以及高校青睐利用机构知识库保存数据,实现数据共享。我国近年来机构知识库的建设也日趋完善,可以为科研数据提供良好的保存环境。但是目前我国高校机构知识库对科研人员数据操作行为的规范还不够明确,如数据上传、删除、修改、获取和共享等。较具有代表性的机构知识库,如中国科学院文献情报中心的NSL OpenIR,北京大学的PKUIR知识库,清华大学以及中国人民大学的学科知识库等,虽然已经具有一定的影响力,但还未形成成熟的保存政策,仅有北京大学在检索界面中明确给出使用指南和规定^[51]。因此,我国应积极搭建科研数据的保存环境,制定有效的政策规范科研数据的创建、保存、共享和再利用行为。

4.4 严格明确数据管理中的责权利

英国科研数据政策的制定都是经由学校研究通过,由专门机构负责管理,如上文提到的PI,同时明确规定管理部门、服务部门、数据创建者和使用者的责任和权益。责权利的清晰界定有利于各部门和角色协调一致,有效保存和管理科研数据。我国高校在科研数据管理方面,一般由研究者独立承担,经常会出现权责不明、

版权纠纷、数据管理不善等问题,且研究者在研究完成之后,对科研数据的管理和保存工作不够重视,导致大量有价值的研究数据损坏和丢失。因此,有必要采用英国的管理方式,成立专门的数据管理部门,不仅有利于对科研数据统一集中管理,而且能够更好地保存科学研究成果,搭建交叉学科之间数据交流的平台,挖掘研究创新点。其次,成立专门的数据管理部门,代表学校与其他院校或科研机构进行交流与联系,促进科研数据联盟或者统一平台的建立,扩大数据的使用范围和利用效率。此外,明确学校、管理部门、研究者和数据使用者的权责关系,按照保存政策的要求规范自己的行为,以更好地推动科研数据的共享和再利用。

4.5 建立和完善版权保护和隐私安全政策体系

通过对英国高校保护政策的分析,可以发现:大多数院校都提出了科研数据的版权归属和隐私安全问题,但政策中规定的内容并不全面,不是以一个独立完整的版权保护和隐私安全政策体系出现的,而是零星散落在数据管理、知识产权、信息安全和个人隐私等政策中。其

次,由于科研数据具有不同于其他智力成果的周期性和重复利用性,在版权归属和安全使用方面,必然与一般的智力成果存在差异,不能简单地运用知识产权等相关政策内容进行规定,所以采用不健全、不配套的版权和隐私安全保存政策,会给科研数据的保护和再利用带来不必要的困扰和隐患。因此,我国在建立保存政策时,应针对科研数据的特性建立配套的版权保护和隐私安全政策体系,或是对原有的知识产权、信息安全政策进行修改和完善,明确科研数据的保存规定。此外,针对资助合同中对科研数据的版权和隐私有特殊规定的,要单独制定特殊的条例或协议,这样科研数据的版权和隐私保护就可以做到有章可循,有法可依。

总之,保存政策可以为规范科研数据的管理和利用提供必要的保障,应受到社会各界和研究者的高度重视。高等学校作为科研数据产生的主力机构,应提高科研数据的保存意识,积极探索科研数据的保存需求,制定符合我国实际的保存政策,推动科研数据保存工作,促进科学研究良性发展。

参考文献

- [1] Milner I. A UK research data service (UKRDS): the way forward for research data management? [J]. Serials, 2009, 22(1): 83-85.
- [2] Simberloff D, Barish B C, Droegemeier K K, et al. Long-lived digital data collections: enabling research and education in the 21st century [R]. Technical Report NSB-05-40, National Science Foundation, Washington D C, USA, 2005.
- [3] ShilobaGbaje E, Mohammed Z. Digital preservation policy in National Information Centres in Nigeria [J]. The Electronic Library, 2013, 31(4): 483-492.
- [4] Santos H M, Flores D. Digital preservation policies for archival documents [J]. Perspectivas em Ciência da Informação, 2015, 20(4): 197-217.
- [5] Sheldon M. Analysis of current digital preservation policies [EB/OL]. [2016-03-05]. <http://blogs.loc.gov/digitalpreservation/2013/08/analysis-of-current-digital-preservation-policies-archives-libraries-and-museums/>.
- [6] Moore R. Towards a theory of digital preservation [J]. International Journal of Digital Curation, 2008, 3(1): 63-75.

- [7] Schottlaender B E C. The digital preservation imperative; an ecosystem view[J]. Library Resources & Technical Services, 2014, 58(1): 2-3.
- [8] 杨云秀, 顾立平, 张瑶, 等. 国外科研教育机构数据政策的调研与分析——以英国10所高校为例[J]. 图书情报工作, 2015, 59(5): 53-59. (Yang Yunxiu, Ku Liping, Zhang Yao, et al. Research on data policies of research and education institutions; case study of British universities[J]. Library and Information Service, 2015, 59(5): 53-59.)
- [9] 周晓燕, 宰冰欣. 澳大利亚高校科研数据保存政策分析[J]. 图书情报知识, 2016(2): 122-128. (Zhou Xiaoyan, Zai Bingxin. Analysis on the research data preservation policies in Australian universities[J]. Knowledge, Learning & Management, 2016(2): 122-128.)
- [10] 司莉, 李月婷. 国外科学数据保存政策的分析与启示[J]. 信息资源管理学报, 2014, 4(2): 45-50. (Si Li, Li Yueting. Analysis and revelation on research data preservation policy abroad[J]. Journal of Information Resources Management, 2014, 4(2): 45-50.)
- [11] 丁培. 国外大学科研数据管理政策研究[J]. 图书馆论坛, 2014, 34(5): 99-106. (Ding Pei. Data management policy for scientific research in overseas universities[J]. Library Tribune, 2014, 34(5): 99-106.)
- [12] 颜颇邓邓, 高峰. 英美澳科学数据存储与共享平台建设现状调查及启示[J]. 图书馆建设, 2016(3): 29-34. (Wayan Dengdeng, Gao Feng. Study on the status and enlightenment of scientific data storage and sharing platform in England, America and Australia[J]. Library Development, 2016(3): 29-34.)
- [13] Permanent access to the records of science in Europe[EB/OL]. [2016-03-05]. <http://www.parse-insight.eu>.
- [14] Digital preservation policies study[EB/OL]. [2016-03-11]. http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf.
- [15] Adrian B M. National geological and geophysical data preservation program: successes and lessons learned[C]// AGU Fall Meeting Abstracts, 2014, 1: 3723.
- [16] How we curate data[EB/OL]. [2016-03-07]. <http://www.data-archive.ac.uk/curate/preservation-policy>.
- [17] 刘霞, 饶艳. 高校图书馆科学数据管理与服务初探——武汉大学图书馆案例分析[J]. 图书情报工作, 2013(3): 33-38. (Liu Xia, Rao Yan. Scientific data management and service in university library; a case study of Wuhan University Library[J]. Library and Information Service, 2013(3): 33-38.)
- [18] 高校科学数据共享平台[EB/OL]. [2016-06-12]. <http://sdm.lib.whu.edu.cn/jspui/aboutus.jsp>. (Scientific data sharing platform[EB/OL]. [2016-06-12]. <http://sdm.lib.whu.edu.cn/jspui/aboutus.jsp>.)
- [19] 茹加, 王晓辉. 数据密集型网格示范站点建设与应用取得良好效果[J]. 中国科学院院刊, 2012(4): 16. (Ru Jia, Wang Xiaohui. Data intensive grid demonstration site construction and application has achieved good results[J]. Bulletin of Chinese Academy of Sciences, 2012(4): 16.)
- [20] 胡永生, 刘颖. 基于用户调查的高校科学数据管理需求分析[J]. 图书情报工作, 2013, 57(6): 28-32. (Hu Yongsheng, Liu Ying. Demand analysis on scientific data management in universities based on user survey[J]. Library and Information Service, 2013, 57(6): 28-32.)
- [21] 谢春枝, 燕今伟. 国内外高校科学数据管理和机制建设研究[J]. 图书情报工作, 2013, 57(6): 12-17. (Xie Chunzhi, Yan Jinwei. Study on construction of scientific data management mechanism in universities at home and abroad[J]. Library and Information Service, 2013, 57(6): 12-17.)

-
- [22] UK institutional data policies[EB/OL].[2016-03-17].<http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>.
- [23] Policy to govern the management of research data [EB/OL].[2016-02-03].http://www.dundee.ac.uk/media/dundeewebsite/library/documents/pdf/2015_Data_Management_Policy.pdf.
- [24] Research data management policy[EB/OL].[2016-02-03].http://www.reading.ac.uk/web/FILES/reas/RDM_Policy_1-0.pdf.
- [25] University guide to research data management[EB/OL].[2016-03-17].<http://sitem.herts.ac.uk/secreg/upr/pdf/IM12-apx%20III-University%20Guide%20to%20Research%20Data%20Management-v03%201.pdf>.
- [26] Research data policy guidance [EB/OL].[2016-02-03].<http://www.bath.ac.uk/research/data/policy/research-data-policy-guidance.html>.
- [27] Research data management policy[EB/OL].[2016-03-17].<https://www.dur.ac.uk/resources/research.office/RDMPolicy.pdf>.
- [28] Policy on the management of research data and records[EB/OL].[2016-03-15].http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/local/sites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf.
- [29] Research data policy[EB/OL].[2016-03-21].http://www.northampton.ac.uk/wp-content/uploads/2015/11/research_data_policy.docx.
- [30] Research data management policy[EB/OL].[2016-03-12].<http://www.kcl.ac.uk/governancezone/Research/Research-Data-Management-Policy.aspx>.
- [31] Code of practice for research[EB/OL].[2016-03-19].<http://www.birmingham.ac.uk/Documents/university/legal/research.pdf>.
- [32] 张民.科研数据的迁移和保存元数据研究[D].兰州大学,2013.(Zhang Min. Research on migration and preservation metadata of scientific data[D].Lanzhou University,2013.)
- [33] University of Manchester research data management policy[EB/OL].[2016-03-14].<http://www.library.manchester.ac.uk/services-and-support/staff/research/services/research-data-management/policy/>.
- [34] Research data policy [EB/OL].[2016-04-06].<http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>.
- [35] Research data management policy[EB/OL].[2016-03-17].<https://www.shu.ac.uk/research/ethics-integrity-and-practice/research-data-management-policy>.
- [36] Section IV; research data management policy[EB/OL].[2016-03-18].<http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html>.
- [37] Research data management and open data policy guidance[EB/OL].[2016-03-07].<https://data.bris.ac.uk/rdm-policy-guidance/#C10>.
- [38] Choosing file formats[EB/OL].[2016-03-10].<http://www.ucl.ac.uk/library/research-support/research-data/best-practices/guides/formats>.
- [39] Research data management policy[EB/OL].[2016-03-15].<http://www.arts.ac.uk/media/arts/research/documents/UAL-Research-Data-Management-Policy-2014.pdf>.
-

- [40] Research data management policy [EB/OL]. [2016-03-16]. <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>.
- [41] University of Leeds research data management policy [EB/OL]. [2016-03-16]. <http://library.leeds.ac.uk/research-data-policies>.
- [42] Open access research and research data management policy [EB/OL]. [2016-03-17]. https://ore.exeter.ac.uk/repository/bitstream/handle/10036/4280/OA_RDM_Policy_Final.pdf?sequence=4.
- [43] UK data service; about us [EB/OL]. [2016-04-90]. <https://www.ukdataservice.ac.uk/about-us>.
- [44] Storing & preserving data [EB/OL]. [2016-03-19]. <http://www.ucl.ac.uk/library/research-support/research-data/best-practices/guides/storing>.
- [45] Research data management policy [EB/OL]. [2016-03-17]. <https://www.liverpool.ac.uk/media/livacuk/computingservices/regulations/research-data-management-policy.pdf>.
- [46] Research data management [EB/OL]. [2016-03-20]. <http://www.nottingham.ac.uk/research/research-data-management/creating-data/policies.aspx>.
- [47] Research data management policy [EB/OL]. [2016-04-10]. <http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy>.
- [48] Public sector data sharing; guidance on the law [EB/OL]. [2016-04-10]. <http://www.gov.scot/Resource/Doc/933/0016820.pdf>.
- [49] Research data lifecycle [EB/OL]. [2016-04-10]. <http://www.data-archive.ac.uk/create-manage/life-cycle>.
- [50] 关于中国科学院文献情报中心机构知识库 (NSL OpenIR) [EB/OL]. [2016-08-03]. <http://ir.las.ac.cn/guiter?id=1>. (About National Science Library, Chinese Academy of Sciences institutional repository (NSL OpenIR) [EB/OL]. [2016-08-03]. <http://ir.las.ac.cn/guiter?id=1>.)
- [51] 北大机构知识库使用手册 [EB/OL]. [2016-04-12]. http://ir.pku.edu.cn/help/index_zh_CN.html. (Institutional repository handbook of Peking university [EB/OL]. [2016-04-12]. http://ir.pku.edu.cn/help/index_zh_CN.html.)

王 琼 北京师范大学政府管理学院研究馆员。北京 100875。

曹 冉 北京师范大学政府管理学院硕士研究生。北京 100875。

(收稿日期:2016-04-27;修回日期:2016-06-20)