

科研模式变革中的数据管理服务: 实现开放获取、开放数据、开放科学的途径

顾立平

摘要 随着信息技术的发展与普及,全球学术交流模式总体朝着开放获取的趋势发展,信息时代的科学活动需要更多的数据开放和共享。开放数据是公众参与科学、推动社会进步的必要条件。无论是对于图书馆工作还是对于信息服务、知识服务来说,科研模式的变革都是一种新的挑战。与此同时,在此背景下,也产生了一种新的机遇,即与科研数据管理相关的一系列新内容,包括:明确科研数据管理各利益方;明确数据管理服务的主要内容;塑造科研数据开放共享文化,开展各个利益相关者之间的数据对话;进行科学数据开放共享各个环节的政策研究;明确数据馆员发挥的作用和职业价值。表6。参考文献70。

关键词 科研模式变革 数据管理服务 开放获取 开放数据 开放科学 数据对话

分类号 G252.8

Data Management Services in the Transition of Research Model: An Approach of Implementing Open Access ,Open Data and Open Science

KU Liping

ABSTRACT

With the development and popularization of information technology, the global academic exchange model has generally developed towards open access. Scientific activity in the information age requires more data to be opened and shared. From open access to open data and to open science, each step is fulfilled with innovation and challenge.

Whether it is for library or for information service and knowledge service, the transition of research model is a new challenge. At the same time, a new opportunity is created in this environment, which is a series of new contents related to the research data management, including: 1) To determine each stakeholder in data management, including research funding institutions, research education institutions, research teams and their members, data centers of key project, Cross-border Internet operators, international publisher/association or society publishing houses and libraries, information centers, informationization centers, and so on. 2) To determine content of data management, including data creation, data processing, data analysis, data preservation, data acquisition and data reuse. 3) Building open and sharing culture of research data and promoting data conversation among stakeholders. 4) To conduct policy research on scientific data openness, and conduct data management standards during life cycle of scientific research data. Scientific data open policy including, but not limited to state administrative level policy, research

通信作者: 顾立平, Email: gulp@mail.las.as.cn, ORCID: 0000-0002-2284-3856 (Correspondence should be addressed to KU Liping, Email: gulp@mail.las.as.cn, ORCID: 0000-0002-2284-3856)

2018年11月 November 2018

funding institution policy, research education institution policy, library and data center policy, publisher group and academic journal policy, and policy on institutional repository, academic repository and data repository, data management plan, data management service, data curation policy, and data level metric. Data management standards during life cycle of scientific research data have five nodes (data backup, data gathering, data selection, data collection, data sharing), four policy categories (project data, multi-source data, fixed elements and variable elements), and three simple implementation steps (data deposit layer, data storage layer, data usage layer). 5) To definite role and value of data librarian. Data librarians implement data management service with professional knowledge, Library Science and Information Science knowledge, and promote open access, open data and open science. 6 tabs. 70 refs.

KEY WORDS

Research mode transition. Data management service. Open access. Open data. Open science. Data conversation.

0 引言

随着信息技术的发展与普及,全球学术交流模式总体朝着开放获取的趋势发展。信息时代的科学活动若要更好地进行,更好地促进科研成果的利用并保护相关权益,需要采取更为广泛的数据开放共享战略。从开放获取到开放数据再到开放科学,各个环节都充满变革与挑战。无论是对于图书馆工作还是对于信息服务和知识服务来说,科研模式的变革都是一种新的挑战,与此同时,在这种大背景下,也产生了一种新的机遇,即与科研数据管理相关的一系列新内容。

信息时代的发展使得科研人员的研究、教学和学术交流方式正在发生重大转变,科学研究的对象也日益复杂,现有的科研方法和手段面临新的挑战。在一系列科研创新的改革当中,力求在不同国家、不同地区的科研机构 and 科研人员之间加强交流,避免重复劳动以及更妥善地进行科研资源配置等成为全球科研社群以及国家政府的核心政策之一。普遍的认识是人们应当充分利用数字化科研带来的机遇,最大程度地加强科研机构和科研人员之间的交流与协作,这要求在各个方面调整工作方式,其中的

关键之一即是数据管理服务。

数据管理服务的核心是数据权益,首要问题是明确科研数据管理的各利益方,包括科研成果的创作者、科研数据贡献者、科研团队和成员,以及政府行政机构、科研资助机构、科研教育机构、出版团体、图书馆界乃至跨界而来的网络运营服务商;比如,数据提交者的科学贡献度的认可机制,不只是某类科学家普遍性的个人诉求,而是应当放在众多利益相关者权益博弈的不平衡的框架上去看待这个问题。其次,明确数据管理服务的主要内容,包括创建数据、处理数据、分析数据、保存数据、访问数据以及明晰重复使用的数据在不同学科之间的共性和差异等。再者,在实施细节上,为更好地实现数据管理,有效提升工作效率,需要明确科学数据开放共享的政策研究。政策是为了鼓励或者抑制人们的某些行为,因此,如果从科研数据生命周期的角度进行考虑,需要对数据备份、数据采集、数据遴选、数据典藏、数据共享各个环节的政策进行研究。最后,在政策执行层面,要明确协助执行的具体人员,协助的内容,以及协助者应该具备的知识技能等,在这方面需要明确数据馆员发挥的作用和职业价值,并且随着社会经济发展和科研模式的需求变化,以及组织单位的战略调整,同时进行调整。

1 科研模式变革

1.1 开放获取、开放数据与开放科学的发展趋势

科研模式变革已经不仅仅是一种特定的工具技术升级或者更换,更是适应全球政治经济局势发展演进中的科技政策和管理机制的变革。目前,通过激励科研创新带动社会经济发展的国家战略主要有两种:第一种是知识产权的所有权控制战略,主要是利用文献、专利、软件著作权等方式,给予创作者财货和名誉上的回报,作为激励科研创新的途径。核心取决于中介者的权益管理。第二种是开放共享的主动权控制战略,主要手段是通过科技论文的开放获取、科研数据的开放共享、科研工具的开源发布、资金人员的众筹众包等途径,使创作者快速扩大影响力和宣传力,从而拥有更多的跟进者、参与者、投入资助等,以此作为激励科研创新的途径。它体现在标准规格的制定、价格成本的谈判、版本的迭代更替等多方面、全领域、各途径的组织动员能力,其根本核心是全球科技交流制度重构中的话语权和制定权,核心不再取决于中介者的权益管理,而是由众多中介者集体参与的权益博弈。

在过去很长一段时间内,全球科技交流制度主要由美国、西欧等少数科技发达国家和地区政府行政机构和科研资助机构掌握。近期,欧盟研究理事推动了科学出版物和科研数据的开放获取^[1-2],该战略是一种有别于二战以及冷战时期延续的科技传播交流机制,即:将论文及数据存储于机构知识库或指定的数据知识库中,或者通过发表在开放获取期刊或存储于数据中心,实现开放共享;同时,以用户明确同意或者知晓默认的方式,按照发布者和中介者所要求的使用目的、途径、场景、方式等给予作者一定形式的回馈,来界定数据权益的边界和使用范围。此种交流机制下,一方面,创作者的权益得到更完整的保护;另一方面,对局势和战略有着深刻理解的组织机构能够更快速、更大范围地掌握主动权、主导权、话语权甚至制定权。

当前,欧盟、美国、日本纷纷出台了相应的开放科学战略。日本内阁府2018年6月15日发布的《创新集成战略》指出,充分利用大数据、人工智能、云计算等,进行跨界整合,实施基于数据共享平台的数据驱动发展战略,该战略是为了支持 Society 5.0 国家发展战略的实施^[3]。日本机构知识库联盟提出的《关于创新集成战略的理解报告》强调,从机构知识库和数据管理服务角度,可以支持该战略第二章“知识泉源”的数据基础设施^[4]。该联盟由日本国立情报研究所牵头组织,从图书馆与信息中心的角度为国家发展战略提供科技信息政策建言以及组织等相关工作。在美国执行“开放政府”计划之后,美国国家科学院于2018年7月17日发布了《设计开放科学》报告^[5],该报告组织学科专家进行研讨,对美国社会和政府提出政策建议:开放科学的基础是科技论文开放获取与科学数据开放共享,需要充分利用已有的大数据共享平台和措施,促进更多科研人员从事科学数据管理计划和科学数据共享,结合数据科学和公众科学等,为社会经济创造新的知识和科研价值。事实上,美国高校图书馆以及研究型图书馆已经从事了相关工作,例如,美国图书馆协会受谷歌公司赞助发布了图书馆开放代码馆藏(Libraries Ready to Code)的测试版^[6],加利福尼亚数字图书馆与跨界互联网运营服务者开展了数据共享和保存试点项目^[7]。欧洲开放科学云(European Open Science Cloud, EOSC)于2018年2月发布政策报告^[8],指出在知识经济和全球数据价值链背景下,开放科学政策起到了促进作用,提出了在科研部门、公共部门、私营部门和公民社会的四重螺旋中自由流动的与数据信息、知识相关的整合政策及各种具体措施。值得注意的是,每年资助经费合计约76亿欧元(约人民币616亿元)的欧洲11个国家科研资助组织,于2018年9月4日推出一项规定,要求到2020年,它们所资助科研项目的每篇论文发表后立即开放获取^[9-10]。从英国、荷兰、法国的国家资助机构以及意大利国家核物理研究所等11所研究机

构获得经费资助的科研人员,其研究成果不能发表在有6个月或12个月开放获取时滞期的期刊、混合开放获取期刊和非开放获取期刊上。这些科研人员将不得不放弃在 *Nature* 和 *Science* 上发表论文,除非这些期刊改变自己的商业模式。

从情报分析的角度看,我们可以发现这些趋势的外在规律:随着时间推移,出现了从合同内容的一系列变化到越来越多的开放许可协议的出台,以及愈来愈高级别的政策颁布,乃至法律制定和体制改革的现象,说明开放科学政策一直在发生改变,相应的数据管理服务也在进行各种调整。然而,从政策研究的角度看,我们能更深入地找到它的内在逻辑,即在无法避免数字化复制、互联网传播、修改删选极为便利的时代里,科研人员对于科研成果的开放程度越高,就越可以得到更好的传播和利用,对于自身利益的保护就越全面;而且,科研资助机构和科研教育机构可以进行更效率的科研管理,从而吸纳全球范围内的科技人才,这也是为什么各组织机构要控制开放共享主动权的原因之一。越

来越多证据显示,国外图书馆和信息中心的专家参与了这些政策建议与实施方案的制定和实现。

1.2 科研模式的数字化、开放化和社群化发展

开放科学是一个较为广泛和不断发展的概念,它是促进科研人员之间合作的一种新的科研方式,使得研究人员可以将自己的研究成果中的每个要素与他人共享^[11]。开放科学所包含的内容比较全面和广泛,如“Data Sharing”“Open Research Data”“Science 2.0+”“Citizen Science”“Open Innovation”等^[12]。英国皇家学会在2012年5月发布的《科学:开放事业》报告中指出,开放科学不仅能有效促进科学发现,还有助于发现、遏制和清理“坏的科学”,开放性促进系统化的科学诚信,如果开放数据能被置于合适的治理框架下,这将符合公共利益^[13]。Web 2.0时代的到来,使得信息的传播方式由单向输出变为双向传播,科研模式的发展产生了各种各样的组合配置:传统模式、现代模式、开放科学模式和创新模式(见表1)。传统科研模式是研究

表1 科研模式的数字化、开放化和社群化发展

模式	环节					
	发现	分析	创作	出版	扩展	评价
传统模式	权威文摘索引系统(如WoS)	商用套装软件(如SPSS和SAS等)	文书软件和文献管理(如MSWord和Endnote)	顶级期刊(如 <i>Nature</i> 和 <i>Science</i> 等)	作者唯一识别符(如Research ID)	期刊影响因子
现代模式	搜索引擎(Google Scholar)	开源软件(如R和Python)	云端办公系统(如Google Docs和RefWorks等)	预印本和完全开放获取期刊(如ArXiv和PLoS)	机构知识库(如ir.las.ac.cn)	期刊级别计量、论文级别计量(如Eigenfactor和ALM等)
开放科学模式	开放资源汇聚池(如Paperity)	开放分析平台(如ROpenSci Science Exchange Zooniverse)	云群组协作和共享权限(如Hypothes.is和Google Drive和Zotero等)	数据、代码、软件共享(如F1000Res、PLOS ONE和Figshare、Dryad等)	唯一标识符和文献关联推荐(如ORCID、Research Gate、SlideShare等)	公众评估(如Publons和ImpactStory等)
创新模式	科研群组共享模式	科研众包(如Zooniverse和Hivebench)	群体协作(如Authorea和Collwiz)	课件共享与论文草稿注册(如Figshare和PeerJ等)	唯一标识符的开放接口运用(如Research Gate和ORCID等)	学术与社会影响力评估(如Peerage of Science和Altmetric)

(据 Innovation 101 相关资料整理)

者将发现的信息进行分析研究后撰写论文并发表在优质期刊上,将其作为一种科研产出或一种研究记录,之后送交权威性的机构进行科研评审。现代模式最大的变化是这种传统意义上的“出版即完成”正在转变为“发布才开始”,未来的科学研究者所要考虑的是其内容的存储位置、开放程度以及对全社会的贡献大小^[14]。这也是真正意义上的革命性变革,是新时代的新挑战。开放科学模式是指研究者将产生的科学数据呈现在互联网上进行开放共享,众筹、众包、共享经济是开放科学模式的典型代表^[15]。创新模式则依靠大众的力量来获取数据,并进行深入研究,这是另一种突破传统模式的思考方式^[16]。

当前,正在发展和逐步完善的一种机制是“同行评审评价”(Peer Review Evaluation)模型^[17],当新的文章出现或发布时,不单单依靠少数的专家进行长达3个月到1年的审稿,而是直接将其投放到互联网上,请具有相关资格的专家或者拥有相关专长的工作人员进行评述,并对这些评述进行分析,来判断哪些人可以真正成为专业的评论员,这些评论也可以成为最终科研成果的一部分。

OpenAIRE的调查内容提出,未来科研模式的发展趋势之一是Open Peer Review^[18]模式,它所涉及的具体概念有以下几点:第一,开放身份,即作者和评论者知道彼此的身份,了解彼此;第二,开放报告,是指评论报告与相关文章一起发布;第三,开放参与,需要更开放的社区为评审过程提供平台;第四,开放式互动,允许并鼓励作者和评论者之间,评论者与评论者之间的互惠式讨论;第五,开放预审稿,手稿即刻创作,在任何正式的同行评审程序之前可供开放预审(例如通过像ArXiv这样的预印本服务);第六,开放最终版本评论,审查或评论最终的“记录版本”出版物;第七,开放平台,由出版地不同的组织实体促成,将评论与出版先做脱钩再做连接。

科研模式的数字化、开放化和社群化发展,

不仅仅是文献和文献信息的数字化,或资源开放利用以及通过社交网络的传播,而且是在知识创造的阶段就开始了;知识管理的方式也不遵循创作、生产、分配、传播、存储的流程,而是环绕文本(Text)本身所处的当前情境(Context)进行知识、信息、数据的再创作、再生产、再分配、再传播、再存储,选择最佳方案开展。

2 科研数据管理

2.1 科研数据管理所涉及的利益相关者

所有数据皆有其价值,但是相对于科学数据(Scientific Data)而言,科研数据(Research Data)是经过处理且在当时被视为具有高度价值的科学数据。科研数据也是数字化科研时代的重要战略资源,为了更好地管理科研项目的成果,使其得到更好的保护和传播,各国都先后出台了国家层面的数据管理政策及管理机构实施办法,学术领域就相关问题达成了共识,科研团队也积累了丰富的实践。我国已经在科研数据管理方面进行了许多尝试,积累了一定的经验和基础,并于2018年3月出台了国家层面的指导意见《科学数据管理办法》^[19],目前,各相关机构正在依照此管理办法进行实施方案的设计、运行、执行和评估考核。

在实践层面上,主要分为两种范式。第一种是数据集中式管理,需要提高数据管理的性能,比如分层分级的办法、人员制度和考核、数量质量的指标,其着力点是提高单个机关的组织动员能力;第二种是数据分布式管理,需要提高节点之间互联互通的能力,比如团队乃至不同院所之间的协作、标准、规范以及分工,其着力点是提高数据之间互相流动的能力。如果仅仅考虑数据的采集和利用,集中式管理下的数据互联互通能力就会略显不足;如果顾虑多个利益主体在发布和共享过程中的诉求,分布式管理下的组织动员效率就会大打折扣^[20-21]。

理想情况下,既希望组织机关具有数据质

量和数量的资产增长,又希望协作网络能够畅通运行,促进数据利用和共享;而现实情况是,数据权益主体或者利益相关者的诉求全部满足才能达到组织动员和协作能力的最佳效果。诉求包括两个方面,一是满足激励因素,二是消除阻碍因素。因此,科学数据管理政策是很有必要

参考的框架。然而在实际处理科研数据管理服务的具体问题上,不得不考虑利益相关者的诉求^[22-24]。笔者通过长期文献阅读、政策解读、跟踪扫描,并对国内外各家机构的专家进行访问的基础上,整理汇总出利益相关者的主要诉求(见表2)。

表2 科研数据管理利益相关者的诉求

利益相关者	主要诉求
科研资助机构	促进意愿: 便于科研项目成果和知识产权的有效、便捷、准确管理; 消除疑虑: 在产权不清晰的情况下额外增加资助经费或者难以进行审计。
科研教育机构	促进意愿: 扩大机构在学术地位和社会认可方面的双重影响力; 消除疑虑: 在经费投入和人员权责不清晰的情况下增加无效的行政任务。
科研团队及其成员	促进意愿: 体制内的奖励、晋升、聘任、资质认定等的具体办法与落实; 消除疑虑: 署名权、成果转移转化优先权、无显著结论的数据贡献认可。
重点项目的数据中心	促进意愿: 增加存储在数据中心的高质量数据的总量,发挥国际影响力; 消除疑虑: 数据质量究责的不可追溯性以及可共享数据持续性的不稳定。
跨界的互联网运营商	促进意愿: 在云计算、迁移计算和高性能计算的设施上强化应用面; 消除疑虑: 所保存和托管的数据具有使用边界和权益归属不清的情况。
国际出版商与学/协会出版社	促进意愿: 论文、数据、代码、附加材料等数字对象成为新的产品服务内容; 消除疑虑: 负责的创新服务增加了额外义务,但是缺乏政治经济方面的明确支持。
图书馆、信息中心、信息化中心等	促进意愿: 成为国家数据战略的一部分并且产生新的职业发展前景,对知识技能进行转型和升级等; 消除疑虑: 科研数据管理服务的“脏活(Dirty job)”受到蔑视和忽视。

数据管理服务的首要任务是了解各方的核心利益。在拥有共同目标即开放科学或者类似理念的基础上,公平合理地满足各方诉求,才能尽可能地平衡利益相关方的权益,尤其要关注数据生产者、贡献者和提交者的核心权益;确保根据各个利益相关者的核心利益所设计的政策实施方案能有效落实,并且据此产生和设计的科研数据管理服务会有相对可靠的用户需求和可检验的任务规划。

2.2 科研数据管理的核心要点

基于数据管理的数据服务主要包括数据创建、数据处理、数据分析、数据保存、数据获取和数据重用等六项内容(见表3)。此项清单可以为数据管理提供一定的经验和基础。在此过程

中需要着重注意以下三点内容^[25-30]:第一,科研数据的生命周期,包括备份、采集、遴选、典藏、共享等各个环节;第二,参与者和权益者,包括科研人员、学术社群、机构、研究资助者、出版商、公众等各个利益群体;第三,记录、存储、保存、发布文件的措施,考虑数据重用能力、数据授权许可、开放性、实施条件等。

表3所列科研数据管理的核心要点,是笔者根据连续七年的中国开放获取推介周、连续五年的中国机构知识库学术研讨会、连续三年的中国合理使用周和中国数据馆员研讨班的众多专家就数据共享、数据管理、数据获取、开放数据、开放科学等主题的报告内容,汇聚整理而成^[31],并融汇了笔者在中国科学院大学开设的公共选修课100多位研究生的课题交流内容,

表 3 科研数据管理服务的核心要点

节点	核心要点
数据创建	<ul style="list-style-type: none"> 项目设计中的数据采集、整理、管理和存储的设计; 符合“数据管理计划”的格式、存缴、共享设定的方式; 现有数据的搜寻、探索、预先判断等; 通过实验、观察、测量、模拟等方式可以产生的数据集; 对于数据集的描述,包括如何创建元数据等。
数据处理	<ul style="list-style-type: none"> 数据导入、导出、数据存储与读写的格式转换; 数据的列表、矩阵、数组、框以及行列的转换; 异常值、缺失值、空值的处理依据和实际处理; 数据的去敏感化、匿名化、置换化和加密解密; 数据集、元数据、描述数据; 数据的版本管理、数据的集中式管理或者分布式管理。
数据分析	<ul style="list-style-type: none"> 数据预览、描述统计、假设检验、参数统计; 数据抽取、理论模型测试、比较评估、模型拟合、模型检验; 数据模型、模型解释和数据解释、数据使用说明; 学术论文、数据论文、代码论文; 原始数据、衍生数据、科研数据、发表型数据的管理。
数据保存	<ul style="list-style-type: none"> 数据的原始格式、数据的使用格式、数据的迁移格式; 数据源的载体、数据处理和分析的载体、数据迁移的载体; 数据备份存储、数据异地存储、数据的暗存储; 元数据和文件、元数据以及各级数据集的使用条件; 数据读取权限管理、数据获取权限管理、数据档案管理。
数据访问	<ul style="list-style-type: none"> 数据流的人员管理、数据转移的合约、数据接口的标准; 数据交换的协议、数据的抽取方式与发布方式; 数据访问权限的服务控制、网络信息安全防范、数据节点管理; 数据产权从源头到中介到释放的使用边界与权益确立; 数据公共利用、商用和交换的各种宣传推广。
数据重用	<ul style="list-style-type: none"> 数据发布后的持续性科研成果、数据持续性的产出与增加; 数据产生的新研究、新课题、新应用; 数据产值的估算、计算、核对与审计; 数据版本的溯源、数据质量的评估、数据影响力的评价; 数据素质教育的课程、培训、咨询与线上服务。

以及 2017 年进行的 10 次数据馆员沙龙的交流讨论内容,旨在探索和发现我国科研教育环境中数据管理服务的具体需求。

2.3 科研数据开放共享的价值

对于科研数据的开放共享,最重要的基础是要有共同的目标和价值观,如果没有核心价值观和共识,就很难达成一致,也很难平衡好各

方的核心利益。利用政策驱动知识与技能、技术与人文基础设施发展的同时,促使科研人员的利益得以改变,利用第三方共同推动文化实践,最终实现数据开放的共同目标^[32]。

随着数字经济的发展,科研数据开放共享已经成为一项必要的公共服务,是否在同一机构内的不同团队间(组内)进行共享,是否在更大的业务平台让各方参与者共享,是否能在公

共服务平台给予各方包括竞争对手在内的无差别共享,甚至在日益推进的开放数据运动中,与政府信息一道为公民、公众和国际友人共享等,在考验新时期的智慧。科研数据开放共享使得各个组织充分利用他们的核心资源,产生了基于现存数据的新的洞见,但同时也带来了新的道德伦理方面的风险^[33]。如何识别和降低这类风险,主要依靠两种手段:一种是行业内以及社会中的科研数据开放共享文化,另一种是图书馆或相关机构合作开展各个利益相关者之间的数据对话。

塑造科研数据开放共享文化,在制定使用政策时可参考国外已有的实践和经验^[24,34-36]:一部分来自科研资助机构的倡导,一部分来自组织单位的宣传,也有委托图书馆和信息中心来进行文化建设的。然而,数据共享文化的实际成效,仍然来自学科领域内不断增长的共识。以医学领域为例^[37],数据共享文化的范围至少包括以下方面。

(1) 数据共享能够增强对于单个临床试验结果的理解,并且相对单一研究的数据而言,数据共享促进来自多个试验的数据结果汇集,可能促进进一步的科学发现。

(2) 促进数据共享中心在科研参与者、最小化安全风险以及尊重医学研究作为公共利益本质等方面的道德与伦理讨论,产生新的探索领域。

(3) 数据共享的实践和科学讨论包括:提高研究的精确性,告知治疗方案选择的风险或益处分析,增强合作,加速生物医学研究,恢复对于临床研究企业的信任等,产生新的探索领域。

(4) 数据共享参与者的层级划分尤其重要,避免因数据再分析可能得出的不精确或者错误结论,但是数据再分析也可能产生新的探索领域。

(5) 当产业界、学术界和监管机构的领导者已经意识到不断增长的透明度和数据共享方面的价值时,文化上的转变已经开始,即主要关注数据应该如何被共享,而非数据被共享

的原因。

数据对话是另一种建设数据共享文化的方式。数据对话^[38-39]是让科研人员一起分享他们的数据故事,讨论问题,交换想法。无论是关于数据的创建、收集、使用、处理、共享和管理,还是从他们自己的失败中吸取的经验教训(如数据故事),或是在数据处理方面的成功,数据对话的目的是促进数据共享的经验交流,最终达到数据共享的目的。人们甚至开始考虑这种来自图书馆和信息中心的数据对话,可以作为开放科学的重要组成部分^[40]。因此,科研数据管理服务在这些实践上的缺失,往往是系统项目建设所没有考虑到的核心关键问题,目前国外正在逐步建立这类服务的执行、评估和检验机制,值得国内同行关注和参考。

2.4 科研数据开放共享的政策

科研数据开放共享的一项关键问题是如何促进数据利用并且保护数据权益。数据权益是分析和提出问题解决的核心,需要系统梳理各利益相关者的角色、权益诉求、政策博弈与政策细节,相关研究对象和研究内容包括:国家政策、科研资助机构、科研教育机构、图书馆与信息中心、国际出版商、数据出版、数据知识库、数据管理服务、数据管理计划、数据监管、数据级别计量等。这些政策研究的目的是为了更好地实现对科研数据的有效管理。

从系统性、全面性和完整性的角度看,政策研究可以粗略分为两个部分:一是宏观的政策内容,二是微观的政策实施细节。目前主要包括但不限于以下内容^[41]:

(1) 国家行政级别的政策:数据驱动创新发展战略。

(2) 研究资助机构的政策:资助成果的最大化利用。

(3) 研究教育机构的政策:知识资产的管理与传播。

(4) 图书馆和信息中心政策:开放信息资源整合利用。

(5) 出版团体和期刊政策: 新兴学术交流体系基础。

(6) 机构知识库、学术知识库、数据知识库: 内容长期可持续。

(7) 数据管理计划: 研究工作流程中的数据整理备份。

(8) 数据管理服务: 处理数据的工具、流程、方法。

(9) 数据监管政策: 遴选具有优秀质量和价值的数据库。

(10) 数据级别计量: 承认以及评估数据共享的贡献。

科研资助与教育机构、学术出版界、信息服务机构等学术社群成员目前正在积极联合起来, 将数据治理、科研资助、科研管理与数据服务深度整合, 充分依托机构知识库, 进行科学数据管理, 借此有效地提升数据资源的利用效率, 发挥“数据驱动科研”的重要价值^[42-46]。助力联合行动的有效落实, 一方面需要实践经验, 并且要持续不断地改善和更新, 以及新的需求引导下的新型试验; 另一方面需要国家科技信息政策研究, 以及政策执行层面的咨询与服务。

2.5 科研数据生命周期中的数据管理规范

在科研数据开放共享的政策框架之下, 在执行层面需要将数据备份、数据采集、数据遴选、数据典藏、数据共享作为各自节点考虑, 而不是作为各个环节进行服务设计。即数据采集也需要考虑数据共享, 反之亦然; 不是把两者割裂为一个项目的不同组成部分, 或者一条流水线上的不同分工, 而是同为数据开放共享政策中数据管理服务的一个整体, 只是这项服务是依托数据生命周期的架构进行理论上的拆解, 但在实践层面几乎没有割裂的情况。

国外的相关机构对于科研数据管理服务的经验和实践案例, 可给我们提供借鉴。例如, 加利福尼亚大学图书馆 (University of California Library) 的学术交流、数据策展和科研数据管理方案和计划^[47-48], 约翰霍普金斯大学图书馆 (Johns

Hopkins University Libraries)、密歇根大学图书馆 (University of Michigan Libraries)、罗格斯大学图书馆 (Rutgers University Libraries) 的科研数据服务等内容^[49-51]。笔者在对已有分析框架深度解析的基础上, 又走访了一线科研工作岗位人员, 其对于科学数据服务的反馈结果仍然是过于抽象, 具体反映为宽泛而不接地气, 或者过于细致而缺乏优先次序。所以在政策执行层面的规范化主要有两个方面的考虑: 一是如何细致完整地描述规范, 这有利于提高质量和准确度, 但不利于增加数量和推广力; 二是如何完成框架性的原则内容, 这有利于提高普适性和接纳特殊情况, 但不利于精准核查和量化评估。这两种考虑方案都具有客观性, 都能成立, 而且可以付诸实现, 但是在科研数据管理服务的领域实施, 其协调性往往欠佳。因此, 我们梳理了“五一四一三”方案, 即五个数据生命周期的节点: 数据备份、数据采集、数据遴选、数据典藏、数据共享, 四种政策规范要素: 固定要素、变动要素、项目数据、多源数据 (见表 4), 以及三项简单明了的实施步骤: 数据存缴层、数据存储层和数据使用层。

综上所述, 对于有效的科研数据管理, 首先要明确数据存缴层, 即“人”的问题——数据的创建者/提供者 (Data Creator/Provider), 他的任务应该包括创建数据、搜集数据、处理数据、分析数据和提供可重用的数据。他的权利和责任应该包括^[52]: ①管理数据、存储数据、长期保存数据; ②遵守科研机构与资助机构的政策: 为共享和重用而备受数据; ③符合学科社群与科研机构的标准: 遴选、部署、转换数据; ④考虑数据在未来可视化中的法律与伦理问题。

其次是数据存储层: 知识库 (Repository), 它的任务是监管数据, 提供数据访问, 提供合适的工具以及整合各项应用。知识库应该能够详述数据管理政策, 确保系统的日常维护, 管理数据使之能够长期使用, 保存数据的真实性并符合标准规范, 保护数据提供者的权益, 确保流量用户的合理获取并且提供有效的数据访问、重用的工

表4 科研数据生命周期各节点的规范研究

节点	科研数据生命周期的规范要素	
	固定要素	变动要素
数据备份 ^[53-54]	存储空间方案、数据备份方案、数据还原方案、数据复原方案、财务支持方案	数据安全风险的级别、数据访问控制的手段、数据用户的保护措施、数据迁移方案
数据采集 ^[53, 55]	数据采集标准方法、结构化文件命名方式、版本区别方案、品质控管方案(校准、重复取样或者测量、采集或记录标准化数据、数据输入验证、数据的同行评议、受控词表等)	数据规模预估方案(类型、格式和量级)、数据存储设施(支持数据共享和长期访问的格式与软件)、现存可重用数据
数据遴选 ^[53, 56]	法规支持(合同、法律、监管作业)、财务支持、人员支持、政策支持(知识库或者档案馆)	遴选标准(科研用途、教学用途)、保存时限、销毁条件、保留条件、格式变化、载体变化
数据典藏 ^[53, 57]	元数据标准、数据文件创建、元数据创建、文件与元数据收割方式	数据说明信息(方法、分析和处理的信息、变量定义、词汇、计量单位、假设、格式和数据的文件类型)、元数据标准规范(创造者、贡献者、标题、日期)
数据共享 ^[53, 58]	数据专属使用期限、数据共享协议、数据保密协定、数据共享机制(如机构知识库)、数据共享方式(类型、大小、复杂度和敏感性)	潜在数据用户、数据共享时机、数据记录回溯、数据引用方式、数据标示符(永久识别符、系统识别符)
环节	科研项目数据权益管理的规范要素	
	固定要素	变动要素
项目数据 ^[59-60]	工作小组的数据管理指南、部门的数据保护/安全政策、科研机构的数据管理政策、资助机构的科研数据管理政策、业界的标准作业程序	科研项目的本质、正在解决的科研问题、产生数据的原因
多源数据 ^[61-62]	数据延迟共享条件、数据限制共享条件、数据知识产权(合作研究项目)、第三方数据权限	数据管理修正方案、数据管理评估方案、数据权责分配(合作研究项目)

具和可视化工具,以及提供用户教育培训等。

第三是明确数据使用层,即流量用户(Downstream user/reuser)的角色、权利与责任。流量用户可以搜集数据,通过浏览、下载、抽取等行为获取数据,或者抽取、操控、重用数据、访问元数据以使用、引用和可视化数据。但是流量用户必须遵守许可条款,向数据的提供者致谢并有效地管理衍生数据。

为了解决理论和实践相结合的问题,笔者在长期研究和走访的基础上,以固定要素和变

动要素进行规范要点的主次区分,以及项目数据和多源数据的情境划分,重新界定科学数据权益分析的基本框架,探索在数据治理^[63]中尚未解决的问题。

3 数据馆员的发展

3.1 数据馆员的角色

开放科学的发展趋势以及《科学数据管理办法》的颁布,促进了我国科研数据管理服务的

发展,同时也带动了学术界、教育界、出版界和图书馆界的发展。数据管理服务的核心基础是数据权益,需要从事这项工作的人具有相应的知识、技能、理念和经验,执行各项协作的任务,完成许多尚未预见的法律、信息技术、服务转型等方面的攻关工作。

众所周知,社会经济的发展已从信息时代过渡到大数据时代,人类社会所拥有的数据量和统计量的增长比以往任何时候都要快,科研人员对于“大智移云”(大数据、人工智能、移动服务、云计算云存储)的需求,促使国外许多图书馆和信息中心设置了数据馆员的新职位,在美国和西欧,数据馆员已经成为研究型图书馆的一个重要职位。

图书馆员在科研数据重用和数据中心中发挥重要的作用,有学者基于过去处理 eScience、Science 2.0 乃至 Open Science 的发展历程,对图书馆员在其中发挥的作用进行系统梳理,进而提出了数据馆员的新角色和新定位^[64]。美国图书馆协会(ALA)正在就这一新兴职业开发核心竞争力框架,在 2018 年的国际图书馆大会(IFLA)期间,也有报告建议图书馆和信息科学(LIS)应开设数据馆员的相关课程,以快速促进这个领域的专业化^[65]。有的学者从数据生命周期考虑,基于图书馆员能够做的事情,提出数据馆员可以担当的角色有:机构数据知识库管理人员、政策制定者、数据管理计划咨询者、数据咨询者、数据出版者等^[66];也有学者认为应该是以下角色^[67]:数据管理者(data manager)、数据工程师、数据管家(data steward)、数据监护者、数据科学家、数据档案管理员等。

我国学者^[68]认为,数据馆员应当与原先的图书馆工作相衔接,可以在大数据时代培养几种不同类型的数据馆员,如元数据馆员、科学数据馆员、数据分析馆员等。也有学者^[69]从终端用户的需求考虑,认为在图书馆员走向数据馆员的道路上,可以有选择性地遴选计算机科学与统计科学中与未来发展的数据科学相关的内容进行培养。在 2016 年的中国机构知识库学术研讨会上,笔者曾经做题为《在开放科学中探索数据馆员的发展》的报告^[70],提出核心目标是“促进开放科学政策成真”,三项任务是:①充分了解和利用开放许可协议来协助人们分享和使用文献;②完全熟练掌握数据管理计划来协助人们组织数据;③妥善运用和推广期刊出版规范指南来协助人们选择发表刊物。

3.2 数据馆员的培训

数据科学家、数据公民、数据馆员正在成为创新社会的新兴力量,也是推动开放获取、开放数据、开放科学发展的主要力量。数据馆员的专业培训有助于提升广大图书馆员的职业技能,适应信息社会发展的新态势,满足面向国家科技信息政策的实践工作的需要。笔者通过中国科学院文献情报中心开展的科研模式变革与数据管理服务方面的研究、咨询与实践,根据定期举办数据馆员培训班以及每月举办的数据馆员沙龙的总结和经验,整理数据馆员培训的课程体系,具体应包括基础课程和进阶课程两个方面(分别见表 5 和表 6)。

表 5 数据馆员基础课程

主题	内容
数据政策	数据治理、数据计量、数据馆员、数据出版发布与引用、开放数据的权益问题与管理方式、数字资源的开放内容管理机制
数据存储	DSpace 安装、调试、测试、配置、注意事项 Dataverse 安装、调试、测试、配置、注意事项
数据处理	科学研究与 R 语言、统计(R)、数据处理(R)、可视化(R)
数据分析	结合项目案例分析(R)
机器学习	机器学习经典案例探讨(R)、无监督学习(R)、有监督学习(R)、半监督学习(R)

表6 数据馆员进阶课程

主题	内容
大数据工程	大数据工程基本概论与工具、Hadoop 原理与结构演示
大数据存储	Hadoop 安装、调试、测试(VM)、HDFS 数据导入导出(VM)
大数据计算	MapReduce 概论(VM)、MapReduce 大数据计算(VM)
大数据管理	Hive 数据查询(VM)、Hive 报表导出(VM)
大数据分析	大数据分析基本概论与工具、Python 基础、安装、调试
大数据采集	爬虫与反爬虫的基础知识、网页抓取(Python)
大数据处理	数据读写和操作(Python)
大数据科学	结合项目案例分析
深度学习	深度学习概论与基础、Tensorflow 介绍、安装、基本操作、图像处理(VM) 影像处理(VM)

4 结语

科研模式的变革与科研数据管理为图书馆事业的发展带来了更多的挑战,也提出了更高要求。图书馆是社会发展的基础设施,既要作为公共知识供给的服务平台,又要成为各种科技创新投入施放的表演舞台。科研数据管理是图书馆发展的新兴机遇,图书馆对科研数据的发现、评价与遴选的能力,以及描述、组织、检索或者分析科研数据的能力,在未来至关重要。

在开放获取、开放数据以及开放科学的发展趋势上,未来可能的发展方向依然是持续:①消除知识鸿沟,使得科研人员、公众和企业都可以获得公共资金资助产生的科研成果;②促进知识创新,科研成果公开共享,有利于人们检视、利用和宣传优秀的科研成果;③打破知识高墙,压制付费高墙乃至打破数据共享壁垒。开放共享战略可以吸引优秀人才,可以在数字化时代里掌握规范、规格、国际学术交流的话语权,在这一过程中,不能缺少图书馆和信息中心

的专家的参与。

国家社会经济发展战略和科学数据共享、科研成果开放获取正在紧密关联,数据科学家、公众科学家、数据工程师、数据分析师、数据管理员等正成为新的职业。科研数据管理服务是开放和共享的主要推动力量,尽管人人手上皆有数据,而且都有所谓的数据管理经验,但是整体推进行业乃至社会经济发展却并不容易,在实施过程中可能出现尚未遇到的问题。在促进科研数据管理服务以及数据共享的过程中,关键是要明确目标,清晰权益以及厘清风险,才能使人们更好地认识数据的价值、数据共享的价值以及科研数据管理服务的价值。

数据馆员的职业价值应该得到认可。数据馆员服务于科研数据管理,他们的职责不仅是管理数据,而且是利用学科专业知识和图书馆学情报学的知识协助学科馆员更好地进行数据管理服务;通过数据管理服务,图书馆员能更好地服务科研人员,在科研生命周期的各个节点发挥重要作用,最终促进开放获取、开放数据、开放科学的实现。

参考文献

- [1] Directorate-General for Research & Innovation of European Commission. Guideline on open access to scientific publication and research data in Horizon 2020 [EB/OL]. [2018-10-27]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hioa-pilot-guide_en.pdf.

- [2] The Expert Panel on Open Science based on Global Perspectives , Cabinet Office of Government of Japan. Promoting open science in Japan [EB/OL]. [2018-04-07]. http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_summary_en.pdf.
- [3] 日本内閣府. 統合イノベーション戦略 [EB/OL]. [2018-06-25]. <http://www8.cao.go.jp/cstp/tougosenryaku/index.html>. (Cabinet Office of Government of Japan. Integrated innovation strategy [EB/OL]. [2018-06-25]. <http://www8.cao.go.jp/cstp/tougosenryaku/index.html>.)
- [4] JPCOAR. JPCOAR 統合イノベーション戦略についての見解 [EB/OL]. [2018-08-16]. https://jpcoar.repo.nii.ac.jp/?action=repository_uri&item_id=126&file_id=22&file_no=1. (JPCOAR. Comprehension report on *Integrated innovation strategy* [EB/OL]. [2018-08-16]. https://jpcoar.repo.nii.ac.jp/?action=repository_uri&item_id=126&file_id=22&file_no=1.)
- [5] National Academies. Open science by design: realizing a vision for 21st century research [EB/OL]. [2018-08-16]. <https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century>.
- [6] ALA. ALA's libraries ready to code to release beta collection at 2018 annual conference [EB/OL]. [2018-07-07]. <http://www.ala.org/news/member-news/2018/06/alas-libraries-ready-code-release-beta-collection-2018-annual-conference>.
- [7] California Digital Library. Internet archive , code for science and society , and California Digital Library to partner on a data sharing and preservation pilot project [EB/OL]. [2018-07-07]. <http://blog.archive.org/2018/06/05/internet-archive-code-for-science-and-society-and-california-digital-library-to-partner-on-a-data-sharing-and-preservation-pilot-project>.
- [8] EOSC. Policy landscape review [EB/OL]. [2018-02-21]. <https://www.eosepilot.eu/content/d31-policy-landscape-review>.
- [9] European Commission. 'Plan S' and 'eOAlition S': accelerating the transition to full and immediate open access to scientific publications [EB/OL]. [2018-09-22]. https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/plan-s-and-coalition-s-accelerating-transition-full-and-immediate-open-access-scientific_en.
- [10] Enserink M. European science funders ban grantees from publishing in paywalled journals [EB/OL]. [2018-09-22]. <http://www.sciencemag.org/news/2018/09/european-science-funders-ban-grantees-publishing-paywalled-journals>.
- [11] National Institute of Informatics. NII 2016 catalogue [EB/OL]. [2018-10-27]. <http://www.nii.ac.jp/en/about/upload/catalogue2016.pdf>.
- [12] Foster. Open science [EB/OL]. [2018-10-27]. <https://www.fosteropenscience.eu>.
- [13] The Royal Society. Science as an open enterprise [EB/OL]. [2012-06-21]. <http://royalsociety.org/policy/projects/science-public-enterprise/report>.
- [14] Innovation101. Traditional model [EB/OL]. [2018-10-27]. <http://innoscholcomm.silk.co/page/Traditional>.
- [15] Innovation101. Open science model [EB/OL]. [2018-10-27]. <http://innoscholcomm.silk.co/page/Open-Science>.
- [16] Innovation101. Innovative model [EB/OL]. [2018-10-27]. <http://innoscholcomm.silk.co/page/Innovative>.
- [17] Highwire. Peer Review Evaluation (PRE) and high wire press partner to increase transparency for peer review [EB/OL]. [2018-09-22]. <http://home.highwire.org/news/peer-review-evaluation-pre-and-highwire-press>

- partner-increase-transparency-peer-review.
- [18] OpenAIRE. OpenAIRE survey: open peer review is moving mainstream [EB/OL]. [2018-09-22]. <https://blogs.openaire.eu/?p=1895>.
- [19] 科学数据管理办法 [EB/OL]. [2018-04-02]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm. (Scientific research data management method [EB/OL]. [2018-04-02]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.)
- [20] Ku Liping. Open data policy and best practice [EB/OL]. [2016-12-14]. <http://ir.las.ac.cn/handle/12502/8778>.
- [21] Ku Liping. Research data management and open data in China [EB/OL]. [2016-12-04]. <http://ir.las.ac.cn/handle/12502/8553>.
- [22] 顾立平. 科学数据权益分析的基本框架 [J]. 图书情报知识, 2014(1): 34-51. (Ku Liping. Basic framework of the analysis on research data rights [J]. Document Informaiton & Knowledge, 2014(1): 34-51.)
- [23] DIIS of Australian Government. Boosting innovation and science [EB/OL]. [2018-08-16]. <https://www.industry.gov.au/strategies-for-the-future/boosting-innovation-and-science>.
- [24] Johnson R. Open access policy dependencies: final project report [EB/OL]. [2018-09-22]. http://repository.jisc.ac.uk/6269/13/OA_Policy_Dependencies_-_Final_project_report.pdf.
- [25] DCC. Checklist for a data management plan. v.4.0. [EB/OL]. [2016-12-14]. <http://www.dcc.ac.uk/resources/data-management-plans>.
- [26] Gower A C, Spira A, Lenburg M E. Discovering biological connections between experimental conditions based on common patterns of differential gene expression [J]. BMC Bioinformatics, 2011, 12: 381.
- [27] Kerrien S, Aranda B, Breuza L, et al. The IntAct molecular interaction database in 2012 [J]. Nucleic ACIDS Research, 2012, 40(D1): D841-D846.
- [28] O'Boyle N M, Guha R, Willighagen E L, et al. Open data, open source and open standards in chemistry: the blue obelisk five years on [J]. Journal of Cheminformatics, 2011, 3: 37.
- [29] Day N, Downing J, Adams S, et al. CrystalEye: automated aggregation, semantification and dissemination of the world's open crystallographic data [J]. Journal of Applied Crystallography, 2012, 45: 316-323.
- [30] Fonseca C G, Backhaus M, Bluemke D A, et al. The Cardiac Atlas Project-an imaging database for computational modeling and statistical atlases of the heart [J]. BIOinformatics, 2011, 27(16): 2288-2295.
- [31] 中国科学院文献情报中心. 文献情报中心组织的学术会议 [EB/OL]. [2018-11-04]. <http://ir.las.ac.cn/handle/12502/5457>. (National Science Library, Chinese Academy of Sciences. Academic references sponsored by National Science Library, Chinese Academy of Sciences [EB/OL]. [2018-11-04]. <http://ir.las.ac.cn/handle/12502/5457>.)
- [32] Gaulton A, Bellis L J, Bento A P, et al. ChEMBL: a large-scale bioactivity database for drug discovery [J]. Nucleic ACIDS Research, 2012, 40(D1): D1100-D1107.
- [33] Accenture technology. The ethics of data sharing: a guide to best practices and governance [EB/OL]. [2018-11-4]. https://www.accenture.com/t20161110T001618Z_w/us-en/_acnmedia/PDF-35/Accenture-The-Ethics-of-Data-Sharing.pdf.
- [34] University of Illinois at Urbana-Champaign Library. Illinois data bank-preservation review procedure [EB/OL]. [2018-11-04]. https://databank.illinois.edu/policies#preservation_review.

- [35] University of Michigan Library. Deep blue data-collections & content [EB/OL]. [2018-11-04]. https://deep-blue.lib.umich.edu/data/agreement#collections_content.
- [36] University of Notre Dame, Hesburgh Libraries. Retention and review of curate ND policy [EB/OL]. [2018-11-04]. <https://curate.nd.edu/policies/retention-review>.
- [37] Institute of Medicine (US). The benefits of data sharing [EB/OL]. [2018-11-04]. <https://www.ncbi.nlm.nih.gov/books/NBK137823>.
- [38] Lancaster University Library. Category: data conversations [EB/OL]. [2018-11-04]. <http://wp.lancs.ac.uk/highly-relevant/category/data-conversations>.
- [39] The University of Luxembourg Library. Data conversations [EB/OL]. [2018-11-04]. https://wwwfr.uni.lu/library/open_science/data_conversations.
- [40] Jonathan England. Data conversations—'GDPR and your research data' [EB/OL]. [2018-11-04]. <https://blogs.openaire.eu/?p=2976>.
- [41] 顾立平. 科学数据开放获取的政策研究[M]. 北京: 科学技术文献出版社, 2016. (KU Liping. Policy research on scientific data open access[M]. Beijing: Scientific and Technical Documentation press 2016.)
- [42] Lord P, Macdonald A, Lyon L, et al. From data dialogue to data curation [EB/OL]. [2018-11-04]. <http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf>.
- [43] Virginia Tech University Libraries. Dataset deposit process [EB/OL]. [2018-11-04]. <http://guides.lib.vt.edu/c.php?g=465788&p=3688869>.
- [44] University of Illinois at Urbana Champaign Library. Illinois data bank curation workflow [EB/OL]. [2018-11-04]. <https://drive.google.com/file/d/0B5Dm3XFQloc4bWF4c0JRTUxLZfK/view?usp=sharing>.
- [45] University of Minnesota Libraries. Curation workflow(DRUM) [EB/OL]. [2018-11-04]. <https://drive.google.com/file/d/0B5Dm3XFQloc4d0Q0Nk5ISW92TGs/view?usp=sharing>.
- [46] Washington University in St. Louis Libraries. Digital research materials repository curation workflow [EB/OL]. [2018-11-04]. <https://drive.google.com/file/d/0B5Dm3XFQloc4UTRtZHZnQ09QNnc/view?usp=sharing>.
- [47] University of California, Irvine Libraries. Digital scholarship—What we do [EB/OL]. [2018-11-04]. <http://www.lib.uci.edu/dss/what-we-do>.
- [48] University of California, Santa Barbara Library. Data curation [EB/OL]. [2018-11-04]. <http://www.library.ucsb.edu/data-curation>.
- [49] Johns Hopkins University Libraries. Data management services [EB/OL]. [2018-11-04]. <http://dms.data.jhu.edu>.
- [50] University of Michigan Library. Research data services [EB/OL]. [2018-11-04]. <https://www.lib.umich.edu/research-data-services>.
- [51] Rutgers University Libraries. About RU research [EB/OL]. [2018-11-04]. <https://rucore.libraries.rutgers.edu/research/about>.
- [52] Solis B S. Factors for enabling sharing and reuse of research data [EB/OL]. [2018-11-04]. <https://fedora.phaidra.univie.ac.at/fedora/objects/o:343651/methods/bdef:Content/get>.
- [53] U K DATA Archive. Create and manage data—research data lifecycle [EB/OL]. [2018-11-04]. <http://www.data-archive.ac.uk/create-manage/life-cycle>.
- [54] ADMIRAL. Data management plan [EB/OL]. [2018-11-04]. <http://imageweb.zoo.ox.ac.uk/wiki/index.php/>

- ADMIRAL_Data_Management_Plan_Template.
- [55] National Institute for Health Research. Research cycle [EB/OL]. [2018-11-04]. <http://www.rds-sc.nihr.ac.uk/public-patient-involvement/the-research-process>.
- [56] ICPSR. Guide to social science data preparation and archiving [EB/OL]. [2018-11-04]. <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>.
- [57] Simcoe T S ,Graham S J H ,Feldman M P. Competing on standards? Entrepreneurship , intellectual property , and platform technologies [J]. Journal of Economics and Management Strategy , 2009 , 18(3) : 775-816.
- [58] Alspaugh T A , Asuncion H U , Scacchi W. Analyzing software licenses in open architecture software systems [C]//Workshop on Emerging Trends in Free/Libre/Open Source Software Research. 2009: 54-57.
- [59] Reichman O J , Jones M B , Schildhauer M P. Challenges and opportunities of open data in ecology [J]. Science , 2011 , 331(6018) : 703-705.
- [60] Goovaerts M. Oceandocs and open science directory: two facets of the information policy of UNESCO/IOC-IODE [C]//Confluence of Ideas: Evolving to Meet the Challenges of Global Change. 2010 , 97: 109.
- [61] Pampel H , Bertelmann R. Data policies im Spannungsfeld zwischen empfehlung und verpflichtung [C]//Handbuch Forschungsdatenmanagement , Bad Honnef: Bock+Herchen Verlag. 2011: 49-61.
- [62] The International Molecular Exchange Consortium. Submit your data [EB/OL]. [2018-11-04]. <http://www.im-exconsortium.org>.
- [63] 顾立平. 数据治理——图书馆事业的发展机遇 [J]. 中国图书馆学报 , 2016(9) : 29-45. (Ku Liping. Data governance: opportunity for the library [J]. Journal of Library Science in China , 2016(9) : 29-45.)
- [64] Heidorn P B. The emerging role of libraries in data curation and E-science [EB/OL]. [2018-11-04]. https://www.researchgate.net/publication/254369307_The_Emerging_Role_of_Libraries_in_Data_Curation_and_E-science.
- [65] Khan H R , DU Y F. What is a data librarian? A content analysis of job advertisements for data librarians in the United States academic libraries [EB/OL]. [2018-11-04]. <http://library.ifla.org/id/eprint/2255>.
- [66] Kim S , Choi M S. Study on data center and data librarian role for reuse of research data [C]// International Conference on Knowledge and Smart Technology. IEEE , 2016: 303-308.
- [67] Thomas C , Urban R. What do data librarians think of the MLIS? Professionals' perceptions of knowledge transfer , trends , and challenges [J]. College & Research Libraries , 2018 , 79(3) : 401-423.
- [68] 曾建勋. 大数据时代需要数据馆员 [J]. 数字图书馆论坛 , 2017(6) : 1. (Zeng Jianxun. Big data era needs data librarians [J]. Digital Library Forum , 2017(6) : 1.)
- [69] 康旭琴,陈锐,程瑾,等. 科学数据管理和数据馆员的发展及关键问题 [J]. 中华医学图书情报杂志 , 2016 , 25(5) : 1-5. (Kang Xuqin , Chen Rui , Cheng Jin , et al. Development of scientific data management and its key problem [J]. Chinese Journal of Medical Library and Information Science , 2016 , 25(5) : 1-5.)
- [70] Ku Liping. Exploring data librarian development in open science [EB/OL]. [2018-11-04]. <http://ir.las.ac.cn/handle/12502/8741>.

顾立平 中国科学院文献情报中心副研究员。北京 100190。

中国科学院大学图书情报与档案管理系岗位教授。北京 100049。

(收稿日期: 2018-08-09; 修回日期: 2018-11-04)