

数据治理——图书馆事业的发展机遇*

顾立平

摘 要 数据驱动发现成为一种新的科学进步路线,产生了新的知识服务对象;愈加重视数据的科学家和工程师、数据科学家、数据工程师、数据分析师、公民科学家等。数据获取需要依靠国家的行政命令、资助机构的政策、研究机构和大学的管理办法等。数据共享存在于不同学科,而且已具备了基础设施,诸如:数据中心、数据云、数据知识库与数据期刊、信息技术中心、数字资源中心等。数据重用需要公共部门、企业、公众、利益相关方在“尽可能地开放,尽责任地封闭”前提下,采取协调一致的行动。在这里,数据馆员扮演重要角色。数据治理生态体系需要图书馆执行数据获取、数据共享、数据重用的政策。通过不断寻找和搜集最佳实践案例,图书馆可以支持数据治理。参考文献 79。

关键词 开放科学 开放数据 数据治理 数据获取 数据共享 数据重用 数据增值服务 数据馆员

分类号 G252.8

Data Governance: Opportunity for the Library

KU Liping

ABSTRACT

The approach of technology-driven development has made a great achievement in the past few years. Nowadays, the data-driven discovery is becoming a new route of scientific progress, which makes “innovation driving development, data driving innovation” possible.

However, opportunity and challenge exist side by side. The key is to put the slogan into practice. We need to address three questions. 1) Who would do it? In other words, whom we are going to serve? 2) What are current obstacles—which means what are currently hindering their development—and what can we do to help them? 3) How can we put it into practice, namely, the distance between imagination and reality? In other words, how can we become more mature?

Data governance will create new employment opportunities and potential economic growth. New protagonists would be scientists and engineers who pay more attention to data, a new generation of data scientists, data engineers who create data values, data analysts who create added value and citizen scientists. Data librarians are probably the most important supporting players.

* 本刊“青年学术论坛”特约稿(Special contribution for the Youth Academic Forum sponsored by this journal)

本文系中国科学院国际合作局项目“支持 GRC 开放获取行动计划的世界开放获取行动评价与监测平台”(编号:Y140321001)的研究成果之一。(The article is an outcome of the project “Global Open Access Evaluation and Monitoring Platform for Supporting GRC-OA Action Plan”(No. Y140321001) supported by Chinese Academy of Sciences.)

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

Data access governance is a form of data governance which depends on national administrative orders, funding agencies' policies and regulations of research institutions and universities to make research data open access. People who can help implement policies are in need. Obviously, it is difficult for librarians to adjust themselves from information management to data management, but not out of the question.

Data sharing environment is another form of data governance which includes different disciplines and is provided with infrastructure. Data centers, data clouds, data repositories, data journals, IT centers, and digital resource centers are tools for data storage and services rather than methods of governance. Librarians who have experience in managing carriers of different disciplines and knowledge in data rights management are able to assist in handling information services under contracts, agreements, commercial licenses and open licenses, and are also able to maximize the effectiveness of existing infrastructure.

Data governance is not only data access, but also data reuse, and its goal is not only data sharing but also open data. It relies on the coordinated action among public sectors, enterprises, publics and stakeholders on the idea "as open as possible, as closed as necessary". In general, data-sets must undergo the cost and risk assessment of national security, individual privacy, data security and business interests before release. Released data-sets are findable, accessible, interoperable and reusable. As a nonprofit agency which does not represent any special interest group, the library is the reliable data filter and release channel.

Data governance is the foundation of knowledge communication, processing and creation. As raw materials of the knowledge industry, data values depend on technologies and social organization of their age. Therefore, they may not be recognized at that time. It is the library's responsibility to store and utilize such raw materials. By searching and collecting best practices, better data governance will be gradually achieved.

KEY WORDS

Open science. Open data. Data governance. Data access. Data sharing. Data reuse. Data value-added service. Data librarian.

0 引言：创新驱动发展，数据驱动创新

技术驱动发展的方式在过去几年取得了巨大成就。如今，数据驱动发现成为一种新的科学进步路线，成为创新驱动社会经济进步的一种方向。数据的指数性增长、数字技术的日益发展、科学社群的全球化，以及社会大众应对时代挑战的需求，是社会持续变革以及开放科学研究的基础^[1]。科学数据的浪潮，不只反映数据的量，也反映了科学实践的根本改变，即：谁实施、谁承担经费、谁从中受益等政策议题。人类历史上从未如此快速、容易地获取如此多的信息^[2]。科学数据共享能力的提高，使得人类

历史上第一次可以实现运用数据和相应的数据分析思想来解决人们共同关心的问题，反映了数字技术应用的核心价值。互联网及其所有相关技术和通信标准，已经从一个科学网络演化成为一股强大的经济力量，所以，科学数据的大规模存储、共享和再利用将形成一个新的巨大财富来源，反映出科学群体的全球化和普及化。

日益成熟的数据基础设施和治理方式，不仅逐渐成为一种改造科学事业的途径，而且使任何人在任何地方、任何时间都可以获取并且重复使用数据，产生新的产品和服务、新的公司和就业机会，从而形成新的贸易流，提升国家竞争力^[3]。当前世界主要经济体的发展战略，均以数据治理作为方向，如美国《大数据研发战

略》^[4]、欧盟《朝向开放科学体系》^[5]、中国《国务院关于促进大数据发展的行动纲要》^[6]等,都强调利用大数据技术、云基础设施、科学信息开放获取、科学数据内容重用等方式,促进知识信息共享,激发社会创新活力。

数据在创造就业机会和刺激经济增长方面具有以下四项特征。

(1)数据的乘数效应。数据经过检索、下载和再利用,不仅直接有利于最初查询数据的人,对于该领域和社会大众,还应当体现在数据后续被传播和被利用的价值。2012年英国政府委托若干研究项目^[7]分析经济和社会数据服务的经济价值,通过对社会科学数据库和档案库的23 000名用户进行调查,发现易于找到并获取合适的数据意味着用户效率的提高,每年可创造超过100万英镑的价值;花在经济和社会数据服务上每1英镑,其经济价值可达5.4英镑。

(2)数据的经济价值。巴特尔研究所2011年的一项研究^[8]表明,1988年至2003年之间,美国政府在“确定人类基因代码序列”的人类基因组计划项目上投资了38亿美元,而由此在新型药品、设备、服务、就业等方面共取得7 960亿美元经济产出。该项目包括一项将序列数据归入公共领域的国际协议,对其他国家的经济发展具有潜在价值。

(3)大科学的商业价值。互联网起源于CERN的粒子加速器,起源于共享数据和信息的实际行动。欧盟成员国每年花费100亿欧元来运行资本密集型的大科学的研究设施,如同步加速器和天文望远镜等。伦敦帝国理工学院Erkko Autio教授的研究指出^[9],欧盟国家通过不同网络的知识投入和知识转移,为其产业供应商提供创新机遇。因为许多大科学项目要求独特的前沿基础设施,从而给提供这些设施的高科技企业带来重大机遇,也能为全球研究网络提供良好的交流平台。

(4)大数据的决策价值。无论是公共事业还是私营企业,当前都在关注如何能更好地利用“大数据”进行服务。德国联邦经济和能源部

2014年3月发布一份报告,坦言“这些颠覆性的变化可能会为德国贸易带来大量的机会和竞争优势”^[10],认为大数据将促进市场和公司发生重大变化,对国家竞争力产生重要影响。芬兰经济研究所在2014年8月发表的一项研究报告中认为,对谷歌的搜索查询数据进行分析,可以提高就业数据的准确性^[11]。在经济学领域,如何获得更为精确的就业数据至关重要,不仅政府依靠这类数据进行政策规划和制定,而且企业也需要这类数据了解市场发展和动向,但这类数据往往因调查时间与发布时间、调查方法与统计方式等诸多原因而不能准确地反映现实情况,从而影响政策决策。然而,通过估算当下人们关于重要热门词汇以及就业岗位的查询频率,可以适当弥补原有数据的不足。

图书馆作为社会知识资源的聚集和传播交流中心,既有发展新型知识加值服务的优势,也有肩负知识传播交流的责任。当前,大量的科研数据散布在不同的科研单位、数据中心、国际领域组织、研究中心等地方,如何挖掘数据的潜在价值,更为充分地利用它的特性达到促进科技创新的目的,是图书馆在收集、保存、提供知识获取(Knowledge Access)之外的另一项功能与职责。为适应当前以及未来的社会经济发展需求,图书馆需适当培养新型知识服务能力。

1 新的知识服务对象

数字技术推动了新一代的科学交流方式,且在科学研究的全球化发展和科学合作增长的带动下,产生了各种各样解决问题的新方法,改变了对科学研究的质量评估和影响力评价。2014年7月,欧盟委员会就Science 2.0发布公共咨询报告,声明:“科学2.0定义了科研系统功能目前正在发生的系统性变化。它的特点是开放、协同、网络化的研究方式……尽管科学2.0的原料是海量数据,但它仍然需要许多人力投入……科学2.0影响着从科学研究的初期规划到科学成果出版的整个科学研究周期,并且将

以这种方式来管理整个研究周期。”^[12]在数据驱动科学的时代,数字技术的最大优势通过科学社群的科技信息开放获取和科学数据内容重用而得到发展。从兼顾数字技术和数字内容的两方面开放性而言,数据治理对社会有四项益处。

(1) 科学家:随着越来越多的数据和工具的触手可及,科研人员跨越学科、机构或国家之间的合作交流将会更不受阻碍,有利于开拓新的研究途径。通过与公民一起探索发现,将若干实验室可以委托的工作,经过明确任务需求和执行步骤,释放到社会上,创造新的就业机会;通过与企业家一同解决具体问题,能够减少科技成果转移转化的时间和成本,使得该项工作变得更容易更快。总体效果是提高科研人员的社会地位和社会价值。

(2) 企业家:良好的数据治理为新产品和新服务提供原材料以及可供预实验的机能。预实验能够缩短产品开发周期,如气象数据、人类基因组的数据等,已有大量“互联网+”的应用。这可以在很大程度上减轻科技创新所需要担负的技术研发和市场开拓成本。通过与科研人员合作,企业家可直接取得技术方案或者间接取得发展线路,充分利用现有研究资源;通过为公民提供各种优惠措施,企业家能够直接贴近用户需求或者间接得到廉价劳动力。数据治理能在资金投入以及专业知识两方面发现找寻和提取数据的价值。

(3) 公民:提高企业和政府的效益和效率,可以使得公民直接或者间接得到新产品和新服务。公民也有权利和能力得到生活各个领域所需要做决策的信息;公民参与,是一种更为清晰的政策理解和实施方式,能够更好地结合公民需求和政府治理,达到从数据治理迈向社会治理的效果。

(4) 公务员:公共部门的数据治理,能直接激励企业以创新的方式使用这类信息。通过数据格式标准化,满足高质量、准确性和可访问性的需求,促进商业创新和社会经济发展。数据

的开放共享不仅对科学界产生影响,也对整个社会经济产生影响,开发其巨大的经济价值存在着明确的可能性^[13]。关乎社会经济发展的七大行业:教育、交通运输、消费品、电力、石油与天然气、医疗保健以及消费金融等,如果能够进行良好的数据治理,将会消除各行业的信息差距,形成规范,提高生产力,帮助企业建立细分市场,开发新的产品和服务,提高运营效率和效益。

数据治理能创造新的就业机会,带来可能的经济增长。在数据获取、数据共享、数据重用和数据加值的数据治理实践中,原有的工作岗位更加重视数据的获取、保存、利用和传播,新的岗位更加凸显创造数据价值的功能。未来学术交流的新角色有以下几类^[14]。

(1) 数据工程师(Data Engineer):具有机器学习的知识和技能,能够编写代码进行数据处理。大型公司拥有一批数据工程师来进行最低程度的数据清洗工作,或者委托第三方处理。

(2) 数据分析师(Data Analyst):了解程序运行方式以及分析工具,善于处理经过整理后的数据表格,能够对数据进行建模,接触和检查数据时具有敏感度,能够从数据描述的结果中提出有待检验的假设问题。

(3) 数据管家(Data Steward):信息专家、数据存储归档者、图书馆员或者其他类似岗位的职员,管理和发现有价值的数据,并确保可用性。

从知识和技能的角度,数据专家(Data Specialist)有时泛指上述三者;从社会结构的组成群体角度,数据专家也可用于尊称具有一定的数据采集、挖掘和利用、共享和开放经验的科学家、企业家、公务员和公众。

图书馆和信息中心是支撑社会知识获取、知识利用、知识创新、知识管理的机构。为了服务各种数据专家,数据馆员的内涵和职责也发生了变化。简言之,数据馆员(Data Librarian)是为科研数据管理(Research Data Management, RDM)而服务。不同方面的科研数据管理活动

往往分布在不同的支持服务和学术部门(如科研办公室、IT 服务、图书馆),然而,科研人员在整个科研生命周期内需要得到许多领域的支持,包括在规划、组织、安全、记录和共享、数据集存储和长期保存方面的准备工作,以及更为普遍的关于著作权、许可和知识产权等问题的咨询和解答方面^[15]。数据馆员是图书馆员职业生涯的一个发展方向,而不单指具体的岗位名称,它可以是:数据管理专员(Data Management Specialist)、数据支持人员(Data Support Officer)、科研数据支持人员(Research Data Support Officer)、数据管理规划顾问(Data Management Planning Consultant)、科研数据管理项目专员(Research Data Management Project Officer)、科研数据管理服务开发人员(Research Data Management Service Developer)、数字研究图书馆员(Digital Research Librarian)、科研数据管理协调员(Research Data Management Coordinator)等。中国科学院文献情报中心在2016年3月举办的中国数据馆员培训班,正式启动了数据馆员在我国发展的第一步。培训内容包括数据的开放许可协议、数据的知识产权、数据权益管理、数据管理计划、数据资源利用、数据分析 R Language 应用、大数据 Hadoop 架构与部署、机构知识库的数据管理、数据引用、数据发表等多个新内容^[16]。但能力培养仅仅是开始,重要的是服务。

图书馆和信息中心的知识服务对象,有各种不同需求,一是社会结构不同角色的需求,二是不同专业领域的需求,三是不同服务层次的需求。本文讨论的数据治理可以分为四个部分:数据获取、数据共享、数据重用、数据加值,在不同的服务情景和治理框架中,图书馆和信息中心的侧重点和角色也有所不同。

2 数据获取治理

数据治理的其中一种形式是数据获取(Data Access),其目的在于促进数据的可获取性和可

利用性。将科研数据开放获取设为默认状态的意义在于:从立场上,数据获取不再是一种特殊要求,而是普遍遵循的规范,数据获取成为科学传播的重要组成部分,它往往依靠国家的行政命令、资助机构的政策、研究机构和管理办法等,形成一套健全的科研数据开放获取政策体系;在实践上,协助执行政策的支撑机构和成员,往往是图书馆和信息中心协同信息基础设施的管理单位和研究单位的数据中心等,在信息管理走向数据管理的协作过程中予以落实。

科研资助机构以其清晰的角色和责任,能够为科研数据开放获取制定明确的政策,这类政策颁布者层级较高,数量较少,影响力较大,具有原则性和纲领性的特点。其对于科研文化的影响,在促进科研群体的行为变化过程中发挥着关键作用。美国白宫科技政策办公室2013年颁布“提高联邦资助科学研究成果获取的备忘录”,备忘录针对科学数据的公开获取制定了清晰的目标,要求由联邦机构全部或部分支持的非保密研究产生的数字格式的科学数据应被存储并提供公开获取,支持查询、检索和分析^[17]。欧盟委员会2012年通过了“关于获取、保存和传播科学信息的建议”^[18],并在“地平线2020”背景下论述了科研数据开放获取的试点行动。建议呼吁各成员国制定全面一致的政策策略,确保公共资助研究的出版物和科研数据开放获取。2014至2015年间,欧盟在7个领域试点实施项目产生的科研数据最大化开放获取和再利用政策^[19]。尽管欧洲尚有若干科研资助机构还未颁布科研数据开放获取政策,且大多数政策集中在政府信息领域的的数据获取,而不是科研领域的的数据获取^[20]。然而,已有若干国家在朝着科研数据开放获取的方向努力,如爱尔兰^[21]、奥地利^[22]、德国^[23]、挪威^[24]等,纷纷制定了科研数据开放获取政策或科研数据存储政策。最为显著的是 EPSRC(英国工程与自然科学研究理事会)的政策,要求英国研究机构和大学制定强制性的开放数据政策,并为科研数

据的开放获取建立基础设施^[25]。良好的数据获取政策与实践应当充分考虑和明确表达每一类利益相关者的角色定位和职责,包括资助者、受资助者、管理科研数据的存储库和数据中心。

诚然,政策是重要的促进方式,但政策实践更为重要,需要采取综合办法,包括:数据获取和长期保存的协同、综合规划关于可扩展的基础设施和服务、促进研究和创新领域数据重用的实践进展、科研人员与信息专家的能力发展、课程教育或者培训班等。例如,英国2015年的开放数据政策路线图,采取了一系列激励行动,包括:持续创建一致协调的开放数据战略,开放更多(面向社会、环境、经济效益)的数据,提供培训和技能发展(为政府、企业、市民),刺激消费开放数据产品,以及路线图整合开放数据研究工作与发展框架等^[26]。在这类综合方法的实施过程中,重要的是鼓励人们调用互补的工具、资源、社群,以实现数据获取从原则理念到实际应用。再如,塔尔图大学图书馆(University of Tartu Library, UTL)2014年公布了图书馆电子信息协会(Electronic Information for Libraries, EIFL)资助项目成果,该项目是关于实施开放科研数据支撑服务,优化开放获取支撑服务,并讨论塔尔图大学开放获取政策,为科研人员分享其开放科研数据创建基础设施与支撑系统^[27]。这类综合方法强调科研社群之间的协作,并理解他们对于基础设施和支撑性服务的需求;另外,在机构内部建立与科研资助机构密切合作的委员会,将会帮助科研社群处理政策规范的要求,从而使数据获取逐渐成为默认状态,而非需要遵守的新规定。

图书馆和信息中心在数据获取治理上扮演着居中协调的重要角色。数据共享规范、基础设施、数据资源通常已经存在于不同学科的研究机构或者大学学院的数据中心中,支撑科研工作长期以来的管理制度。因此,与得到认可的知识库或数据中心合作,对于精简数据提交、存储、利用、开放、保存等的沟通协调和协助推广至关重要。另外,科研人员更为重视出版商

的政策。所以,鼓励出版商与符合信任度接受标准的知识库和数据中心合作,甚至通过图书馆和信息中心的协调沟通,与机构知识库进行协作等日益重要。例如,开放获取出版商Ubiquity与OpenAIRE实现兼容,通过链接到科研数据的DOI丰富了其期刊元数据^[28];2014年4月,PLoS启动PLoS数据知识库集成合作伙伴计划,将投稿过程与经过挑选的数据知识库进行集成,便于科研人员简化工作流程,完成从最初投稿到最终数据开放的行为等^[29]。此外,就PLoS而言,还要求作者对稿件中关于证明研究成果的所有数据完全开放,作者在线提交稿件时必须提供一个符合PLoS政策的数据可用性声明;一旦被录用,数据可用性声明将会与文章一起发表^[30]。这类编辑政策通过数据可用性声明来进一步促进关于文档、元数据、已发布数据的格式、许可以及引用等问题的解决方案的形成。

图书馆和信息中心具有支持实现数据管理计划(Data Management Plan, DPM)的作用。科研人员需要集中在研究事务上,行政工作和数据管理也很重要,但如果有人从旁协助规划科研生命周期中每个阶段的数据存储、保存、发布和标识工作,则无论是研究还是数据获取治理本身,都会更有成效。图书馆和信息中心就有发挥这样作用的空间。例如,欧洲研究图书馆协会(Ligue des Bibliothèques Européennes de Recherche, LIBER)宣布启动促进开放科学培训欧洲研究(FOSTER)项目,旨在支持不同的利益相关者(特别是年轻的科研人员)参与开放获取、开放数据共享和开放科学活动^[31],通过面向科研人员和技术人员的培养方案来支持数据管理。另外,联合国教科文组织(UNESCO)创建面向科研人员 and 图书馆员的开放获取课程,以帮助促进图书馆、信息专业人员和科研人员的能力建设^[32]。针对专业问题的特定需要,图书馆员需要提供知识产权、许可协议、数据治理以及学术伦理与道德等的咨询服务或者辅助工作。

图书馆和信息中心具有妥善处理数据获取边界的作用。与其他社会组织或者利益团体不

同,在数据保护和隐私泛政治化的情况下,图书馆和信息中心具有公益服务的性质,更具备进行数据获取治理的中立立场。数据获取治理需要处理的权益问题是:如何保护数据共享所带来的好处,既不破坏科研人员和资助者的权利,又不损害公民的隐私权。这些问题涉及法律条款与合同、科研伦理与道德、数据获取与保护等,是图书馆和信息中心的咨询服务内容之一。若干数据获取治理的方式,已有成例。

(1)在线隐私。欧洲法院曾就遵循在线“被遗忘权”强制改变 Google 和其他搜索引擎的运行方式^[33]。许多社会科学研究涉及如何在线处理个人数据的问题,这类数据获取治理的判断原则和边界不断进行调整,科研人员需要提供适当的咨询服务。

(2)病历记录。在科研界,病历记录带来了最大的政策挑战。没有人愿意在网络空间或者员工档案上看到自己的就诊记录,但每个人都希望科研人员找到新药物和更好的治疗方案。欧盟委员会提出的数据保护框架,其用意就在于完成研究的同时进行隐私保护^[34]。

(3)知识产权。公众理应免费使用公共资金资助的科研成果,但同时,社会上又需要鼓励产业投资以及商业化研究成果。激励措施有两方面:一是鼓励开放和透明化,因为这将促使更加进步和得到认可;二是主张保护和产权化,以追求合理的市场独占或者寡占利润空间。这既是数据共享的两难,也是科技创新的双轨。

科研机构应系统地支持其科研人员解决由科研数据开放获取产生的法律和道德问题。例如,哈佛大学定量社会科学研究所(IQSS)和学术交流办公室推出开放数据协助计划(ODAP),为如何将数据文件存入哈佛大学的数据知识库提供咨询和指导,包括数据在限制允许的条件下开放获取的咨询建议^[35]等。通过图书馆和信息中心的实践,能够支持科研教育机构在数据获取上的政策实现,包括:丰富学术道德委员会的职能,提供正式的培训,开发适用的许可协议,审理和总结关于隐私和机密性等问题的解

决方案和良好实践等。

3 数据共享治理

数据共享不完全等于数据获取,前者是在已有科研基础设施基础上,如:数据中心、数据云、数据知识库与数据期刊、信息技术中心、数字资源中心等,提供数据存储、保存以及有条件地提供服务。它要求数据质量,要求使用者也是贡献者,或者是内容贡献者或者是资金资助者。数据共享已在不同学科领域内形成各自的数据规范和共享氛围。随着应用需求的变化和信息技术的进步,数据共享已经成为跨学科、跨机构、跨越研究机构和社会界限的重要措施。以下是关于数据共享的实例。

(1)数据空间:哥白尼计划。欧洲航天局发射的哥白尼 5 系列^[36]地球观测卫星中,第一颗卫星自 2014 年 4 月起将采集到的图像发送回地球,每个地球轨道将产生几百万兆字节的数据流,提供连续性和高分辨率的时空数据,所有这些信息仅供参考提供给公共部门、科学和商业用户及公众。研究表明,预计 2030 年,该系统能够产生大约 300 亿欧元的经济效益和 50 000 个新的就业机会^[37]。高端仪器设备可以由欧盟境内的企业开发和制作,并且逐步推广到世界其他地区,而数据可以用于公民科学。例如,Tiger Nation^[38]项目追踪地球仅剩的 1 700 只存活在印度野外的孟加拉虎的记录,每年禁猎区的部分游客拍摄照片后,可上传至中央数据库,利用图像识别软件通过老虎条纹来识别每一只老虎,从而利用卫星位置数据进行跟踪。

(2)生物和生态体系 LifeWatch 项目^[39]提供生物多样性和生态系统研究的数字化科研基础设施。该项目关注入侵物种蔓延,与本地物种竞争,使得本地物种面临灭绝威胁的现象。例如,监测欧洲地区的湿地,以及通过单点记录器所收集的候鸟信息。这些所得数据可以与其他来源的数据交叉引用,如有关天气、气候和地理的数据集等。

(3)内海长期生态研究^[40]。为获得荷兰重要内海的综合生态系统视图,瓦尔特集团倡议除了网络监测之外,还需建设数据池以填补空白。

(4)生命科学研究 Elixir 项目^[41]。对公共资助的生命科学研究所产生的海量数据进行管理和维护的多个机构的合作项目。

(5)欧洲医学信息框架(European Medical Information Framework, EMIF)^[42]。该项目为数据共享提供了一个通用架构,有来自7个国家4800万病人的病历记录。众多项目之中,包括EMIF-AD阿尔茨海默氏病的项目,利用文本和数据挖掘病例记录,以查找在持续性神经系统退化的情况下,基因、生物标志物和结果之间的联系。

(6)临床试验数据 Pharmacog 项目^[43]。旨在通过改善临床研究和模型,协助在药物发现的早期阶段找出可能治疗阿尔茨海默氏症的化合物。该项目涉及15家学术机构、12家全球性制药企业和5家中小企业,是产学合作的平台,目的是确保只有有效的生物化合物才能进入临床开发阶段,减少药物开发过程中的昂贵药物在后期失败的现象。

(7)脑损伤共享数据库。创伤性脑损伤是导致45岁以下的人患残疾的主要原因,在美国照顾脑损伤患者的成本一年可达600亿美元,在欧洲成本至少为330亿欧元^[44]。因此,该项目旨在共享60间医院和38所科学基础设施的信息,收集2万至3万例脑外伤患者的数据,并且提供一系列可能的治疗方法,目的是帮助医生决定在何种情况下使用何种方法。

(8)人类大脑项目^[45]。工作内容覆盖众多学科:高性能计算、医疗信息、模拟人类大脑等,是一个跨国家多合作伙伴的平台,涉及26个国家的135个研究机构或者团体。当前,关于人类大脑研究的数据尚未具有标准化地、系统地收集与存储的方式和规范,因此需要这类项目不断进行数据共享的探索和整合。

(9)文化遗产 CLARIN 项目^[46]。它将欧洲

的大学、图书馆和其他公共机构的档案信息进行整合,关注文字、声音、图像等文化工艺品,通过注释和标记,使得数据易于检索,而且数据和科研成果通过永久标识符以长期可持续的方式保存。人文和社会科学学者可以登录访问所有数字数据集,或将数据提供给公共机构;科研人员能够利用不同国家的不同资源建造自己的虚拟馆藏,利用语言工具对数据进行注释、探索、开发、增强、分析、处理和可视化。

图书馆和信息中心的作用在于管理各学科不同载体的知识内容,特别是长期累积的数据权益管理经验,能够协助处理合同、协议、商业许可和开放许可下的信息服务,从而最大程度地发挥现有基础设施的效用。

图书馆和信息中心可以通过积极找寻和宣传数据共享的最佳实践,参与和促进数据共享文化。目前已有许多权益管理的实践案例,通过数据共享建立跨区域合作,或者实现领域内的共享规则与数据管理。例如,英国开放数据协会(ODI)和中国台湾开放数据联盟(ODA)在2013年签署开放数据协同协议,旨在促进和探索适用于公众、个人和学术领域的潜在开放数据^[47]。古生物学数据库(Paleobiology Database, PBDB)是最大的化石数据库之一,只要使用前确定遵守知识共享(Creation Commons)协议,其数据库内所有数据均可公开使用^[48]。数字化管理中心(Digital Curation Centre)制定详细指南来帮助机构开发数据管理服务^[49]。这类良好的数据共享实践可以引领科学研究方式和交流方式的转变。

科研资助机构、科研人员、科研机构政策制定者、数据管理者、出版商应当团结合作,制定协调一致的政策和可持续发展的战略,建立相关的基础设施。例如:世界银行联合开放数据机构(ODI)和开放知识基金(Open Knowledge Foundation)帮助发展中国家的决策者和民众理解和利用开放数据的项目^[50]。在美国农业部召开的国际农业开放数据会议中,有400位食品安全专家、数据科学家、科技专家、政策制定者等,

共同倡议和推动数据共享和科技信息开放获取^[51]。美国国立卫生研究院(NIH)和美国癌症研究所(NCI)等科研资助机构不仅吸引许多研究项目,也要求受资助者遵守有关政策^[52]。图书馆和信息中心尽管不是各个学科发展和制定科技政策的主要推动者,但是在同有关的利益相关者和网络平台培养合作关系上,是提供交流渠道的最佳平台。

在形成基础设施和服务规模经济方面,合作共享科学数据服务非常重要。为促进科研数据为社会增益,科学技术数据委员会(The Committee on Data for Science and Technology, CODATA)、世界数据系统(World Data System)和科学数据联盟(Research Data Alliance, RDA)于2015年3月达成合作协议,其中CODATA关注宣传、政策和流程,WDS关注数据供给和管理,RDA关注自下而上的数据共享工具和实践实现^[53]。这类数据共享的支持服务要求科研机构评估自己当前的能力,并在提供服务和支持可持续、可扩展的学术信息交流生态系统方面与其他专业机构开展合作。

数据共享的生态系统和基础设施,不仅仅是新型信息技术的应用,更重要的是法律法规的制度建设。它需要建立在机构内部和机构之间积极追求合作的成功案例的基础上,才能确保有效实现科研数据的开放获取和长期保存。例如,开放数据研究所(ODI)创建的“数据结点”项目提供培训、研究和发展规划,成为世界各地公司、大学和非政府组织支持开放数据项目和团队的基础数据枢纽^[54]。这类支持服务鼓励专业图书馆和数据中心与其他的利益相关者(科研机构、出版商和资助机构)和科研群体合作,从开发合作基础设施和服务的角度来评估它们在开放获取生态系统中的整体能力和地位,以及评定将要提供的服务类型。

4 数据重用治理

数据重用不完全等于数据共享或者数据获

取。数据重用的目的是最大程度地利用和挖掘数据的价值。这需要公共部门、企业、公众及其他利益相关者在“尽可能地开放,尽责地地封闭”前提下,采取协调一致的行动。数据重用治理的目标清晰而且过程严谨:释出之前的数据集必须经过国家安全、个体隐私、数据安全和商业利益的成本与风险检查,释出的数据集具有可发现性、可获取性、可交互性、可重用性。图书馆是具有公益性质的第三方机构,不代表特殊利益团体,是可靠的数据过滤闸口和数据释出渠道。

数据重用改变了“做”科学的方式,并且改变了从事科学的主体。通过在线工具可以使科研人员引导成千上万的群众主动报告野生动物的踪迹,观察星系,帮助跟踪传染性疾病等。科研群体更加具有开放文化的特质:共享初步的科研成果,进行大规模的国际合作,使其后台数据可访问,引领公民社会注重知识和参与科学研究等。其结果是直接产生社会贡献,而不仅仅是成果转化产品、产品投入市场、改变民生的间接性社会化过程。例如,2014年初英格兰南部洪水泛滥,英国政府启动了一个为期一天的“洪水黑客”活动^[55],由200多名程序开发者自愿为被洪水袭击的社区开发应用程序。其中,“不要惊慌”允许当时的人们不论是否通过Web访问,都可以得到从传递沙袋到传递洪灾信息的咨询帮助,事后所有数据被记录下来,供日后进行分析和制定实时响应计划。而“洪水反馈线”则是可视化洪水和相关数据反馈的聚合工具。充分利用实时更新的环境数据能有效帮助管理,控制损失。

数据重用促进公民科学的发展。欧洲部分国家在2014年5月进行的“荷马史诗”项目,组织群众开发地中海国家公共部门的信息价值^[56]。其中,应用程序“Geostep”能为智能手机提供相应的指南,它根据用户的位置来推荐参观景点,然后在到达该景点后提供进一步的信息。基于低成本传感技术的“Everyaware”项目^[57]鼓励公民收集环境数据并融入他们个人主

观感受的事实,来进行数据资源建设;通过智能手机的麦克风记录噪音污染,通过电子设备的温度计(传感器)来记录温度,通过社交网络工具收集和传播来自数千部智能手机的其他数据等。该项目进行环境调查和结果分析的方式,是通过各种移动用户的数据上传并整合在一个单一的技术平台上,对其进行分析、解释和可视化而实现的。在数据重用的基础上,公民科学的项目得以蔓延开来。

数据重用对于科学 2.0 至关重要。如今,科学研究不仅仅是有关离散假设的检验,而是与庞大数据集的持续互动,产生了一系列日新月异的新假设,并用不同级别的信度来评估每一个假设。根据欧盟委员会的调查,科学 2.0 的合作方式正在快速增加^[58]。例如,在哈佛史密森天体物理中心的科学家 2014 年发表宇宙膨胀(俗称宇宙大爆炸理论)的直接证据六周后,出现了 200 余篇论文,这是其他科学家借鉴哈佛的数据所产生的新见解。同理,在 CERN 的粒子加速器中发现希格斯玻色子的著名实验,则是由数千名科研人员和工程师一起共享和分析数据来实现的。这些事例表明,可随时获取庞大的数据集会改变科学工作的方法,具有深刻的认识论意义。

图书馆和信息中心可能不是主导各种数据重用带来公民科学和科学 2.0 浪潮的主角,但是在提升数据管理计划的工作和任务,则是辅助数据专家的重要方式。国际农业研究磋商组织联盟(Consortium of International Agricultural Research Centers, CGIAR)通过的“开放获取和开放数据政策”提倡几类有代表性的信息产品实施开放获取,包括:数据和数据库、数据收集和分析工具、元数据等,并指出应制定开放获取和数据管理计划,以保证这一政策的实施^[59]。图书馆和信息中心近年来的主要发展方向之一,就是通过 DMP(数据管理计划)嵌入和促进数据管理过程,鼓励科研资助机构和科研教育机构认可 DMP 是一项伴随项目的独特活动,应有适当资源进行分配。

图书馆和信息中心具有协作建立数据质量机制来确保数据重用以及长期保存的作用。为了确保数据质量,数据管理者往往设计一系列的质量保证和控制策略,包括:提供过程文档,完整性和一致性检验,数据管理和共享培训,文件格式验证,元数据检查,存储完整性验证和标注质量信息的工具等。例如,欧盟支持的由 OpenAIRE 和 CERN 联合开发的一站式研究成果网站 Zenodo,能将欧洲的研究成果数据集和资助信息链接在一起^[60]。但是政策实施需要与科研机构、学术社群和科研人员建立引用标准,以及与资助机构、资助者、出版商和期刊编辑合作,确保政策执行的效果。图书馆和信息中心可以充分发挥支持作用。

图书馆和信息中心具有数据监管(Data Curation)的作用,更重要的是提供社会服务的职责。例如,德国数字图书馆(DDB)发布的应用程序接口(API)可将所有馆藏数据实施开放获取和数据重用,用户通过 API 可以更容易地重用自由共享的 DDB 文化遗产数据^[61]。新西兰国家图书馆(NLNZ)和 Alexander Turnbull 图书馆联合发布支持馆藏内容及其元数据使用和再利用的新政策,该政策可以保障图书馆政策制定、活动和信息发布方面的一致性和透明性^[62]。

图书馆和信息中心可发展对数据重用和开放数据的观测。例如,在开放数据的观测活动上,可以通过开放数据监测(Open Data Monitor, ODM)平台得到有关信息:目录和数据集的数量、数据随时间推移的发展情况、机器可处理数据比例、数据授权信息、元数据、发布数据的组织等。加州数字图书馆(CDL)和 DataONE 与 PLoS 合作的项目“数据级别计量”旨在产生一系列跟踪和测量数据利用的指标^[63]。这类指标能够观测数据重用的学术价值和社会价值,从而评估和考虑数据馆藏的存储对象和方式。这些信息有利于图书馆和信息中心的数据采集、评估和管理。

数据重用的价值应当体现在数据可被视作一流的学术成果。为科研数据建立同行评审过

程是进一步加强高质量数据产出的一种方法。数据同行评议及数据质量保证的一般作法^[64]包括:对数据可获取性的需求(通过数据中心或知识库),为被审核数据提供充分的信息,为数据同行评议者就如何进行数据审核和应该检查哪些特征制定清晰的指南等。科学价值通常通过出版前和出版后的同行评议进行评估,所采用的技术和质量标准包括:数据集的完整性和一致性、数据描述规范、数据产生的方法和工具(软件)等。

图书馆和信息中心需要明确告知数据引用的处理方式,以解答读者咨询数据重用所需要的学术伦理和道德问题。这类信息包括:数字资源唯一标识符(DOI)、开放许可协议(例如CC)、可操作软硬件和仪器,以及告诉用户能够用科研数据所做的案例介绍。开放许可协议CC修订并发布的新版本CC-BY 4.0和CC-BY-SA 4.0协议极大地简化了开放数据和内容发布的授权,允许出版商为其所有开放作品从一系列协议中进行挑选^[65]。数据知识库High Wire平台则是采用出版集团和图书馆界的COUNTER规范,该规范为用户提供如何使用订阅内容的信息,包括某个时间段使用的新数据、机构存档获取、期刊的数字资源唯一标识符(DOI)和专有标识符^[66]等。在CrossRef和DataCite宣布合作后,两家机构致力于加速科研人员采用DOI进行数据出版和引用^[67]。出版商应当保证其出版物的附加数据可引用、可获取、可重用,而且对如何引用提供明确的指导。数据重用需要政策指引,特别是数据引用以及数据计量,这项任务需要图书馆和信息中心与出版团体进行协作。

图书馆和信息中心可以通过宣传推介活动,提高人们对数据重用的认识。参考已经被翻译成18种语言的开放数据手册(新版在2015年5月发布),图书馆员可以了解如何指导开放数据新人和富有经验的老手。该手册从不同视角阐述了开放数据,即哪些数据应该开放,开放数据的社会和经济效益是什么,以及一旦开放

如何有效使用它^[68]。在研讨会或者培训班上,通过介绍利用开放数据进行创新性使用的可能性,可以使人们了解如何探索数据驱动解决社会、经济和环境问题^[69]。图书馆和信息中心的宣传活动可以有不同的形式,如研讨会、网络会议、宣传册、传单等,而且应该与培训项目相结合。

5 数据加值治理

数据是知识工业的原料,其价值取决于当时的技术手段和社会组织,因此,数据治理是知识传播、知识加工、知识创作的重要基础。数据加值涉及从原始数据或者源生数字资源,经过一定程度的数据处理,形成衍生数据或者数据产品的过程。数据加值是图书馆和信息中心在存储和利用具有质量保障的数据,通过不断寻找和搜集最佳实践案例,逐步实现良好的数据治理的过程中新增的一项社会服务功能。

数据加值与数据获取、数据共享、数据重用的最大不同在于,更加重视数据资产管理。因此,获得第三方认证以及认可,来保证长期高质量的数据加值服务,至关重要。例如数据认证印章以及其他适当的ISO认证。第三方机构诸如DANS的数据认证印章(Data Seal of Approval, DSA)在国际上得到广泛应用,在完成DSA所提供的相关指导和审查之后,将授予可信任和可持续的数据知识库进行实名认证,可为用户提供可靠性的识别^[70]。建立质量保证机制不仅对数据的可信度非常重要,对于数据加值也同样重要。

数据加值需要可持续的商业模式作为长期提供服务的保障。例如,受FP7资助的4C项目,帮助欧盟机构解决数字化管理的成本问题,使得机构在数字化管理和保存方面的投资更加高效^[71]。收入来源的计划在数据加值服务开始之初就尽可能被有效地解决,同时,也需要定期对策略进行审核。如获得收入,可能需要将收入来源多样化,并将所提供的服务进行分级,从

而对用户进行收费。

图书馆和信息中心应当倡议文本和数据挖掘的合理使用以及开放许可政策。国际出版集团 Elsevier 意识到科研人员的需求后,于 2014 年 1 月正式推出新的文本和数据挖掘政策:科研人员通过 API 完成文本挖掘过程后,允许研究成果以 CC BY-NC 许可方式(用于非商业目的)发布^[72-73]。欧洲研究图书馆联盟(LIBER)则一直积极倡导著作权法改革,并于 2015 年 5 月发布了有关数字时代知识发现的《海牙宣言》^[74],以促进文本数据挖掘(TDM),该宣言旨在强调通过数据驱动创新的便利化加快科学发现的步伐,从而促使科研机构研究成果达到国际一流水平^[75]。考虑到利用 TDM 工具可以带来显著的经济效益,应鼓励出版集团调整他们的著作权政策,以鼓励在各类研究情景中使用这类技术,而出版社的编辑政策应当明确地声明数据被出版和再利用所遵循的许可(默认的和推荐的)。

图书馆和信息中心的数据增值服务,应当首先保障数据获取和数据重用的公益性质,在此基础上,若有其他数据处理要求,则进行数据增值服务。例如,德国国家图书馆(DNB)自 2014 年 1 月 1 日起,根据知识共享 Zero 协议(Creative Commons Zero,CC0),使其大约 95%的数据都处于完全开放的情况^[76]。哈佛大学定量社会科学研究所以(QSS)和学术交流办公室所推出的开放数据协助计划(ODAP),就如何将数据文件存入哈佛大学的 Dataverse(数据知识库)中提供咨询和指导,其主要咨询内容是如何使数据文件在限制条件下进行开放获取^[35]。科研机构应当系统地解决科研数据开放获取可能产生的法律和道德问题,包括许可、隐私和机密性等。图书馆和信息中心可以提供法律和学术伦理道德的咨询服务,来支持科研群体、中小企业、公众等认识数据增值产品的数据资产管理。

数据加值的技能和知识是未来信息管理专家和数据馆员的核心竞争力。欧盟 2014 年 11 月 4 日宣布投资 290 万欧元用于建立一所新的

欧洲数据科学研究院(European Data Science Academy,ESDA),旨在培育具备技能和详尽知识的新一代数据科学家^[77]。这类为科研人员和工作人员开发的教育性培训计划,能够提高数据管理技能并增强数据密集型研究的潜能。伊利诺伊大学图书情报研究生学院则开设了数据管理专业,对图书馆员、技术人员和信息专家提供课程培训,帮助其掌握相关技能,反过来也会满足科研人员的需求^[78]。在设计此类项目时,科研机构应注意学科的特殊性和实际情况,同时避免出现一刀切的现象。图书馆和信息中心应当与科研机构和数据治理者探索开展联合课程的可能性,特别是针对不同专业的实际应用和未来发展方向的探索。

6 结语:走向可以创造幸福的时代

数据治理的基础是开放获取,图书馆和信息中心可以通过以下几种方式参与到数据开放获取中^[79]:①发展知识加值型服务:提供科研人员发表开放获取期刊论文或者图书著作的便捷渠道;②进行机构知识库管理:支持学者们将科研产出进行有效存储和开放获取;③编制最佳实践指南:提供关于著作权和开放获取的咨询和实施指南;④开展学术研究:长期关注开放获取和数据治理的政策研究以及参与中国开放获取推介周等交流活动;⑤管理出版基金和资助基金:参加开放学术社群,推广开放许可协议等;⑥服务科研人员和教职员工:就学术交流的问题给予具体解答或者提供问题解决方案等。在科技信息开放获取之后,新的挑战是科学数据内容重用,其后再是数据治理。

图书馆和信息中心可以通过开设课程或者培训班来推行数据治理:面向科研人员和支持开放科学的数据管理者进行培训和教育,面向图书馆员进行最新的数据权益和产权问题的案例分析,开发对数据科学和信息管理发展有用的课程,作为能力建设的同时,铺设合情合理的职业发展道路。

图书馆和信息中心可为数据治理制定长期可持续的管理和保存计划;采取协作模式,与各个利益相关者合作,利用资源,发展适当的服务和基础设施;共同应对数据调用、发现和获取、保存、技术老化、文件和元数据规范、质量和相关指标以及安全等问题。在信息技术方案上,解决不同学科要求、数据多样化以及元数据和数据标准化的问题。

图书馆和信息中心可与利益相关者和相关网络合作,建立良好的数据治理生态体系,为不同利益团体设置合适的角色和责任,以避免重复工作和资源流失。共同制定最优的资源配置方案,协调一致的政策建议,共同的质量标准等。

图书馆和信息中心应有能力解决数据治理中的法律和道德问题。凝聚利益相关者的共识,需要立场相对中立的图书馆和信息中心作为中介。通过数据治理流程的制度化以及论坛

和培训,可系统地解决各自独立的法律和道德事件。开放许可协议,描述了哪些科研数据应该被获取、共享和重用。机械可读的开放许可协议,是基于可互操作的技术解决方案,最终以政策形式系统性地呈现更有效率的实现方案。

图书馆的服务对象是读者,图书馆事业的发展核心是图书馆员。每个时代都有图书馆事业发展的困难和挑战。生活在这个时代,是可以选择生活方式和职业理想的时代,偶尔也有困难和委屈,但是可以通过努力取得收获。数据治理是图书馆事业的发展机遇之一,在数据获取、数据共享、数据重用、数据加值上,处处皆有可能,在数据馆员的发展方向上,也有一群图书馆员正在努力。

最重要的是,永不放弃对用户的关心和支持,因为那才是图书馆员和图书馆事业的生长动力,以及面对任何时候的发展机遇(包括数据治理在内)所应具有的良好心态和发展理念。

参考文献

- [1] Council of the European Union. Open science[EB/OL].(2015-12-16)[2016-08-10]. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.
- [2] Silva D. Internet has only just begun, say founders. [EB/OL].(2009-04-22)[2016-08-10]. <http://phys.org/news159644537.html#jCp>.
- [3] RDA Europe. The data harvest: how sharing research data can yield knowledge, jobs and growth[EB/OL].(2014-12-27)[2015-09-22]. <https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf>.
- [4] NITRD. The federal big data research and development strategic plan[EB/OL].(2016-05-20)[2016-05-20]. <https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=63>.
- [5] Council of the European Union. The transition towards an open science system-council conclusions[EB/OL].(2016-03-15)[2016-05-27]. <http://data.consilium.europa.eu/doc/document/ST-7082-2016-INIT/en/pdf>.
- [6] 国务院. 关于促进大数据发展的行动纲要[EB/OL].(2015-09-15)[2015-09-22]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm. (The State Council of the People's Republic of China. Note of action guideline of big data development[EB/OL].(2015-09-15)[2015-09-22]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.)
- [7] Houghton J. Economic impact evaluation of the economic and social data service[EB/OL].(2012-03-28)[2016-04-07]. <http://www.esrc.ac.uk/files/research/evaluation-and-impact/economic-impact-evaluation-of-the-economic-and-social-data-service>.

-
- [8] Gitlin J M. Calculating the economic impact of the Human Genome Project [EB/OL]. (2013-06-12) [2016-05-15]. <https://www.genome.gov/27544383/calculating-the-economic-impact-of-the-human-genome-project>.
- [9] Department for Business, Innovation & Skills. Innovation from big science enhancing big science impact agenda [EB/OL]. (2014-05-21) [2016-08-16]. <http://dera.ioe.ac.uk/19649/1/bis-14-618-innovation-from-big-science-enhancing-big-science-impact-agenda.pdf>.
- [10] Markl V, Hoeren T, Krcmar H. Innovations potenzial analyse für die neuen technologien für das verwalten und analysieren von großen Datenmengen [EB/OL]. (2013-01-11) [2015-11-04]. http://www.dima.tu-berlin.de/fileadmin/fg131/Publikation/BDM_Studie/StudieBiDaMa-online-v2.pdf.
- [11] Tuhkuri J. Big data Google searches predict unemployment in Finland [EB/OL]. (2014-08-14) [2015-11-04]. <http://www.etla.fi/wp-content/uploads/ETLA-Raportit-Reports-31.pdf>.
- [12] European Commission. Science 2.0; science in transition [EB/OL]. (2015-08-26) [2016-01-10]. <http://ec.europa.eu/research/consultations/science-2.0/background.pdf>.
- [13] Manyika J, Chui M, Farrell D, et al. Open data; unlocking innovation and performance with liquid information [EB/OL]. (2013-10-27) [2015-11-04]. http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.
- [14] Boulton G. Open data and the future of science [EB/OL]. (2014-11-04) [2015-04-01]. <https://www.ub.uit.no/baser/septentrio/index.php/SCS/article/view/3231>.
- [15] Schmidt B, Shearer K. Librarians' competencies profile for research data management [EB/OL]. (2016-06-20) [2016-06-20]. https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf.
- [16] 武汉大学信息管理学院期刊中心. 首届中国数据馆员培训班在京顺利举办 [EB/OL]. (2016-04-01) [2016-04-07]. <http://www.dik.whu.edu.cn/CN/column/item167.shtml>. (Journal Center of School of Information Management Wuhan University. The first data librarian training workshop is success in Beijing [EB/OL]. (2016-04-01) [2016-04-07]. <http://www.dik.whu.edu.cn/CN/column/item167.shtml>)
- [17] OSTP. Memorandum for the heads of executive departments and agencies-increasing access to the results of federally funded scientific research [EB/OL]. (2013-02-22) [2013-02-26]. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
- [18] European Commission. Towards better access to scientific information; boosting the benefits of public investments in research [EB/OL]. (2012-07-23) [2016-04-16]. https://era.gv.at/object/document/485/attach/ST12847_EN12.pdf.
- [19] European Commission. H2020 programme guidelines on FAIR data management in Horizon 2020 [EB/OL]. (2016-07-26) [2016-08-08]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- [20] Sturges P, Bamkin M, Anders J, et al. Access to research data; addressing the problem through journal data sharing policies [EB/OL]. (2014-06-04) [2016-08-10]. <http://docs.lib.purdue.edu/iatul/2014/openaccess/3>.
- [21] IUA. Government launches national open access statement [EB/OL]. (2012-10-23) [2015-11-04]. <http://www.iua.ie/government-launches-national-open-access-statement-23-oct-2012>.
- [22] FWF. Pilot programme "Open Research Data" [EB/OL]. (2016-01-18) [2016-03-08]. http://www.fwf.ac.at/en/news-and-media-relations/news/detail/nid/20160118-2166/?tx_rsmnews_detail%5Bref%5D=1&cHash=5ee68842ef88402c950153bda397ace4.
- [23] German Research Foundation. Empfehlungen zur gesicherten aufbewahrung und bereitstellung digitaler forschungsprimärdaten [EB/OL]. (2009-01-17) [2016-03-08]. <http://www.dfg.de/download/pdf/foerderung/>
-

programme/lis/ua_inf_empfehlungen_200901.pdf.

- [24] Research Council of Norway. Research data must be shared[EB/OL].(2014-10-10)[2015-11-04]. http://www.forskningssradet.no/en/Newsarticle/Research_data_must_be_shared/1254000848864/p1177315753918.
- [25] EPSRC. Clarifications of EPSRC expectations on research data management[EB/OL].(2014-10-09)[2015-11-04].<http://www.epsrc.ac.uk/about/standards/researchdata/expectations/>.
- [26] Open Data Institute. Open data roadmap for the UK-2015[EB/OL].(2014-12-11)[2015-12-27]. <http://theodi.org/roadmap-uk-2015>.
- [27] EIFL. Tartu university supports OA and open data sharing[EB/OL].(2014-09-18)[2015-09-22]. <http://www.eifl.net/news/tartu-university-supports-oa-and-open-data-sharing>.
- [28] Prestigious OA publisher OpenAIRE compliant[EB/OL].(2015-07-28)[2015-10-27]. <https://www.openaire.eu/prestigious-oa-publisher-openaire-compliant>.
- [29] PLoS. Make data sharing easy: PLoS launches its data repository integration partner program[EB/OL].(2014-11-05)[2015-11-27]. <http://blogs.plos.org/tech/make-data-sharing-easy-plos-launches-its-data-repository-integration-partner-program>.
- [30] PLoS. Data access for the open access literature[EB/OL].(2013-12-12)[2015-12-27].<http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy>.
- [31] LIBER. Call for proposals to host FOSTER training events on open access, open data and open science[EB/OL].(2014-02-05)[2016-03-08].<http://www.libereurope.eu/news/call-for-proposals-to-host-foster-training-events-on-open-access-open-data-and-open-science>.
- [32] UNESCO. UNESCO launches open access curricula for researchers and librarians[EB/OL].(2014-12-19)[2015-12-27]. http://www.unesco.org/new/en/communication-and-information/resources/news-and-in-focus-articles/all-news/news/unesco_launches_open_access_curricula_for_researchers_and_librarians/#.VKqxm-QSw11.
- [33] Court of Justice of the European Union. An internet search engine operator is responsible for the processing that it carries out of personal data which appear on web pages published by third parties[EB/OL].(2014-05-13)[2016-06-22]. <http://curia.europa.eu/jcms/upload/docs/application/pdf/2014-05/cp140070en.pdf>.
- [34] League of European Research Universities. The EP's position on the general data protection regulation threatens EU research[EB/OL].(2014-10-06)[2016-04-27]. <http://www.leru.org/index.php/public/news/the-eps-position-on-the-general-data-protection-regulation-threatens-eu-research>.
- [35] Crosas M. Harvard will help researchers open up their data[EB/OL].(2014-10-17)[2016-04-27].<https://osc.hul.harvard.edu/oaweek2014#etds>.
- [36] COPERNICUS. Overview[EB/OL].(2014-11-24)[2016-03-07]. <http://www.copernicus.eu/main/copernicus-brief>.
- [37] ESA. Free access to copernicus sentinel data[EB/OL].(2013-11-15)[2016-03-08].http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Free_access_to_Copernicus_Sentinel_satellite_data.
- [38] Tiger Nation. Follow wild live[EB/OL].(2013-11-15)[2016-03-08].<http://www.tigernation.org>.
- [39] LifeWatch[EB/OL].[2016-03-08].<http://www.lifewatch.eu>.
- [40] Walter project[EB/OL].[2016-03-08].<http://www.walterproject.nl>.
- [41] Elixir[EB/OL].[2016-03-08].<http://www.elixir-europe.org/>.
- [42] European medical information framework[EB/OL].[2016-03-08].<http://http://www.emif.eu>.
- [43] Pharmacog[EB/OL].[2016-03-08].<http://www.imi.europa.eu/content/pharma-cog>.
- [44] CENTER-TBI. Project[EB/OL].[2016-01-02].<http://www.center-tbi.eu/project/background>.

-
- [45] Human brain project[EB/OL]. [2016-03-08]. <http://www.humanbrainproject.eu>.
- [46] CLARIN ERIC[EB/OL]. [2016-03-08]. <http://www.clarin.eu>.
- [47] Open Data Institute. ODI and Chinese Taiwan's Open Data Alliance sign up to open data collaboration[EB/OL]. (2013-12-11) [2015-12-27]. <http://theodi.org/news/odi-and-taiwans-open-data-alliance-sign-up-to-open-data-collaboration>.
- [48] Creative Commons. Paleobiology database now CC BY[EB/OL]. (2013-12-19) [2016-03-07]. <http://creativecommons.org/weblog/entry/41216>.
- [49] Jones S, Pryor G, Whyte A. How to develop research data management services[EB/OL]. (2011-09-08) [2016-05-22]. <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>.
- [50] World Bank. New partnership seeks to bring benefits of open data to developing countries[EB/OL]. (2013-09-18) [2016-06-22]. <http://www.worldbank.org/en/news/press-release/2013/09/18/new-partnership-seeks-bring-benefits-open-data-developing-countries>.
- [51] Townsend K. Why open data matters: G-8 and African nations increase open data for food security[EB/OL]. (2013-06-22) [2016-02-02]. <https://www.data.gov/development/why-open-data-matters-g-8-and-african-nations-increase-open-data-for-food-security>.
- [52] Hayden E C. Geneticists push for global data-sharing[J]. *Nature*, 2013, 498: 16-17.
- [53] CODATA. International organizations form partnership to increase the benefit of research data for society[EB/OL]. (2015-03-08) [2016-03-08]. http://www.codata.org/uploads/Press%20Release%20CODATA_RDA_WDS-FINAL.pdf.
- [54] Open Data Institutes. ODI Nodes — collaborative reach, influence and impact[EB/OL]. (2013-10-23) [2016-03-07]. <http://theodi.org/nodes>.
- [55] Jongman B, Hochrainer-Stigler S, Feyen L, et al. Increasing stress on disaster-risk finance due to large floods[J]. *Nature Climate Change*, 2014(4): 264-268.
- [56] Hack for med[EB/OL]. [2016-07-16]. <http://www.hack4med.homerproject.eu/info>.
- [57] Everyaware[EB/OL]. [2016-03-08]. <http://www.everyaware.eu>.
- [58] European Commission. Have your say on the future of science: public consultation on Science 2.0[EB/OL]. (2014-7-3) [2016-04-16]. <https://ec.europa.eu/digital-single-market/en/news/have-your-say-future-science-public-consultation-science-20>.
- [59] CGIAR. Open access and open data[EB/OL]. (2013-11-27) [2015-12-27]. <http://www.cgiar.org/resources/open>.
- [60] Zenodo-Sharing research data across Europe-making science more visible[EB/OL]. (2013-5-8) [2016-07-16]. <http://www.openaire.eu/en/component/content/article/9-news-news-events/457-zenodo-is-launched>.
- [61] Open GLAM. German Digital Library releases API[EB/OL]. (2013-12-04) [2015-12-27]. <http://openglam.org/2013/12/04/german-digital-library-releases-api>.
- [62] Crookston M. Our new use and reuse policy[EB/OL]. (2014-05-20) [2016-04-22]. <http://natlib.govt.nz/blog/posts/our-new-use-and-reuse-policy>.
- [63] PLoS. Making data count: PLoS, CDL, and DataONE join forces to build incentives for data sharing[EB/OL]. (2014-10-07) [2015-11-04]. <http://blogs.plos.org/tech/making-data-count-plos-cdl-and-dataone-join-forces-to-build-incentives-for-data-sharing>.
- [64] Mayernik M S, Callaghan S, Leigh R, et al. Peer review of datasets: when, why and how[EB/OL]. (2014-05-07) [2016-04-16]. <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00083.1>.
- [65] Dodds L. CC 4.0 and open data[EB/OL]. (2013-12-06) [2015-12-27]. <http://theodi.org/blog/cc-40-and->
-

open-data.

- [66] Highwire. HighWire releases COUNTER 4 usage reports[EB/OL].(2014-02-09)[2016-03-07].<http://highwire.stanford.edu/PR/COUNTER4HighWire.pdf>.
- [67] DataCite. CrossRef and DataCite announce new initiative to accelerate the adoption of DOIs for data publication and citation[EB/OL].(2014-11-10)[2015-12-27].<https://www.datacite.org/CrossRefDataCiteinitiative/>.
- [68] Open Knowledge Institutes. Announcing the new open data handbook[EB/OL]. [2016-05-13]. <http://blog.okfn.org/2015/05/13/announcing-the-new-open-data-handbook>.
- [69] Open Data institutes. Data jam! Exploring creative uses of open data[EB/OL].(2014-05-29)[2016-05-16]. <http://theodi.org/blog/guest-post-data-jam-exploring-creative-uses-of-open-data>.
- [70] Data Seal of Approval. Implementation of the data seal of approval[EB/OL].(2014-01-16)[2016-03-08].https://assessment.datasealofapproval.org/assessment_101/seal/html.
- [71] 4C Project Draft Roadmap. Investing in curation;a shared path to sustainability[EB/OL].(2015-01-30)[2016-03-08]. <http://wiki.lib.sun.ac.za/images/d/dc/2015-digital-curation-roadmap-eu.pdf>.
- [72] Elsevier updates text-mining policy to improve access for researchers [EB/OL].(2014-02-03)[2016-03-07]. <http://www.elsevier.com/connect/elsevier-updates-text-mining-policy-to-improve-access-for-researchers>.
- [73] Elsevier opens its papers to text-mining[EB/OL].(2014-02-03)[2016-03-07].<http://www.nature.com/news/elsevier-opens-its-papers-to-text-mining-1.14659>.
- [74] The Hague Declaration.The Hague declaration on knowledge discovery in the digital age[EB/OL].(2015-05-06)[2016-07-16]. http://thehaguedeclaration.com/wp-content/uploads/sites/2/2015/04/Liber_Declaration_A4_2015.pdf.
- [75] LIBER. LIBER statement on the digital single market strategy for Europe[EB/OL].(2015-05-18)[2016-05-18].<http://libereurope.eu/blog/2015/05/18/liber-statement-on-the-digital-single-market-strategy-for-europe>.
- [76] Nearly all German National Library metadata now available under CC0 license[EB/OL].(2014-01-16)[2016-04-07]. <http://openglam.org/2014/01/16/nearly-all-german-national-library-metadata-now-available-under-cc0-license>.
- [77] Open Data Institute. ODI helps unlock £11m for open data innovation in Europe[EB/OL]. [2016-05-04]. <http://theodi.org/news/odi-helps-unlock-11m-for-open-data-innovation-in-europe>.
- [78] European Data Science Academy. Brief description[EB/OL].(2015-02-25)[2016-03-03].<http://edsa-project.eu/overview/about-edsa>.
- [79] Calarco P, Shearer K, Schmidt B. Librarians' competencies profile for scholarly communication and open access [EB/OL].(2016-06-20)[2016-08-10]. https://www.coar-repositories.org/files/Competencies-for-Scholar-Comm-and-OA_June-2016.pdf.

顾立平 中国科学院文献情报中心科技信息政策研究咨询与服务中心执行副主任,中国科学院大学经济与管理学院岗位教授。北京 100190。

(收稿日期:2016-08-03)