

·专辑:人工智能与情报学·

国内外开放数据的研究进展与述评*

濮静蓉 刘桂锋 钱锦琳

(1.江苏大学科技信息研究所 江苏镇江 212013)

摘要:文章采用文献调研和网络调查相结合的方法,从开放数据相关理论、许可协议的制定、开放数据的开放内容、开放途径以及实践案例等方面对国内外开放数据的研究进展进行梳理和分析。文章认为国外开放数据发展在理论和实践研究方面已趋于完善,国内研究则侧重于理论以及对国外开放数据的实践分析,缺少符合本国实际情况的开放数据实践。通过分析也发现目前国内外开放数据主要以政府数据开放为主,科研数据开放与共享相对较少,在研究过程中未充分考虑全生命周期理论的阶段开放。

关键词:开放数据;政府数据;数据管理;科研数据;数据科学;数据共享

中图分类号:G255.76

文献标识码:A

DOI:10.11968/tsyqb.1003-6938.2017123

Research Progress and Enlightenment of Open Data at Home and Abroad

Abstract Through analyzing theories about open data, the formulation of the license agreement, content of open data, way of open data and practice case deeply, the paper provides advices and inspiration for China's all-round development of open data. Using the method of literature research and network investigation, this paper systematically combs the current situation of open data at home and abroad. The development of open data in foreign countries has been perfected in theory and practice research. Domestic research focuses on the theory and the practice analysis of foreign open data, and lacks open data practice in line with the actual situation of the country. At present, the open data development at home and abroad is dominated by government data, while the open and shared data of scientific research data are relatively few. However, the whole life cycle theory is not fully considered in the research process.

Key words open data; government data; data management; research data; data science; data sharing

在大数据的时代背景下,开放数据的浪潮席卷全球,标志性事件是2009年美国政府门户网站的正式上线,自此之后,各国政府、企业、高校等都纷纷加入这场浪潮中。目前,国外的理论和实践研究都取得了一定进展,如基于开放数据倡议,研究人员从可重复性标准的制定、“智能开放”科学的发展以及数据开放带来的透明度不同角度分析了数据开放的必要性^[1-2];Wessels B^[3]和Peng G^[4]剖析开放数据的兴起为相关利益者带来的利益和挑战等。国内学者^[5-6]从各个角度对国外典型国家开放数据发展现状进行对比分析,发现国内的不足并提出相应的措施。不容忽视的是,开放数据带来机遇的同时也带来了巨大挑战,普遍存在着“不愿开放、不敢开放、不知开放什么与不会开放”的实际现状。鉴于此,笔者采用文献调

研和网络调查相结合的方法,系统梳理国内外开放数据的研究现状,从理论基础、许可协议、开放内容、开放途径和开放实践角度,深入总结与剖析国外的研究成果和实践经验,并结合我国数据开放的实际情况提出对策与建议。

1 国内外开放数据的理论探索

基础理论的发展是推动开放数据实践成熟的重要基础和研究重点。随着开放数据运动的持续发展,研究人员对开放数据的理论探索已基本趋于成熟。近年来,学界在开放思想的指导下,围绕政府、企业、高校等领域的数据开放展开大量理论与实践探索,笔者将代表性的理论研究成果整理(见表1)。

众所周知,数据是已发布的科学知识的佐证,而

* 本文系国家社科基金项目“开放科学理念下的科研数据治理研究”(项目编号:17BTQ025)研究成果之一。

收稿日期:2017-12-20;责任编辑:魏志鹏

数据的开放将提高透明度和再现性。开放数据理论丰富,涉及的内容层面较多,如相关概念、图书馆角色的转换等,虽涉及领域较多,如政府、高校、企业等,但其本质特点是不变的、透明和可再利用。此外,研究人员还进一步探究开放数据的模型框架与阻碍其不断发展的影响因素。同时,在大量的理论探索中,部分学者对图书馆与开放数据的联系进行了研究,指出图书馆与图书馆员都应变更传统模式,顺应数据潮流,在开放数据发展中扮演好自身角色。影响数据开放的最关键因素是科研人员自身意愿和行为,而影响其开放意愿和行为的因素以及数据共享的障碍又可归纳为个人利益和风险的感知。研究过程中重点分析影响科研数据开放的内外要素,细化并识别科研数据开放的关键要素,从而推动开放数据的发展,促进数据的再利用。国内外理论探索成果已经相当丰富,但在对开放数据研究过程中未能充分考虑全生命周期理论的阶段,不同阶段图书馆和图书馆员职能不一样,影响因素也不尽相同。

2 国内外开放数据许可协议和政策研究

许可协议和政策是开放数据发展的重要机制和保障。许可协议不仅是开放数据的核心,更是开放数据增值的保障,同时对促进开放数据的发布和今后的再利用具有重要意义。国内已有研究人员在国外图书馆领域和政府领域制定的开放数据许可协议的背景下展开研究,并与我国的制定现状进行对比,发现存在的问题和不足,为我国各领域制定许可协议提供借鉴和经验。由于目的不同,目前开放数据的许可协议类型可概括为政府制定和非政府制定两种类型,由英国政府制定的开放政府许可协议(Open Government License,OGL)是政府制定的许可协议代表,而非政府制定协议的代表有知识共享组织(Creative Commons,CC)和开放数据共享(Open Data Commons,ODC)。研究人员发现国外部分图书馆、学术出版与学术资源联盟制定许可协议都参考和引用了CC许可类型^[24],而国外政府制定开放数据许可

表1 国内外开放数据理论研究概览

研究角度	研究人员	研究内容
基本概念与理论	Ayre L B ^[7]	开放数据是由公共实体发布的可获取和再利用的数据。
	Rouder J N ^[8]	开放科学能使研究尽可能透明、开放数据将原始数据提供给任何人、天生数据是由自动化系统产生的数据。
	马海群 ^[9]	通过“数据”和“开放”两大要素阐释开放数据的内涵,研究大数据、开放源代码、开放获取以及信息公开相关概念。
模型与框架	赵龙文 ^[10]	基于国外相关性模型的研究,提出开放数据的相关性四维模型的构建,包括数据资源(A、B两维)、时间以及RDF三元组四个维度。
	蒲攀 ^[11]	提出国内开放数据政策的“系统-制度-过程(S-R-P)理论模型”。
	Traverso-Ribón I ^[12]	提出面向学习的协作评估方法的开放数据框架,应对项目产生大量数据无法进行可持续评估现象。
	Davies T ^[13]	重新审视开放数据的新兴影响(ODDC)的概念框架。
	Sieber R E ^[14]	提出改变公民和政府关系的四种开放数据模型,以提高数据的利用率。
图书馆角色与职能	刘春丽 ^[15]	在开放数据和开放科学环境下,专业图书馆将扮演知识服务中心以及开放数据的管理与保存中心的角色。
	MacDonald S ^[16]	研究人员、图书馆员、技术人员为管理数据泛滥现象需变革原有职能,同时图书馆在电子研究活动中扮演数据管理角色。
科研人员的意愿与行为	张晋朝 ^[17]	科研人员的数据共享意愿决定着他们的共享行为,主观规范对数据共享意愿存在直接显著性影响,其他因素只是间接影响。
	周姗姗 ^[18]	在文献分析和访谈研究基础上,发现科研人员数据共享意愿中存在邻避现象。
	Kim Y ^[19]	提出影响STEM(科学、技术、工程和数学)研究人员数据共享行为的因素可概括为信仰、态度、规范以及资源因素。
障碍与影响因素	Janssen M ^[20]	归纳出阻碍科研人员的数据开放行为的主要因素包括机构、处理数据的复杂性、开放数据的使用和开放数据的参与、立法、信息质量和技术水平。
	Conradie P ^[21]	提出影响数据发布共享的关键指标有:存储数据、获取数据以及部门使用数据的方式。
	Williamson K ^[22]	指出数据共享的障碍包括:缺乏共同的学科背景、信任、责任感和受控的数据访问,以及描述数据以实现再次利用。
	Kim Y ^[23]	提出影响不同学科背景的研究人员数据共享的因素包括制度因素和个人因素。

协议时同时借鉴了 CC 和 ODC 两种协议^[25]。根据调查,开放数据政策研究主要集中在政府领域,张晓娟^[26]、程银桂^[27]在研究中都指出在政府数据开放发展过程中隐私权的保护越来越受到公众的关注,隐私权政策的制定是推进政府数据开放有序进行以及平衡各方利益的重要保障。无论是许可协议的制定,还是政策的制定,都将会推动数据开放的发展以及数据的再利用。尽管与国外研究相比,我国在开放数据许可协议和政策上的制定还不够完善,相关研究也不够深入,但是国内相关研究工作势头依然保持良好,可在借鉴国外发展提供的经验和启示的基础上,结合我国实际情况对已制定的政策和许可协议进行调整和完善。

3 国内外开放数据内容的探索

开放数据内容的研究是整个开放数据运动发展中的重中之重,可分为开放对象、开放类型和开放格式的研究。开放数据的开放不仅指数据的开放,还应该包括相关服务的开放,如在开放数据服务方面,图书馆需要向用户提供馆藏目录、数据监管、开放获取、知识发现、语义分析等服务^[28]。数据开放内容的研究目前重点集中于对元数据的开放,数据是各个研究领域不可缺少的元素之一,而元数据是描述数据的数据,更是与科研过程息息相关。于梦月^[29]基于元数据存在方案不同、对数据描述不全面、格式单一等问题,设计出以 W3C DCAT 标准为主的核心元数据方案;马海群^[9]认为,按数据权利主体的划分,开放数据类型可归纳为政府数据、公共数据、科学数据、商业数据和个人数据,部分数据类型可重合或者在一定的条件下进行转换。值得注意的是,尽管开放数据的内涵是被任何人免费获取、再利用,但是数据开放类型实质上是受到限制的,凡涉及国家安全、商业机密、个人隐私安全等重要数据并不能向任何人公开,只允许部分指定人员能获取。数据格式的多样性和开放性将有利于降低用户对数据再利用的阻碍,常见的开放数据格式有 CSV、RDF、JSON、KML/KMZ、XML、HTML、ZIP 等^[9]。对开放数据内容的探索将会满足用户需求、有助于科研人员对数据的再次利用,并且促进开放数据运动的发展。

4 开放数据的途径与模式

随着互联网的迅速发展,建立合理安全的开放途径是推动数据开放、共享和再利用的必要条件。基于国内外研究成果,笔者将从开放途径与模式涉及到的三个方面展开分析:数据开放平台、出版与机构知识库、模式与机制。

4.1 数据开放平台

数据开放平台的适用范围较广泛,如政府数据、高校科研数据、商业数据、网络数据和个人数据等多种类型的数据都可通过开放平台进行开放,供用户获取和再利用。Van Schalkwyk F^[30]指出生态领域的主要利益相关者正在使用数据平台提供的政府数据进行再次使用,推动生态领域的发展。而国内已设有北京、上海等 7 个具有代表性的政府数据开放平台^[31]。同时研究人员对用于科研数据开放的数据平台也进行了深入探讨,如对图书馆数据开放平台系统构建的研究,并提出由开放数据安全管理与防御平台基础结构、数据类型转换与标准统一平台、数据中心存储系统、数据开放与应用技术、数据开放核心服务层以及开放数据服务平台六部分组成构建的平台系统^[32]。罗鹏程^[33]则从开放数据平台的应用角度出发,结合研究数据管理平台 Dataverse,构建提供数据存储、共享等服务的北京大学开放研究数据平台。此外,信息在国际发展中扮演着重要角色,探讨援助机构如何借助开放数据平台,对数据进行再利用,以提高援助的有效性^[34]。大数据背景下,用户对数据获取的期望值正不断提高,对获取数据的便利需求也提出要求,因此开放平台的设立一方面为用户查询和获取数据提供途径,另一方面为政府数据、科研数据以及机构数据等类型数据的开放提供一个平台。但通过对比和分析发现,国内数据开放平台的发展存在诸多问题,如开放的数据量欠缺、规范管理差、数据更新落后、技术支持不完善等,国外政府数据开放平台发展的成熟,可为我国各种用途的开放数据平台的构建提供借鉴。

4.2 出版与机构知识库

出版与机构知识库开放途径多被用来开放科研过程中产生的数据。张静蓓^[35]指出,数据出版将会提高数据的再利用,通过综合分析,现有 4 种出版模

式:数据独立出版、数据论文出版、合作出版以及期刊自行出版,数据出版的一般流程需要经历5个阶段,分别是数据提交、同行评议、数据发布和永久存储、数据引用、影响评价。迄今为止,机构知识库未有一个确切的定义,但目前国内很多高校设置了机构知识库,如北京大学、厦门大学、中国科学院文献情报中心,高校认为机构知识库的设立是为其各部门及其成员产生的数字化产品进行管理和传播。同时国外已开展相关研究,如Konstantinou N^[36]研究发现,将现有机构存储库安装转换为链接打开数据存储库的转换引擎,能将学术信息以关联开放数据的形式进行公开。印度是第二个在机构存储库(IRs)和数字图书馆(DLs)采用DSpace开源软件的国家^[37]。基于学者的分析,可见机构知识库是由机构建立,以网络为载体,对在工作中所产生的各种数字化内容进行收集、整理、保存、检索以及提供利用。而出版不仅以网络为载体,还可以纸质为载体,同样对数据进行收集、开放和利用。

4.3 模式与机制

开放共享能够促进科学数据的有效交流和转换,因此开放共享模式以及保障机制在开放数据发展进程中是必不可少的途径,并且它的适用范围相对较宽。王晴^[38]指出一个国家或地区的科学数据开放共享一般由政府主导下的公益性运行模式和市场主导下的商业化运行模式运行,我国由于受到相应的政策法规、数据结构和社会环境等因素的影响,目前运行模式是以公益性为主。在大数据环境下,数据开放共享的公益性模式和商业化模式可具体概括为国家政策驱动模式、部门之间交换模式、企业发展带动模式、国际组织参与模式,它们相互交叉和渗透,满足公众对信息共享的需求^[39]。Janssen M^[40]则分析商业领域数据开放利用的模式,认为新的和信息性的商业模式的兴起是由公共机构对开放数据刺激创新引起的,通过研究荷兰的利用模式确定6种商业模式,发现混合商业模式的公共和私人组织都有助于创造价值。数据开放共享需要一系列的保障机制来运行,研究人员发现,在推动开放数据运动合法运行以及保障数据生产者和利用者合法利益的前提下,法律许可机制、合作机制、监督机制以及问责制

这四种机制能贯穿开放数据的整个过程,推进开放数据的发展。岳丽欣^[41]从控制论的角度探讨政府数据开放保障机制,并分析美、英、印三国的保障机制是由政策和法律保障、全国统一的政府数据开放平台、完善的执行与合作机制、监督机制共同构成。数据的开放共享不仅改变了学术交流环境,也为社会带来了一定的经济价值。一方面开放共享模式的构建为用户提供了获取和再利用数据的渠道,另一方面数据开放共享的发展必然离不开相关机制的保障,两者协作发展。

数据是一种宝贵的资源,更是一种具有战略性的资产,无论是政府、高校,还是机构、企业只有将自己拥有的数据借助相关的开放途径进行开放才能发挥其潜在的价值,实现数据的增值,提高政府透明度和公众参与度,推动科研内外部的交流和共享,促进社会经济的可持续发展。

5 国内外开放数据案例研究概览

5.1 国外案例研究

研究人员发现,对开放数据实践的研究将有助于理论的完善,从而再次指导实践的开展。自2009年美国数据开放门户正式上线后,全球范围内的政府数据开放门户逐渐上线,数据开放实践研究日益增加(见表2)。

开放数据由于其重要性和必要性受到越来越多的关注,相应的实践研究逐渐在美洲、欧洲和非洲等地区开展。尽管文化背景、经济基础等各方面存在差异,但实践开展主要以政府数据的开放为主。并且政府的重视程度较高,不同国家、地区层面的政府为提高政府工作效率、增加数据透明度以及加强政府与公民之间的互动性,相应提出或借鉴适合自身实际情况的开放数据项目和计划。目前美洲的研究对象较丰富,不仅包括政府数据,还包括医学数据,研究指出医学领域数据开放并不积极,导致该领域出现内部数据交流和共享不畅现象,不利于医学的发展和传播,开放数据项目的研究对研究人员传统观念产生冲击;非洲则通过政府数据门户和平台的建设推动政府数据开放,提高数据透明度和问责制;欧洲侧重通过实践研究解决目前政府数据开放中遇到的

数据质量、数据分布、开放水平以及数据访问等问题。根据国外学者的研究可知,政府工作效率、政府透明度、公民参与度、数据质量、数据的共享和利用等都与数据开放紧密联系。

5.2 国内实践研究

相较于国外各种开放数据实践的开展,目前国内基于开放数据实践的研究成果非常有限,仍处于探索的初级阶段,已有研究主要集中对国外开放数据实践的分析,而针对我国的实践较少(见表3)。

可见,目前国内开放数据的实践内容仍侧重政府数据开放的研究。由于美、英、加拿大等发达国家在政府数据开放的实践上一直处于领先地位,国外有关开放数据实践的研究成果也很丰富,国内研究人员通常对国外的实践进行深入分析和对比,归纳出我国政府数据开放实践的不足,包括数据量缺乏、

数据主题分类不明确、支撑技术欠缺、制度保障不完善。结合国外实践的优劣势以及我国实践现状,为我国今后开放数据的发展提供参考和启示。

开放数据实践研究可以帮助政府、企业和高校在厘清开放数据概念的基础上认识数据开放的必要性和重要性,指导他们开展高效的数据开放共享工作。根据国内外实践的研究,国外更侧重于开放数据计划、项目以及倡议等实践,而国内学者少数是对国内政府数据开放平台的研究,更多的是侧重对国外实践的剖析。对于数据的开放,尽管各国的文化背景、经济发展、政策环境不同,但在实践研究重点上仍有共同点,即首先研究政府数据开放。

6 结论与展望

在 21 世纪逐渐兴起的开放科学运动,使得数据

表 2 国外开放数据案例研究

地区	研究者	案例名称	研究内容
非洲	Bello O ^[42]	开放政府数据门户举措	非洲要求通过公开数据来实现数据的透明度和问责制,在创建数据门户方面正作出努力,部分国家以格式的开放实现开放,并在门户上增加更多的数据集。
	Ohemeng F L K ^[43]	加纳开放数据倡议(GODI)项目	建议政府应该以自主创新为平台建立 ICT 基础设施来改善政府和公民的互动以及确保知识共享和所有社区的平衡,同时刺激需求。
美洲	Kassen M ^[44]	芝加哥地区开放数据项目	在地方上建立一个数据开放的平台有利于政府和公民的沟通、促进公民参与的积极性、加强政府和公民的互动性。
	Peled A ^[45]	美国开放数据项目	美国联邦政府存在不能存储、移动和访问信息的问题,并且多数机构忽略开放数据程序,与政府倡导的透明度产生冲突,数据贸易相撞与现有的机构间协作的做法将解决这一冲突。
	Krumholz H M ^[46]	耶鲁开放数据获取(Yoda)项目	临床医学研究人员未积极共享实验数据,不利于数据的再利用和领域的发展,该项目帮助他们意识到数据开放和共享的重要性。
欧洲	Jetzek T ^[47]	丹麦基础数据计划(BDP)	BDP 的目的在于提高政府数据质量以及数据平台的可访问性,解决目前政府数据开放已经存在的相关问题。
	Martinez-Mendez F J ^[48]	西班牙公共部门开放数据	开放数据集分布不均匀,需要通过元数据分类进行分组从而开展研究和分析,并且可能由于数据免费获取导致相关应用程序的开放水平较低。

表 3 国内开放数据案例研究

研究者	案例	研究内容
武琳 ^[49]	Smart Disclosure	对英美三个应用 Smart Disclosure 的典型项目“midata”、Money Advice Service 和“蓝绿按钮”深入分析,发现政府进行数据开放能获得更多的经济价值。
黄如花 ^[50]	政府数据开放平台	分析英美数据对象、数据组织、数据检索、数据开放利用、数据分享反馈以及数据管理功能,了解英美政府数据开放平台发展现状。
周文泓 ^[51]	政府数据开放	从开放政府动议与行动计划发布与完善、开放数据门户网站的建立、G8《开放数据宪章》签署、《开放数据宪章——加拿大行动计划》发布梳理加拿大政府数据开放的实践,分析加拿大政府开放数据的优势和劣势。
熊易 ^[52]	国际实证资源	基于开放科学数据内涵和意义,探讨“联合国数据”、“经合组织图书馆”、“国际货币基金组织数据库”、“非洲开放数据”开放科学数据实证资源及其利用。
郑磊 ^[53]	政府开放数据平台	利用开放政府数据评估框架分析和比较我国 13 个地方政府开放数据平台的指标,发现目前国内的实践面临各种挑战,有待解决。
晴青 ^[54]	政府数据开放	对开放北京市政府数据的三类网站进行调研,梳理分析其公布的数据/元数据数量、分类、增值利用能力等指标,从技术和管理层面为北京市政府数据开放的发展提出建议。

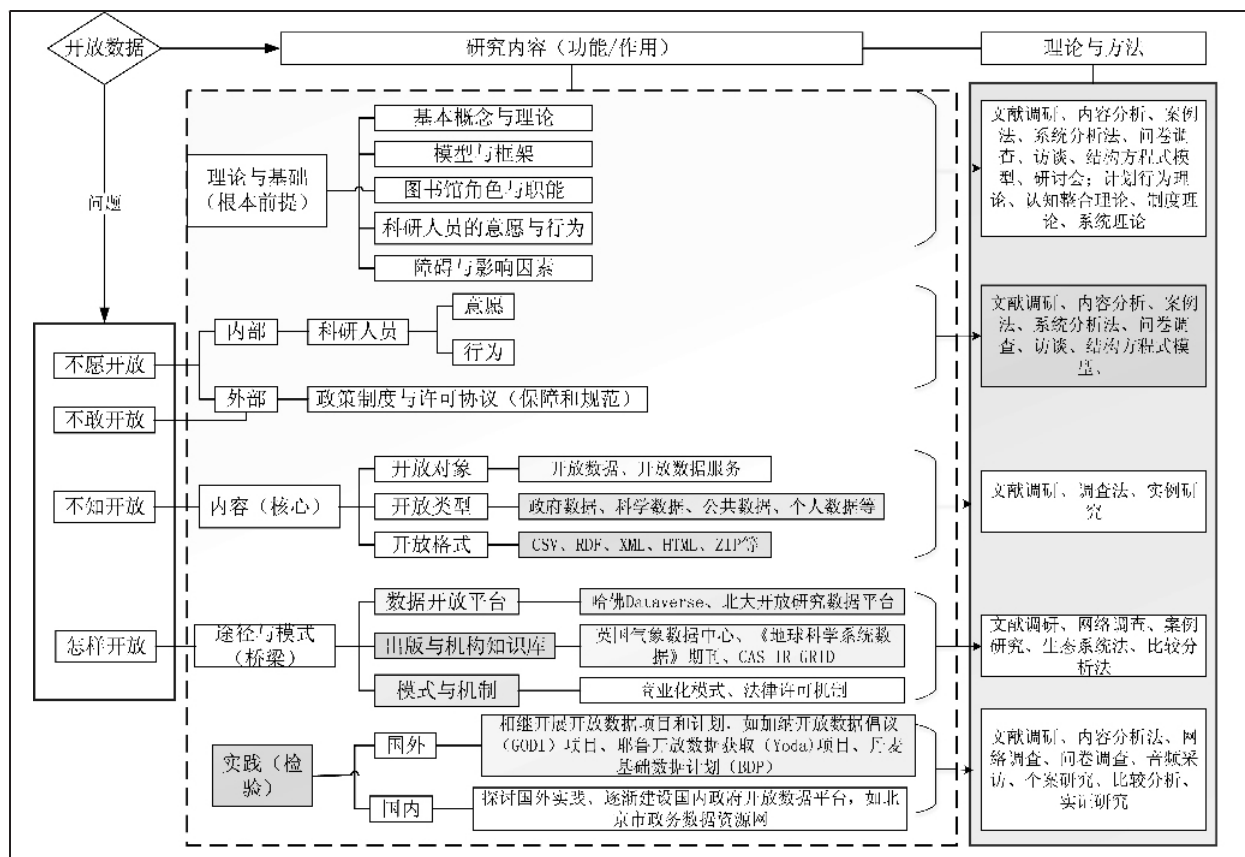


图1 开放数据研究框架

的开放共享成为不同领域的共同关注点。开放科学是由开放数据等不同个体共同组成的，其中开放数据处于开放科学的高级阶段。目前，在生态学、医学以及政府领域，开放数据的重要性已得到广泛认可，而在高校、科研等领域，数据的开放共享仍需提倡。此外，国内学者的研究已不再局限于理论的探索，更多的研究人员倾向于以国外实践研究案例为参考，将我国的实际情况和行业具体特征纳入研究范围，并在此基础之上，完善开放数据的概念内涵，为构建开放数据的模型与框架提供思路与建议等。图1是在对国内外开放数据理论与实践研究的基础上，总结概括了开放数据研究框架。

开放数据研究框架图涵盖了开放数据目前研究的主要问题、并从多角度阐述已有的研究内容，同时对研究过程中使用的相关理论与方法进行分析，内容丰富且较全面。

6.1 开放数据的模型与框架：开放共享的重要保障

通过研究发现，开放数据的模型与框架虽具有普适性，但各领域的行业特点不容忽视，所以该模型

与框架的唯一性，亦称之为特殊性应在构建之处纳入思考范围。模型和框架的构建有利于开放共享过程中对数据的管理、分析和理解，是开放的重要保障。基于国内开放数据运动发展现状，构建开放数据的模型框架存在诸多困难，因此需要结合国内实际情况、借鉴国外模型框架建立的经验，针对所需要使用的领域以及使用用途（数据开放、数据分析、数据组织和管理，以及数据评估等），发挥各研究者的优势以及现有的技术水平和设备，构建更多适用于国内开放数据发展的模型框架。

6.2 许可协议和政策制定：开放共享的驱动力量

随着网络技术的发展和开放数据运动的推广，开放数据思想已经逐渐在国内广为人知，但目前国内针对数据开放共享的实践不容乐观，很多政府部门、研究人员或者各种研究机构认为将自己的研究数据、过程或者科研成果等公开后会影响到自己的研究成果以及自身利益，因此呈现出“不愿开放，不敢开放”的现象。许可协议和政策的制定将有助于保障各群体的利益，指导和规范开放共享。开放数据是基

于科学研究和创新的需求而产生的,国外政府部门在早期就起到推动作用,国内推动开放数据运动发展的领域先驱亦是政府,政府可充分发挥其领导和引领的功能,推动政策和开放许可的制定。

6.3 开放数据的内容:开放数据的实质与基石

在开放数据发展过程中最为重要的是开放数据内容,它是开放数据运动发展的基石。研究发现数据开放应该开放采集到的第一手数据,并不全是被修改或者加工的数据;同时并不是所有的数据都能被开放,部分涉及到国家安全、商业机密层次的以及个人隐私或者特别限制的重要数据不能公布,因此在进行数据开放时需要谨慎,需要掌握能够开放的数据类型、格式以及具体内容。

6.4 开放数据的途径:数据开放的依托与桥梁

国内外学者的研究发现,目前有多种途径支持数据的开放和共享,是数据开放的桥梁。国内政府、高校和研究机构在数据开放平台的建立、开放数据出版和机构知识库等开放数据的模式与机制方面的研究成果仍不够成熟,并不是所有的途径都能允许用户自由、免费的进入并获取使用数据,为研究人员的研究带来了诸多不便。有学者指出,由于政府数据开放平台的特殊性,将会面临数据开放的内容、开放平台的安全、数据访问的权限等问题,对用户进行访问控制、身份认证以及用户授权将有助于保障政府数据开放安全^[55]。该研究发现,为保障政府数据开放的安全性,对用户进行限制,只有授予访问权限的用户才能自由

获取数据。基于国内研究的不足,借鉴国外的实践成果对国内已有的相关途径进行改进和完善,建立更多的开放途径,满足数据开放共享以及用户需求。

6.5 开放数据的实践:开放数据理论的升华和检验

在大数据时代,数据开放实践能够将开放数据理论贯彻到底,解决社会公众的数据需求,为社会创造更多的经济价值,这一研究具有重大现实意义。在政府领域,政府数据开放实践有利于提高政府的透明度,提高政府管理效率,增加政府与公众的互动性;在科研领域,数据开放实践有助于研究人员进行学术交流,为更多专业人士或感兴趣的业余人士提供一个学习和参与的平台,提高科研过程的效率;在商业领域,数据开放的实践可推动各企业的竞争,不仅获得更多的经济利益,而且为用户提供更优质的服务。有效的数据开放实践,能够使得数据对公众开放,达到共享和再利用的目的,实现开放科学“自由、平等、合作、共享”的目标与愿景。

推动开放数据发展已成为国家、高校、企业等的首要任务,开放数据的重要性越来越显著。需要对开放数据的全生命周期理论阶段进行分析,实现阶段开放,减少开放风险。界定开放与共享的概念,认识到开放是整个开放数据发展的基础,共享是开放数据的最终目标之一,将数据开放和共享有机结合。在数据开放阶段,数据提供者要保障数据质量、数据格式、数据可访问;在数据共享阶段,数据利用者要遵循相关制度和政策,推动数据共享和利用。

参考文献:

- [1] Boulton G.Reproducibility:international accord on open data[J].Nature,2016,530(7590):281.
- [2] Naik G.Peer-review activists push psychology journals towards open data[J].Nature,2017,543(7644):161.
- [3] Wessels B,Finn R,Wadhwa K,et al. Open Data and the Knowledge Society[M].Netherlands:Amsterdam university Press,2017.
- [4] Peng G,Ritchey N A,Casey K S,et al.Scientific stewardship in the Open Data and Big Data era—Roles and responsibilities of stewards and other major product stakeholders[EB/OL].[2017-10-10].<http://dlib.org/dlib/may16/peng/05peng.print.html>.
- [5] 刘海房,莫世鸿,范冰冰.开放数据最新进展及趋势[J].情报杂志 2016,35(9):163-167.
- [6] 夏义堃.国际比较视野下我国开放政府数据的现状、问题与对策[J].图书情报工作,2016,60(7):34-40.
- [7] Ayre L B,Craner J.Open Data:What It Is and Why You Should Care[J].Public Library Quarterly,2017,36(2):173-184.
- [8] Rouder J N.The what,why,and how of born-open data[J].Behavior research methods,2016,48(3):1062-1069.
- [9] 马海群,蒲攀.开放数据的内涵认知及其理论基础探析[J].图书馆理论与实践,2016(11):48-54.
- [10] 赵龙文,朱丹,陈明艳.开放数据的相关性四维模型构建研究[J].情报科学,2016,34(9):18-22.
- [11] 蒲攀,马海群.大数据时代我国开放数据政策模型构建[J].情报科学,2017,35(2):3-9.

- [12] Traverso-Ribón I, Balderas-Alberico A, et al. Open data framework for sustainable assessment of project-based learning experiences[J]. Program, 2016, 50(4): 380-398.
- [13] Davies T, Perini F. Researching the emerging impacts of open data: revisiting the ODDC conceptual framework [J]. The Journal of Community Informatics, 2016, 12(2): 148-178.
- [14] Sieber R E, Johnson P A. Civic open data at a crossroads: Dominant models and current challenges [J]. Government information quarterly, 2015, 32(3): 308-315.
- [15] 刘春丽, 徐跃权. 开放科学和开放数据环境中专业图书馆的新角色[J]. 图书馆建设, 2014(2): 83-88.
- [16] MacDonald S, Martinez-Urbe L. Libraries in the converging worlds of open data, e-research, and Web 2.0 [J]. Online, 2008, 32(2): 36-40.
- [17] 张晋朝. 我国高校科研人员科学数据共享意愿研究[J]. 情报理论与实践, 2013, 36(10): 25-30.
- [18] 周姗姗, 翁苏湘, 毕强, 等. 科学数据共享中的邻避现象及应对研究[J]. 图书情报工作, 2015, 59(17): 3-9.
- [19] Kim Y, Zhang P. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories [J]. Library & Information Science Research, 2015, 37(3): 189-200.
- [20] Janssen M, Charalabidis Y, Zuidewijk A. Benefits, adoption barriers and myths of open data and open government [J]. Information systems management, 2012, 29(4): 258-268.
- [21] Conradie P, Choenni S. On the barriers for local government releasing open data [J]. Government Information Quarterly, 2014, 31: S10-S17.
- [22] Williamson K, Kennan M A, Johanson G, et al. Data sharing for the advancement of science: Overcoming barriers for citizen scientists [J]. Journal of the Association for Information Science and Technology, 2016, 67(10): 2392-2403.
- [23] Kim Y, Stanton J M. Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis [J]. Journal of the Association for Information Science and Technology, 2016, 67(4): 776-799.
- [24] 杨敏, 夏翠娟, 徐华博. 开放数据许可协议及其在图书馆领域的应用[J]. 图书馆论坛, 2016(6): 91-98, 141.
- [25] 迪莉娅. 政府数据开放许可适用研究[J]. 图书馆, 2014(6): 91-96.
- [26] 张晓娟, 王文强, 唐长乐. 中美政府数据开放和个人隐私保护的政策法规研究[J]. 情报理论与实践, 2016, 39(1): 38-43.
- [27] 程银桂, 赖彤. 新西兰政府数据开放的政策法规保障及对我国的启示[J]. 图书情报工作, 2016, 60(19): 25-31.
- [28] 张嵘嵘, 刘亚丽. 大数据时代的图书馆开放数据服务探析[J]. 图书与情报, 2014(2): 120-122.
- [29] 于梦月, 翟军, 林岩. 我国地方政府开放数据的核心元数据研究[J]. 情报杂志, 2016, 35(12): 98-104.
- [30] Van Schalkwyk F, Willmers M, McNaughton M. Viscous open data: The roles of intermediaries in an open data ecosystem [J]. Information Technology for Development, 2016, 22(sup1): 68-83.
- [31] 杨瑞仙, 毛春蕾, 左泽. 我国政府数据开放平台建设现状与发展对策研究[J]. 情报理论与实践, 2016, 39(6): 27-31.
- [32] 马晓婷, 尚庆生. 大数据时代图书馆开放数据服务平台与开放数据服务模式研究[J]. 图书馆理论与实践, 2015(5): 72-75, 102.
- [33] 罗鹏程, 朱玲, 崔海媛, 等. 基于 Dataverse 的北京大学开放研究数据平台建设[J]. 图书情报工作, 2016, 60(3): 52-58.
- [34] Linders D. Towards open development: Leveraging open data to improve the planning and coordination of international aid [J]. Government Information Quarterly, 2013, 30(4): 426-434.
- [35] 张静蓓, 任树怀. 科研数据出版模式、流程及引用策略研究[J]. 图书情报工作, 2015, 59(9): 21-27.
- [36] Konstantinou N, Spanos D E, Houssos N, et al. Exposing scholarly information as linked open data: RDFizing DSpace contents [J]. The Electronic Library, 2014, 32(6): 834-851.
- [37] Surendran Cherukodan, Humayoon Kabir S. Adoption of DSpace by Indian libraries [J]. Program, 2016, 50(2): 175-183.
- [38] 王晴. 论科学数据开放共享的运行模式、保障机制及优化策略[J]. 国家图书馆学刊, 2014, (1): 3-9.
- [39] 左建安, 陈雅. 基于大数据环境的科学数据共享模式研究[J]. 情报杂志, 2013, 32(12): 151-154.
- [40] Janssen M, Zuidewijk A. Infomediary business models for connecting open data providers and users [J]. Social Science Computer Review, 2014, 32(5): 694-711.
- [41] 岳丽欣, 刘文云. 我国政府数据开放保障机制的建设研究[J]. 图书情报工作, 2016, 60(19): 40-48, 39. (下转第 107 页)

- [42] Bello O, Akinwande V, Jolayemi O, et al. Open Data Portals in Africa: An Analysis of Open Government Data Initiatives[J]. African Journal of Library, Archives and Information Science, 2016, 26(2): 97-106.
- [43] Ohemeng F L K, Ofosu-Adarkwa K. One way traffic: The open data initiative project and the need for an effective demand side initiative in Ghana[J]. Government Information Quarterly, 2015, 32(4): 419-428.
- [44] Kassen M. A promising phenomenon of open data: A case study of the Chicago open data project[J]. Government Information Quarterly, 2013, 30(4): 508-513.
- [45] Peled A. When transparency and collaboration collide: The USA open data program[J]. Journal of the Association for Information Science and Technology, 2011, 62(11): 2085-2094.
- [46] Krumholz H M, Waldstreicher J. The Yale Open Data Access (YODA) project—a mechanism for data sharing[J]. New England Journal of Medicine, 2016, 375(5): 403-405.
- [47] Jetzek T. Managing complexity across multiple dimensions of liquid open data: The case of the Danish basic data program[J]. Government Information Quarterly, 2016, 33(1): 89-104.
- [48] Martínez-Mendez F J, López-Carreno R, Pastor-Sánchez J A. Open data in the Spanish public administrations: main subject categories and apps[J]. PROFESIONAL DE LA INFORMACION, 2014, 23(4): 415-423.
- [49] 武琳, 伍诗瑜. 开放数据服务价值研究——以 Smart Disclosure 为例[J]. 情报杂志, 2016, 35(1): 171-175.
- [50] 黄如花, 王春迎. 英美政府数据开放平台数据管理功能的调查与分析[J]. 图书情报工作, 2016, 60(19): 24-30.
- [51] 周文泓. 加拿大联邦政府开放数据分析及其对我国的启示[J]. 图书情报知识, 2015(2): 106-114.
- [52] 熊易. 国际开放科学数据实证资源及利用研究[J]. 图书馆理论与实践, 2016(1): 12-14.
- [53] 郑磊, 熊久阳. 中国地方政府开放数据研究: 技术与法律特性[J]. 公共行政评论, 2017(1): 53-73.
- [54] 晴青, 赵荣. 北京市政府数据开放现状研究[J]. 情报杂志, 2016, 35(4): 177-182.
- [55] 宋赢硕. 基于分级访问控制的政府数据开放平台及安全设计[D]. 上海: 上海交通大学, 2014.

作者简介:濮静蓉(1991-),女,江苏大学科技信息研究所硕士研究生;刘桂锋(1980-),男,江苏大学科技信息研究所副研究馆员;钱锦琳(1993-),女,江苏大学科技信息研究所2016级硕士研究生。

(上接第111页)

- [28] 马思言,何宁,李亮.人工智能知识管理系统在制造型企业中的应用[J].机械制造与自动化,2008,37(3):10-13.
- [29] 钟英姿.基于数据挖掘的知识管理系统研究[J].科技管理研究,2006,26(11):173-175.
- [30] 李颖,姚艺.国内外知识管理系统研究:回顾与展望[J].图书情报知识,2010(5):103-110.
- [31] Ramanathan V, Liang P, Fei-Fei L. Video Event Understanding using Natural Language Descriptions[C]. International Conference on Computer Vision (ICCV), 2013.
- [32] 杨阳,张文生.基于深度学习的图像自动标注算法[J].数据采集与处理,2015,30(1):88-98.
- [33] 沈敏,杨新涯,王楷.基于机器学习的高校图书馆用户偏好检索系统研究[J].图书情报工作,2015,59(11):143-148.
- [34] 吴少俊.动态工作流办公自动化系统关键技术的分析[J].数字技术与应用,2015(9):120.
- [35] Shen Y, He X, Gao J, et al. Learning semantic representations using convolutional neural networks for web search[C]. Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014: 373-374.
- [36] Elkahky A M, Song Y, He X. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C]. Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015: 278-288.
- [37] 任宗强,刘冉.人机交互模式下企业知识管理平台研究[J].技术与创新管理,2017(5):526-529.
- [38] 李伟超,牛改芳.智能代理技术分析及应用[J].情报杂志,2003,22(6):29-30.

作者简介:欧阳智(1987-),男,贵州大学贵州省大数据产业发展应用研究院校聘副教授,博士,研究方向:知识管理、大数据治理;魏琴(1983-),女,贵州大学贵州省大数据产业发展应用研究院副教授,研究方向:机器学习、人工智能;肖旭(1994-),男,贵州大学现代制造技术教育部重点实验室硕士研究生,研究方向:机器学习。