

国内外数据治理研究述评*

林伟^{1,2} 周耀铭¹

(1. 福建警察学院, 福州 350007; 2. 西南政法大学, 重庆 401120)

摘要: 随着全球经济发展和互联网的广泛应用, 数据资源在经济运行和资源配置中起着日益重要的作用, 数据治理成为企业治理和国家治理的重点领域和重要方式, 数据治理研究已成为热门的研究领域。本文通过对相关文献的梳理分析, 总结出国内外数据治理研究的成果主要聚焦于法律、技术、管理三个层面, 其中个人信息权益、企业数据权属和政府数据开放是法律层面的研究热点, 数据采集技术、数据存储技术和数据处理技术是技术层面的研究热点, 数据质量管理、数据生命周期管理是管理层面的研究热点。未来还应当更加深入的是对数据安全治理、数据协同治理及数据中台的构建研究。

关键词: 数据治理; 数据安全; 法律规制; 协同治理; 数据中台

中图分类号: G353.1 DOI: 10.3772/j.issn.1673-2286.2022.06.009

引文格式: 林伟, 周耀铭. 国内外数据治理研究述评[J]. 数字图书馆论坛, 2022 (6) : 65-72.

随着全球经济发展和互联网的广泛应用, 国家运行、社会发展和个人生活都高度依赖网络, 数据渗透社会生活的方方面面, 包括金融、交通、医疗、教育、消费等领域, 数据资源已成为经济和社会发展不可或缺的战略资源。世界各国研究发展大数据技术, 运用大数据推动经济发展、完善社会治理、提升政府管理和服务能力已成为趋势。

数据治理的内涵是指以“数据”为对象, 在确保数据安全的前提下, 建立健全规则体系, 理顺各方参与者在数据流通各个环节的权责关系, 形成多方参与者共享共治的数据流通模式, 从而最大限度地释放数据价值^[1]。自大数据时代到来, 国内外学界从不同学科、不同维度对数据治理进行充分研究。有学者从法律视角进行探讨, Zech^[2]提出了“数据生产者权”的概念, 认为应给数据载体上生成的数据承担责任的主体制定可转让的数据专有权。有学者从技术视角进行分析, 如邢春晓^[3]认为智能数据治理需要大数据质量实时分析与控制、算法的可解释性、知识图谱等相关技术的支撑。同时, 还有学者从不同的视角、使用不同的方法对相关研究成果进行综述。从定性研究来看, 学者分别从医

疗^[4]、政务^[5]等不同领域对国内数据治理进行综述。从定量研究来看, 刘强^[6]运用CiteSpace软件量化分析档案数据的国内研究现状及展望。总体来看, 现有的综述成果大多局限于国内某一行业的数据治理研究, 鲜有学者对国内外数据治理研究的整体动态及发展趋势进行系统梳理。为此, 本文通过文献梳理, 从法律、技术、管理三个层面归纳分析国内外数据治理领域的研究状况, 在此基础上对未来仍需深入研究的问题提出展望。

1 数据治理的研究历程

纵观国内外数据治理的研究历程, 大致可以分为三个阶段, 第一阶段与第二阶段以2013年为分界线。在2013年以前(第一阶段), 国内外对数据流动、数据治理等相关问题的研究仍处于相对匮乏的状态。国外学界对数据治理的认识始于2004年, 之后陆续有学者对数据治理展开研究^[7]。该阶段的研究多处于零星状态, 并未呈现规模之势。直到2013年, 大数据技术迅速发展, 有媒体称该年为“大数据元年”。从此数据治理研究也进入第二阶段, 人们逐渐意识到以传统关系型数据库

* 本研究得到教育部人文社会科学研究项目“网络空间中数据安全生态构建研究”(编号: 20YJC820028)资助。

为核心的数据存储与处理技术无法适应多元、异构、海量、高时效等大数据特征和应用需求,因而逐渐加大对数据治理的研究。特别是2013年6月,震惊全球的美国“棱镜门”事件爆发,引发了世界各国对个人数据权利的思考;2015年5月我国工信部电子技术标准化研究院制定《数据治理白皮书》国际标准研究报告。但是,在该阶段,国内外对数据治理的研究仍然处于技术层面上的思考,世界各国仍未对数据治理进行系统性、全面性以及学科交叉性的研究。2017年以后,随着大数据应用的不断深入,数据作为战略资源的地位日益凸显,数据共享与开放、安全与隐私保护、数据确权等问题引发了人们的深度思考,至此数据治理研究也进入了第三阶段。学术界和产业界的学者开始认为,大数据治理应该从管理、规范和技术等多个维度进行综合考虑与研究。在国内,中共中央政治局2017年就实施国家大数据战略进行第二次集体学习;2018年,贵阳大数据交易所首次在交易系统中采用区块链技术,区块链技术开始运用于交易;2019年中共十九届四中全会审议通过《中共中央关于坚持和完善中国特色社会主义制度 推进国家治理体系和治理能力现代化若干重大问题的决定》,提出要健全劳动、资本、土地、知识、技术、管理、数据等生产要素由市场评价贡献、按贡献决定报酬的机制。这是国家层面首次将数据列为生产要素,数据资源的重要地位得到确立。在国外,美国出台了《加州消费者隐私法案》(California Consumer Protection Act)和《澄清合法使用境外数据法案》(Clarifying Lawful Overseas Use of Data Act),欧盟通过了《通用数据保护条例》(General Data Protection Regulation),表明世界主要经济体越来越重视数据治理领域的立法。

2 数据治理研究主题分析

笔者以中国知网(CNKI)、Web of Science(WOS)作为检索数据来源,检索并梳理数据治理领域的研究成果发现,关于数据治理研究的成果主要聚焦于法律、技术、管理三个层面。

2.1 法律层面数据治理研究

法律法规是数据治理的前提和基础。数据治理的法制建设既要促进数据流通,更好地满足多维大数据

利用的需求,又要对数据利用与个人信息保护、企业利益维护、社会安全保障等目标进行更好的平衡^[1]。法律层面数据治理研究可以从个人、企业和国家这三个层面来分析。

2.1.1 个人信息权益

从个人层面看,数据权属问题体现为数据人权问题。随着大数据在国家治理、现代经济体系运行和民生方面的运用日益广阔、影响日益加深,个人信息泄露事件时有发生,网络黑灰产业屡禁不止,给公民个人信息权益造成严重威胁。^①个人信息保护的相关立法研究一直是学界关注的焦点。2021年《个人信息保护法》的出台便是法学理论界与实务界集体智慧的结晶。国内学者周汉华^[8]在《法学研究》上发表论文指出:个人信息保护法于信息控制者而言,应强化其内部治理机制,承担应有的法律责任;于信息主体而言,应确认其信息控制权;而在实施方面,则应当由易到难,循序渐进,推动激励相容机制实现。赵光^[9]从全球数据治理视角下探讨了《个人信息保护法》在个人、企业、国家利益上的平衡,以及个人信息保护和利用的关系平衡。张新宝^[10]也认为应坚持综合立法的思路,充分兼顾信息主体、企业和国家的不同诉求,妥善处理三方关系,平衡好人格尊严、商业价值与公共管理价值的利益诉求。杨震等^[11]认为应该坚持国际接轨原则,立足我国具体国情,主动适应国际化需求,在立法中保障个人信息自决权,包括信息保密权、信息选择权、信息查询权等。^②个人数据权属亦是学者研究的重点。在物理世界中,人们享有人格权与财产权以保护自身肉体尊严;而在虚拟世界中,个人数据权是保护个人数据安全的重要防线。Kerber^[12]认为个人数据财产化在政治与文化上更具吸引力,其主要理由是相较于单纯的宪法隐私基本权,兼具经济诱因的制度设计更能有效减少个人数据被大量非法收集与监控的风险。汪厚冬^[13]认为个人数据应当财产化,这将有利个人数据主体对其数据的控制权,有助于促进数据经济发展等,但个人数据财产权与传统数据财产权应有所不同,除了赋予其私益外,还应根据保护个人数据主体合法权益、社会公共利益、数据经济发展与数据安全等对其设定相关的限制结构。不过,也有学者对此持反对意见,如国外学者Miller^[14]认为不应将个人数据财产化,因为这将不利于数字经济的发展。

2.1.2 企业数据权属

从企业层面看,数据权属问题体现为数据产权问题。数据资源成为相关企业追逐的热点,拥有更庞大数据资源的企业能在行业发展中拥有更多话语权以及更大的竞争优势,因此个人数据在商业化过程中引发了许多数据产权争议问题。相关研究有以下两类。①数据权力平衡研究。学者王磊^[15]认为个人数据商业化利用过程中应平衡数据开发利用方与用户之间,数据开发利用方之间,数据开发利用方与国家、社会公共利益之间的冲突。通过设计分级分类的数据利用规则,建立起保障个人信息权益、数据追溯和共享机制、合理的数据管辖标准的法律框架。Ursic^[16]认为数据可携权可以增加用户自由选择服务商的权力,从而建立更可靠的数字信任环境。②数据滥用分析规制研究。大数据应用中的数据算法可能直接决定或影响预测和决策,因而可能为公民个人权益乃至社会利益、国家安全带来影响和风险。如当前饱受热议的“大数据杀熟”便是利用数据挖掘算法对用户的地理位置、消费偏好等数据的深度挖掘分析而进行差异化定价,导致消费者的知情权、公平交易权等受损。张欣^[17]分析了当前算法决策面临个人主体性的不断丧失、个体不公的结构锁定、传统决策治理框架的频繁失效三重危机,探讨了算法治理一般采用个体赋权、外部问责和平台义务3种范式,提出以技术信任和治理信任根基,以算法监管机制设计保障,有效联结3种治理范式的智慧型型算法治理思路。汤晓莹^[18]指出在职场领域,算法可能对劳动者隐私权、平等就业权等带来挑战,并提出相应法律规制路径。

2.1.3 政府数据开放

面对数据资源带来的智慧城市治理全新赋能,提升政府数据的开放与共享能力便成为推进政府治理体系建设、促进治理能力现代化的重要一环。与此同时,政府层面上的数据在共享中的边界问题和数据在流通时的安全问题也更加复杂。政府数据开放研究主要侧重两个方面。①对政府数据流通的安全性研究。政府数据的累积性、功能的融合性和信息的敏感性逐渐提升。但是,与信息产业界相比,政府缺少足够的技术能力来建构、维护和运用好这些海量的政府数据。因此,在境内外数据流通的安全问题上,应做好相应的安全防范

措施。国内学者程学旗等^[19]表示,要建立数据流通交易规则规范,优化数据共享、交易、流通的相关制度,明确数据权属分配,探索数据交易市场,构建有序的数据流通环境。张铭慎^[20]也认为基于政府数据数量的庞大性、功能的复杂性和内容的敏感性,要求政府数据应分类有序开放,以防止数据的滥用和泄密。②国家数据主权的自主性研究。从国家层面看,数据权属问题体现为数据主权问题。数据主权体现为国家独立自主对其掌控的数据进行管理和使用的权力。以美国为例,2018年3月美国通过《澄清合法使用境外数据法案》,该法案授予美国执法机构单边调取域外数据的权力,建立了执法长臂管辖规则。基于此,国内学者刘天骄^[21]认为我国应当坚持以数据主权为基础构建秩序,但同时还要兼顾数字经济时代效率价值、长臂管辖的立法阻断,从而达到三者之间的有效平衡。吴沈括^[22]也认为我国应立足于维护数据主权的基本价值立场,在统筹研判数据主权、数字经济发展的内在逻辑联系和外在规范支撑基础上,建立行之有效的中国方案。

就法律层面数据治理研究而言,当前研究普遍达成的共识是,数据治理的法制建设关键是在个人数据流通、企业数据利用与政府数据开放三者之间取得平衡的过程。但相关研究也存在如下不足。①数据治理的法学实证研究。对于数据治理模式的创新,无论是宏观还是微观层面,都只是提出了初步的理论研究框架,欠缺法学的实证研究。②比较法视野下的数据治理研究。随着近几年各国数据治理领域立法的不断出台,比较法视野下的数据治理研究明显存在不足。③数据权力的规制研究。数据在生产与创造、记录与传递、处理与使用的数据化过程中产生了对人的支配与控制的权力,即数据权力。在各种利益的驱使下,数据权力不断扩张,从而导致数据权力异化,产生了平台权力私有化等现象,如何有效规制这种数据权力的扩张,还待进一步深入探讨。

2.2 技术层面数据治理研究

技术创新是数据治理的重要保障,能够有效解决数据规范、数据清洗、数据交换、数据集成等问题。从数据使用过程来看,技术层面数据治理研究主要围绕数据采集技术、数据存储技术、数据处理技术进行。

2.2.1 数据采集技术

数据采集是获取数据的首要阶段,获得合法、有效的数据是进行数据治理的前提和基础。长期以来,传统数据采集方式的问题主要集中在采集源、采集终端、采集过程中,包括采集阶段面临的采集效率低下、采集终端安全性低、采集过程的事后监督审计不足等,为数据治理带来了一定的隐患与不便^[23]。因此,数据采集技术创新能够从技术层面上解决数据采集阶段存在的问题与风险,主要体现为:一是通过技术上保障接入安全,如IP设定、连接数量、权限设定等;二是通过数据传输加密的方式入手,如通过HTTPS/SFTP的方式;三是加强数据权限管理,在所操作的账号中进行多权限管理设定;四是采集数据校验,包括数据的一致性校验和合法性校验^[1]。这些方式均在技术上保障了数据采集阶段的合法、有效、高质量,给数据治理提供良好的前提。在提升采集效率方面,张韬等^[24]认为可以通过EtherCAT这种广泛应用的以太网协议,通过映射方式与各个从站设备通信,以提高传输效率和传输速率,达到延迟低、通信时间短、配置灵活的效果。就安全性而言,杜鹏等^[25]认为数据技术在电网运用中可以通过采取安全消息总线,配合标签加密认证机制来提高控制流程的安全性。在数据采集监管方面,徐超^[26]认为可以通过网络爬虫技术,对数据进行预处理,并对数据来源进行实时监控,以保障数据的实时性与安全性。

2.2.2 数据存储技术

数据存储技术是数据治理的关键阶段,保障安全可靠的数据存储是进行数据治理的关键一环,在数据采集和数据应用两个环节之间发挥着承上启下的作用。数据治理过程中数据隐私泄露、数据滥用、数据被篡改等问题频有发生,导致这些问题的主要原因是大数据收集和共享流通过程不透明以及对重要数据所提供的加密存储机制不够完善。基于以上问题,对数据存储技术的研究主要侧重以下两个方面。一是存储方式研究。通过对数据分类分级、数据加密、数据访问控制等技术来解决大数据平台数据在存储阶段的安全风险。如国内学者陈永府等^[27]提出可以通过采用全同态加密算法对数据加密,以增强数据在云端存储和信道传输过程中的安全性。肖亮等^[28]也认为可以通过运用重复数据删除技术、隐藏存储技术、数据加密与密

文搜索技术以及数据完整性审计技术来提升云存储服务的安全性。二是存储效率研究。数据存储的高效性能能够为下一步数据的高效处理运用提供良好的保障与基础,是在数据存储的关键性阶段。如以区块链技术为前沿技术的数据存储技术能够在数据透明性与高效性上提供技术支持,为数据治理提供技术保障。张桐^[29]分析了传统数据中心化的存储模式存在的风险以及采用区块链技术分布式存储优点,指出在区块链技术的加持下我们需突破传统中心化思维,重塑数据治理模式。

2.2.3 数据处理技术

数据处理技术是指组织在内部针对动态数据进行一系列活动的集合的技术,是数据治理的核心过程,确保数据被合适的访问及数据被以正当方式处理是保证数据合理应用、处理安全的基本前提。一是数据整理技术研究。数据整理技术主要包括数据的结构化处理、数据质量评估与数据清洗、数据规范化处理、数据融合与摘取、发布共享等^[30]。国外学者Rzeszotarski等^[31]提出一种称为“Kinetica”的架构,使用该架构可以一次性处理多维度的数据并识别异常值,以支持大规模的机器学习训练。匡俊攀等^[32]提出一种基于深度学习的异常数据清洗算法,通过感知数据的“时-空”相关性和异常值的稀疏性解决异常数据清洗问题。二是数据分析技术研究。数据在社会中扮演重要的角色,但是数据通常并不能直接被人们利用,数据分析是从大量看似杂乱无章的数据中揭示其中隐含的内在规律、发掘有用的知识以指导人们进行科学的推断和决策。何振等^[33]指出在人工智能技术赋能下,政府数据治理呈现出一些新的特征,主要包括数据内容更加丰富、数据处理更加快捷、数据管理更加高效、数据决策更加科学、数据服务更加精准等。Kumar等^[34]提出了一种基主外键的特征选择方法,这样机器学习算法可以同时多个原始数据的子集进行训练,以提高数据分析的效率。

就技术层面数据治理研究而言,学界分别从数据采集高效性、数据存储安全性和数据处理便利性三方面入手研究,阐述了数据治理技术对推动数据治理能力提升的作用。学者普遍对数据处理技术持积极、认可的态度,认为诸如人工智能、区块链等相关技术能够提高数据收集、处理和利用水平,技术的完备性和先进性得到了一定程度的提升。但相关研究仍然存在以下问

题:一是从共享技术的视角来看,由于传统烟囱式的数据技术建设,导致“数据孤岛”仍然存在,即相关数据共享技术仍付诸阙如;二是从集成技术的视角来看,当下仍然缺乏相关技术平台对数据进行整合,导致大数据集成水平不高,而相关共享集成技术的缺位则会导致数据应用效率与水平的低下。因此,如何通过技术消弭“数据孤岛”、打破“数据壁垒”,实现数据更加高效可靠的共享,学界还待进一步深入探讨。

2.3 管理层面数据治理研究

数据管理是数据治理的关键,数据管理能够提供对数据的访问、执行或监视数据存储以控制输入输出操作的整个过程,在整个数据生命周期中,提供符合数据要求的业务数据的规划、获取和管理^[1]。当前数据治理在管理层面的研究主要聚焦数据生命周期管理和数据质量管理。

2.3.1 数据生命周期管理

数据生命周期管理是指对数据在使用过程中的整个生命周期(含数据产生、数据传输、数据存储、数据应用和数据销毁)进行有效管理,发挥数据最大作用的过程。常见的数据生命周期管理模型有英国数据存储中心提出的UKDA模型、美国的雪城大学秦健等人提出的科学数据管理能力成熟度等^[35]。在数据的全生命周期中,可以清洗、转化、合并甚至生成新的数据。在数据流通过程中,赵正等^[36]提出应以数据生命周期为视角,在数据流通情景下数据要素治理应配套促进数据要素合规高效流通,建立数据要素流通管理规则、数据要素流通标识规则、数据要素跨境流通规则等制度。在政府数据生命周期管理方面,夏义堃等^[37]基于数据生命周期理论提出应推进数据资产目录动态管理、数据资产分级分类管理、数据资产质量控制数据治理模式。Shah等^[38]提出了一种数据驱动型政府的数据生命周期管理框架,建立政府大数据生态系统。

2.3.2 数据质量管理

数据质量管理是指对全流程数据的真实性、准确性等特质进行的管理活动,包括数据清洗、数据监测、数据校验等。数据质量管理智能平台能够对数据接入、

数据处理、数据组织等过程的数据进行采样并输入到质量样例数据库,建立数据质量核验任务,自动完成数据质量规范性、一致性、准确性和完整性的检查。不同学者从不同应用场景对数据质量管理进行了探讨。周林兴等^[39]分析智慧城市视域下政府数据质量反馈机制构建、机理模型及运行模式,提出应从反馈组织制度建设、法律遵从、融入人工智能技术、搭建专用反馈平台等方面优化政府数据质量治理。在教育数据质量管理领域,美国加利福尼亚州教育局(California Department of Education)为强化数据质量管理,于2016年1月开始实施“教育数据治理计划”(Educational Data Governance Program),建立数据标准并倡导教育领域利益相关者共同关注教育数据质量管理^[40]。丰佰恒等^[41]结合博弈论模型、改进病毒传播SIR模型(Susceptible Infected Recovered Model)构建科研大数据质量管控模型,并进行仿真模拟。李青等^[42]提出应通过制定和完善数据标准,推进不同层次和维度的数据共享,落实和健全数据隐私保护机制,建设数据治理技术平台来保障数据质量。

就管理层面数据治理研究而言,学者分别从宏观流程管理和微观质量监督两个视角对数据治理进行阐述。通过对数据生命周期的管理,能够加强各环节的高效流通,促进数据要素合规合理配置。对数据质量的管理,能够更好地保证全流程数据的真实性和准确性,通过规范数据应用进而保证数据的质量。然而,数据治理在政府、企业中普遍存在统筹协调能力不足、组织权责不清、组织架构混乱的问题,而目前研究多偏向于针对数据个体特性进行对数据管理客体的管理研究,较少针对政府、企业等管理主体存在的问题进行研究。

3 结论与展望

在数字经济迅猛发展的形势下,数据资源在经济运行和资源配置中起着日益重要的作用。数据治理关涉关系的复杂性使高效的数据治理既要遵从自然、社会规律,也要遵从技术、法律、管理规律。数据治理研究旨在探索发现数据治理规律的科学研究活动,特别是理论研究活动,既无法也不能囿于一两个学科领域去展开,而必须进行多学科跨领域的共同探索研究,因此融入了多学科的研究方法和理论,产生了丰硕的研究成果。纵观国内外数据治理研究的发展历程,它的每一次发展都源于理论与实践相互融合与促进。随着移动

互联网、云计算、人工智能等信息技术的发展,网络空间成为陆、海、空、天之后的第五大空间,与现实空间产生紧密的交互作用。人类固有的社会属性在其开启的网络化生存模式下形成全方位缩影,出现了一个与现实世界平行的“元宇宙”,而数据是连接现实空间与网络空间的桥梁,这注定了数据治理将是未来学术界持续性关注的重要课题。数据治理研究作为一个复杂的系统性工程,今后仍需要从以下三方面进行拓展与完善,以期进一步提升数据治理的效果。

第一,数据安全治理研究。随着大数据的深度应用,网络空间数据安全问题日益凸显。于国家而言,互联网资源储备量已经成为各国实力比拼的新战略要素。某些国家借助自身先进的技术和设备支撑,大肆窃取他国数据,对其他国家信息安全构成极大威胁。可能导致很多国家经济、军事、商业等方面的机密泄露,从而引发重大的国家经济和社会安全问题。于普通民众而言,掌握大量数据资源的企业网络一旦出现非法采集、窃取、贩卖和利用网络个人信息,将给公民的隐私权利、人身和财产安全带来严重威胁。虚拟社会的自由交互也为网络犯罪的滋生提供了绝佳的机会,网络攻击、网络诈骗、网络黑灰产业交易等行为屡禁不止,致使公共网络安全感缺失,引发公众安全危机。于行业层面而言,如何在数据保护和数据的有效利用之间达成平衡是关键。过度的保护显然不利于创新,也不利于数据的挖掘和使用;保护的缺失则易造成侵犯个人隐私,导致失去行业竞争优势,更有可能造成国家数据主权难以有效的维护。大数据作为社会科技进步和发展的重要资料的前提是,只有在安全规范的环境中使用才能整合为战略性资源,否则将成为巨大的公共安全隐患。特别是随着《数据安全法》《个人信息保护法》的相继出台,数据安全治理将是未来数据治理领域持续关注的热点。

第二,数据协同治理研究。前文所述,数据从采集、传输、储存、处理、应用、销毁是处于动态的数据生命周期过程,在这动态流通的过程中其所涉及的主体涵盖个人、企业、机构、组织、政府等。基于此,数据治理应聚合多方主体力量协同进行,通过发挥各方优势、平横各方利益,共同提高治理效率与质量。①多元主体协同。在进行协同治理时应吸纳政府、企业、个人等多方主体参与其中,实现治理主体的多元化。既要发挥政府在公共安全管理事务中的主导作用,保证采集、开发和使用的有序进行,又要充分发挥科技企业、非营利机

构以及社区组织的技术能力,激发创新活力,形成多元主体的协同开发利用模式。②技术协同。数据治理的技术既涉及传统的机器学习技术、加密技术等,又涉及人工智能背景下的隐私计算,如联邦学习、安全多方计算等。引导产学研各界联合推动技术创新相关应用落地,实现技术协同是数据治理的根本。③平台协同。构建以数据为中心的轻量化大数据治理协同平台,通过对数据采集、验证、转换、处理、组织、分析等全过程可视化操作治理,建立全流程闭环的监管体系,定位、分析、跟踪及解决数据治理风险问题,形成数据治理闭环的处理机制,保证数据的持续稳定可靠。

第三,数据中台构建研究。随着信息化的不断发展,各行各业若仍采用传统烟囱式的IT建设方案,“数据孤岛”、数据重复开发建设、数据质量不一致等问题将更加凸显,对计算资源将造成更大的浪费。大数据的本质是数据的融合,把原本各自孤立的数据互相关联、融合,构建数据资产标签类目体系,从而赋予数据更深层次的语义和价值,洞察事物的本质。数据中台的基本理念便是打通“数据孤岛”和“数据烟囱”构建数据共享的统一数据技术架构,让大数据“用起来”“统起来”^[43]。数据中台是通过技术手段对多源异构数据进行采集、集成、存储与处理,同时统一数据标准,形成大数据资产层,为内部和外部提供高效服务的新型数据服务平台。数据中台的功能包括数据融合、数据加工、数据可视化、数据服务化,通过数据融合实现标准化的数据采集以解决“数据孤岛”问题,通过数据加工以统一的数据标准和质量体系实现数据资产化,通过数据可视化展示数据图谱,通过数据服务化实现数据共享、算法共享。2019年是数据中台爆发的元年,数据中台必将依循从概念引爆到迭代试错,再到规模复制的认识路径,对数据中台的研究任重而道远^[43]。

总体而言,我国需要通过数据治理,构建能够保障数据安全、承担数据责任、解决数据问题的多元治理路径,从而进一步提升数据治理的能力;通过多学科融合、全面协调保证决策者遵守组织战略和治理规则,提高数据治理能力,促进数据治理能力的现代化。

参考文献

- [1] 梅宏. 数据治理之论 [M]. 北京: 中国人民大学出版社, 2020.
- [2] ZECH H. A legal framework for a data economy in the European Digital Single Market: rights to use data [J]. Journal

- of Intellectual Property Law & Practice, 2016, 11 (6) : 460-470.
- [3] 邢春晓. 大力推进数据治理技术与系统的学术研究 [J]. 计算机科学, 2021, 48 (9) : 3-4.
- [4] 叶琳, 罗铁清. 医疗数据治理综述 [J]. 计算机时代, 2021 (5) : 10-12.
- [5] 冯淑娟, 黄寿孟. 跨源多维政务数据治理综述 [J]. 网络安全和信息化, 2021 (11) : 76-79.
- [6] 刘强. 档案数据时代: 国内研究现状及展望 [J]. 办公自动化, 2022, 27 (2) : 40-43.
- [7] 刘桂锋, 钱锦琳, 卢章平. 国内外数据治理研究进展: 内涵、要素、模型与框架 [J]. 图书情报工作, 2017, 61 (21) : 137-144.
- [8] 周汉华. 探索激励相容的个人数据治理之道——中国个人信息保护法的立法方向 [J]. 法学研究, 2018, 40 (2) : 3-23.
- [9] 赵光. 全球数据治理视角下的个人信息保护立法 [J]. 探索与争鸣, 2020 (11) : 15-17.
- [10] 张新宝. 我国个人信息保护法立法主要矛盾研讨 [J]. 吉林大学社会科学学报, 2018, 58 (5) : 45-56, 204-205.
- [11] 杨震, 徐雷. 大数据时代我国个人信息保护立法研究 [J]. 南京邮电大学学报 (自然科学版), 2016, 36 (2) : 1-9.
- [12] KERBER W. A new (intellectual) property right for non-personal data? An economic analysis [J]. An Economic Analysis, 2016, 11: 989-999.
- [13] 汪厚冬. 个人数据财产权化的进路研究 [J]. 行政法学研究, 2021 (6) : 90-100.
- [14] MILLER K. Total surveillance, big data, and predictive crime technology: Privacy's perfect storm [J]. Journal of Technology Law & Policy, 2014, 19 (1) : 105.
- [15] 王磊. 个人数据商业化利用的利益冲突及其解决 [J]. 法律科学 (西北政法大学学报), 2021, 39 (5) : 89-100.
- [16] URSIC H. Unfolding the new-born right to data portability: Four gateways to data subject control [J]. SCRIPT-ed, 2018, 15: 42.
- [17] 张欣. 从算法危机到算法信任: 算法治理的多元方案和本土化路径 [J]. 华东政法大学学报, 2019, 22 (6) : 17-30.
- [18] 汤晓莹. 论职场领域算法技术风险法律规制的理念与路径 [J]. 贵州师范大学学报 (社会科学版), 2021 (5) : 137-150.
- [19] 程学旗, 刘盛华, 张儒清. 大数据分析处理技术新体系的思考 [J]. 中国科学院院刊, 2022, 37 (1) : 60-67.
- [20] 张铭慎. 完善我国数据开放、流通和交易制度 [J]. 宏观经济管理, 2022 (4) : 32-39.
- [21] 刘天骄. 数据主权与长臂管辖的理论分野与实践冲突 [J]. 环球法律评论, 2020, 42 (2) : 180-192.
- [22] 吴沈括. 数据治理的全球态势及中国应对策略 [J]. 电子政务, 2019 (1) : 2-10.
- [23] 吴良. 基于大数据平台的信令数据采集技术研究 [J]. 电子技术, 2019, 32 (5) : 89-91, 95.
- [24] 张韬, 刘康, 廖映华. 基于EtherCAT协议的FMS多设备实时数据采集技术 [J]. 机床与液压, 2022, 50 (1) : 112-116.
- [25] 杜鹏, 陶洪铸, 高保成, 等. 面向多应用的通用数据采集技术方案 [J]. 电力系统自动化, 2015, 39 (1) : 26-30.
- [26] 徐超. 大数据背景下审计数据采集技术与方法的研究——以互联网金融企业专项审计为例 [J]. 会计之友, 2020 (19) : 114-119.
- [27] 陈永府, 宋鹏, 王启富, 等. 云环境下的数据防泄密存储技术 [J]. 计算机应用与软件, 2016, 33 (10) : 288-293, 310.
- [28] 肖亮, 李强达, 刘金亮. 云存储安全技术研究进展综述 [J]. 数据采集与处理, 2016, 31 (3) : 464-472.
- [29] 张桐. 数据治理的变革: 来自区块链技术的支持与想象 [J]. 宁夏社会科学, 2019 (6) : 76-82.
- [30] 杜小勇, 陈跃国, 范举, 等. 数据整理——大数据治理的关键技术 [J]. 大数据, 2019, 5 (3) : 13-22.
- [31] RZESZOTARSKI J M, KITTUR A. Kinetica: Naturalistic multi-touch data visualization [C] //Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014: 897-906.
- [32] 匡俊塞, 赵畅, 杨柳, 等. 一种基于深度学习的异常数据清洗算法 [J]. 电子与信息学报, 2022, 44 (2) : 507-513.
- [33] 何振, 彭海艳. 人工智能背景下政府数据治理新挑战、新特征与新路径 [J]. 湘潭大学学报 (哲学社会科学版), 2021, 45 (6) : 82-88.
- [34] KUMAR A, NAUGHTON J, PATEL J M, et al. To join or not to join? Thinking twice about joins before feature selection [C] //Proceedings of the 2016 International Conference on Management of Data. 2016: 19-34.
- [35] 杨蕾. 数据安全治理研究 [M]. 北京: 知识产权出版社, 2020.
- [36] 赵正, 郭明军, 马骁, 等. 数据流通情景下数据要素治理体系及配套制度研究 [J]. 电子政务, 2022 (2) : 40-49.
- [37] 夏义堃, 管茜. 政府数据资产管理的内涵、要素框架与运行模式 [J]. 电子政务, 2022 (1) : 2-13.
- [38] SHAH S I H, PERISTERAS V, MAGNISALIS I. DaLiF: a data lifecycle framework for data-driven governments [J]. Journal of Big Data, 2021, 8 (1) : 1-44.
- [39] 周林兴, 徐承来, 宋大成. 智慧城市视域下政府数据质量优化反

- 馈机制研究[J]. 情报杂志, 2021, 40(7): 146-156.
- [40] 王正青, 但金凤. 大数据时代美国教育数据质量管理流程与保障[J]. 现代远程教育研究, 2019, 31(5): 96-103, 112.
- [41] 丰佰恒, 佟泽华, 韩春花, 等. 科研大数据质量管控模型仿真研究[J]. 情报理论与实践, 2021, 44(9): 33-42.
- [42] 李青, 韩俊红. 数据治理: 提升教育数据质量的方法和途径[J]. 中国远程教育, 2018(8): 45-53, 80.
- [43] 付登坡, 江敏, 任寅姿, 等. 数据中台: 让数据用起来[M]. 北京: 机械工业出版社, 2019.

作者简介

林伟, 男, 1983年生, 博士研究生, 副教授, 研究方向: 信息化侦查、机器学习, E-mail: 190012898@qq.com。
周耀铭, 男, 2001年生, 研究方向: 数据法学、犯罪治理。

Review of Data Governance Research at Home and Abroad

LIN Wei^{1,2} ZHOU YaoMing¹

(1. Fujian Police College, Fuzhou 350007, P. R. China; 2. Southwest University of Political Science and Law, Chongqing 401120, P. R. China)

Abstract: With the development of global economy and the wide application of Internet, data resources play an increasingly important role in economic operation and resource allocation. Data governance has become a key field and important way of enterprise governance and national governance, and data governance research has become a hot research field. Through analyzing related literatures, it is concluded that the research results of data governance at home and abroad mainly focus on three levels: law, technology and management. Among them, personal information rights and interests, enterprise data ownership and government data opening are the research hotspots at the legal level, data collection technology, data storage technology and data processing technology are the research hotspots at the technical level, and data quality management and data lifecycle management are the research hotspots at the management level. In the future, more in-depth research should be done on data security governance, data collaborative governance and the construction of data platform.

Keywords: Data Governance; Data Security; Rules and Regulation; Collaborative Governance; Data Center

(收稿日期: 2022-05-27)

■ 书 讯 ■

《汉语主题词表》

《汉语主题词表》自1980年问世以后, 经1991年进行自然科学版修订, 在我国图书情报界发挥了应有作用, 曾经获得国家科学技术进步二等奖。为适应网络环境下知识组织与数据处理的需要, 由中国科学技术信息研究所主持, 并联合全国图书情报界相关机构, 自2009年开始进行重新编制工作, 拟分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分逐步完成。目前工程技术卷和自然科学卷已出版。

《汉语主题词表(工程技术卷)》共收录优选词19.6万条, 非优选词16.4万条, 等同率0.84, 在体系结构、词汇术语、词间关系等方面进行了改进创新。《汉语主题词表(自然科学卷)》共收录专业术语12.4万条, 包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域, 收词系统、完整, 语义关系丰富、严谨, 每条词汇都有相应的学科分类号表现其专业属性, 并与同义英文术语对应。同时, 建立《汉语主题词表》网络服务系统, 提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表》可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘, 是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版, 分为13个分册, 总定价3 880元。

《汉语主题词表(自然科学卷)》已于2018年5月由科学技术文献出版社出版, 分为5个分册, 总定价1 247元。两卷均可分册购买。