

数字健康产业数据治理体系研究

董焕晴¹ 何树坤¹ 曹高辉^{1,2*}

- 华中师范大学信息管理学院, 湖北 武汉 430079;
- 湖北省数据治理与智能决策研究中心, 湖北 武汉 430079)

摘要: [目的/意义] 构建数字健康产业数据治理体系是推进数字健康产业数据治理活动、提升政府数字化精准产业治理与提高企业数字化能力的基础。[方法/过程] 本文首先通过分析数字健康产业数据现状和数据治理的必要性, 并借鉴 PDCA 循环理论提出了数据治理过程; 其次, 以《数据治理规范》和五要素集成论为理论基础, 从治理主体、治理客体、治理活动、治理工具、治理目标 5 个维度设计了数字健康产业数据治理体系的逻辑框架; 最后, 在逻辑框架的基础上构建了基于数据中台的数字健康产业数据治理体系技术架构。[结果/结论] 本文构建的数字健康产业数据治理体系逻辑框架和技术架构能够提升产业数据治理理论体系研究和技术的价值, 为数字健康产业数据治理实践提供参考, 推进数字健康产业的数字化治理和可持续发展, 提高政府和企业的数字治理能力。

关键词: 数字健康产业; 数据治理; 五要素集成论; 治理体系; 技术架构

DOI: 10.3969/j.issn.1008-0821.2024.09.011

[中图分类号] G256.5 [文献标识码] A [文章编号] 1008-0821 (2024) 09-0131-11

Research on Data Governance System of Digital Health Industry

Dong Huanqing¹ He Shukun¹ Cao Gaohui^{1,2*}

- School of Information Management, Central China Normal University, Wuhan 430079, China;
- Center for Data Governance and Intelligent Decision Making of Hubei Province, Wuhan 430079, China)

Abstract: [Purpose/Significance] The construction of digital health industry data governance system is the basis for promoting digital health industry data governance activities, improving the government's digital precision industry governance and improving the digital capability of enterprises. [Methods/Process] Firstly, the paper analyzed the data status of the digital health industry and the necessity of data governance, and drew on the PDCA cycle theory to put forward the data governance process. Secondly, based on the data governance specification and the five-element integration theory, the paper designed the logical framework of the data governance system for the digital health industry from the five dimensions of governance subject, governance object, governance activity, governance tool and governance target. Finally, on the basis of the logical framework, the paper constructed the technical architecture of the digital health industry data governance system based on the data center. [Result/Conclusion] The logical framework and technical framework of data governance system for digital health industry constructed in the paper can enhance the value of theoretical system research and technical practice of industrial data governance, provide reference for the data governance practice of digital health industry, promote the digital governance and sustainable development of digital health industry, and improve the data governance capabilities of governments and enterprises.

Key words: digital health industry; data governance; five elements integration theory; governance system; technical architecture

收稿日期: 2023-11-27

基金项目: 武汉市知识创新专项项目“数字健康产业大脑系统构建与关键技术研究”(项目编号: 2022010801010277); 国家社会科学基金项目“基于事理图谱的社会化问答知识组织与服务研究”(项目编号: 19BTQ075)。

作者简介: 董焕晴(1995-), 女, 博士研究生, 研究方向: 情报分析、数据治理。何树坤(1998-), 男, 硕士研究生, 研究方向: 情报分析、数据治理。

通信作者: 曹高辉(1980-), 男, 教授, 博士, 博士生导师, 研究方向: 知识组织、知识服务。

全球数字化革命的加速发展，数字技术逐渐赋能医疗健康服务，数字化、网络化、智能化等现代科技手段的应用，正在逐渐改变传统线下医疗的形态，数字健康应运而生。尤其是新冠肺炎疫情期间，人工智能、物联网、云计算等数字信息技术在医疗健康领域方面更是起到了重要支撑作用。习近平总书记强调“要高度重视新一代信息技术在医药卫生领域的应用，重塑医药卫生管理和服务模式，优化资源配置、提升服务效率”。随着数字健康作为一种新型服务模式的不断发展，它已经成为医疗健康领域的新生产力和新发展方向。同时，数字健康产业也进入了高速增长阶段，并逐渐成为数字经济发展的新动能。在我国积极推进“数字中国”和“健康中国”的背景下，数字健康产业正处于数字化转型的关键时期。但是目前我国的数字健康产业市场发展尚不成熟，涉及行业领域较多，政府精准管理存在困难，存在着产业发展方向不明晰、产业数据安全系数低、行业竞争加剧等问题。

随着产业数字化改革的持续进行，数字健康产业中的数据也正在以指数级速度增长，并且数据结构也愈来愈多元化。但是这些海量、多元的数据大多是存储与管理在不同的组织和部门之中，面对跨部门、跨区域、跨层级的数据，出现了数据孤岛、数据共享困难、数据融合困难、数据安全与隐私保护困难、数据赋能率低等问题。为了促进政府数字化产业治理、提升数字健康产业的协同创新、优化产业资源配置和效率，以及提升数字健康产业数据价值，数据治理成为了数字健康产业高质量发展与其数字化转型升级进程中迫切需要解决的问题。因此，本文拟通过分析目前数字健康产业数据现状和数据治理的必要性，规划数字健康产业数据治理过程，提出数字健康产业治理体系逻辑框架以及技术架构，以期为未来数字健康产业数据治理实践提供理论与技术参考。

1 相关研究

1.1 数字健康相关研究

随着数字健康理念的兴起，数字健康相关研究也逐渐成为了一个备受学者们关注的领域。目前，国内外对于数字健康的相关研究主要围绕理论探讨与实践应用两个方面。在理论探讨方面，其主要围绕数字健康的定义、发展阶段等方面，欧盟^[1]将数

字健康定义为利用现代通信技术，满足普通公民、病人、医务工作者和医疗决策者的需要的一系列活 动。国内学者黄如意等^[2]提出了数字健康的含义有狭义与广义之分。从狭义上讲，数字健康是指通过数字化、网络化和智能化技术为医疗卫生服务赋能的过程。从广义上讲，数字健康是将数字化技术和卫生管理相结合。秦建友^[3]基于数字经济与健康产业发展现状探讨数字经济对健康产业发展的重要意义。在数字健康实践应用方面，Maier E 等^[4]认为，当前数字健康实践应用成果多属于数字医疗服务，主要包括电子医疗服务、移动医疗服务和远程医疗服务等。Zheng Y L 等^[5]认为，目前存在的数字健康应用程序可以与腕带传感器、心率传感器、皮带传感器、鞋子传感器等可穿戴设备进行无线同步，从而获取更精准的健康信息数据，为个人健康管理和医疗决策提供支持。Cross S P 等^[6]认为，数字心理健康技术与面对临床护理的整合使用有潜力提高青少年的护理质量，提高了其心理健康服务质量。

1.2 数据治理相关研究

数据治理是以数据为对象，在保证数据安全的基础上，构建完善的规则体系，明确各参与方在数据流通过程中的权责关系，形成多参与方共享、共治的流通模式，实现数据价值的最大化^[7]。大数据时代背景下，国内外学者相继对数据治理展开研究，研究主题主要分布在概念界定、逻辑框架研究、治理工具探究等方面。

数据治理起源于企业管理、IT 管理，近年来已逐渐蔓延到政府、科研院所、产业等较多的机构和应用场景中，并出现了政府数据治理、科研数据治理、产业数据治理等概念。在各类研究中，数据治理概念定义也较为多样化，暂未有统一、明确的定义。夏义堃^[8]从数据的发展过程出发，把数据治理视为信息管理中不可或缺的组成部分。樊振佳^[9]认为，数据治理是为了提高数据质量而采取的一系列管理控制活动的集合，包括对数据资源实施计划、监督和执行等。根据国际数据管理协会^[10] (International Data Management Association, DAMA) 的定义，数据治理是指对数据资源进行规划、监控和执行等方面的权限与控制。数据治理协会^[11] (The Data Governance Institute, DGI) 认为数据治理是一个系统，通过一系列信息相关的流程来实现决策权

和职责分工。Seiner R S^[12]认为,数据治理是对数据以及相关资产进行正式管理和权力实施的过程。

在数据治理框架研究方面,目前较为系统化、具有指导性的框架主要是现存的一些行业标准,如DAMA框架和DGI框架。我国数据治理相关国家标准《数据治理规范》(GB/T 34960.5-2018)和企业数据治理系统标准也参考了DAMA框架、DGI框架、ISO(国际标准化组织)国际标准框架。目前,各机构对于数据治理的定义和数据治理实施办法也各有不同,但其宗旨都是实现数据统一管理、提高数据质量、挖掘数据的潜在价值。其中,DAMA框架主要介绍了数据治理的功能与环境要素,并阐述了功能与环境要素之间的对应关系。而DGI架构则将数据治理的十大要素之间的逻辑联系以一种“路径”的形式呈现出来,从而构成了一个从方法论到实现的完整体系^[13]。ISO/IEC 38505-1将数据治理定义为IT治理的一个子集或领域,而IT治理本身是组织治理或企业治理的一个子集或领域^[14]。

目前,数据治理工具主要分为技术手段工具和行政手段工具,在技术手段工具方面^[15-16],主要是关注数据的集成与处理能力,如数据标准体系的构建、数据资产管理、数据处理技术、数据的算法模型与可视化工具、云计算平台、大数据管理平台等。行政手段方面则侧重于政策法规体系的建设、数据的审计、数据价值评估、数据资产定价等,通过业务驱动,建立数据的流动,进而达到组织的协同。

综上所述,目前国内外对于数字健康相关的研究主要在理论探讨与实践应用方面,对于数据健康产业数据相关的研究较少。对于数据治理的研究主题主要分布在概念界定、逻辑框架研究、治理工具探究、治理实践和经验剖析等方面。为了解决数据治理问题,学术界虽已提出了一些数据治理框架,但目前现有的框架标准均侧重于解决政府与企业的传统数据问题,对于涵盖多个行业、企业方面的产业数据治理涉及较少,在数字健康产业方面的数据治理更是存在空白,尚未形成统一完善的体系,并且在技术实现上也缺乏相应的理论依据。基于此,本文拟通过阐述数字健康产业数据现状和数据治理的必要性,提出产业数据治理流程,基于相关理论构建数据治理体系的逻辑框架,并在此基础上提出数字健康产业数据治理体系的技术架构,以期在未

来数字健康产业数据治理实践提供研究路线与借鉴。

2 数字健康产业数据现状

2.1 数字健康产业数据来源

世界卫生组织^[17]认为,数字健康是指在大数据、云计算、人工智能等新兴技术的引领下,对医疗市场以及健康服务业务供给产生重大影响的新兴业务模式、新技术应用、新产品服务、新监管方式等,是医疗卫生与人们的日常生活和社会活动紧密结合的结果。健康产业是指建立在医疗保健和生物技术、生命科学的基础上,旨在维护、改善人们的身体健康状况,向社会大众提供与健康有直接关系或者密切相关的产品(商品和服务)的一种生产活动集合^[2]。

因此,结合相关学者对于数字健康以及健康产业的定义,本文认为数字健康产业是指在大数据、云计算、人工智能等新型信息技术的引领下,以生物技术和生命科学为先导,涵盖数字医疗、数字医药、医药电商、数字医保、数字医养等健康服务功能的健康产业,旨在通过数字化技术改善健康管理、提高健康服务质量、提升健康资源共享效率,从而改善人们的健康水平。

数字健康产业链节点囊括了数字医疗、医药电商、数字医检、数字健保、数字康养、器械研发、医药研发、信息化厂商、医疗设备经营商、药店、医院、医生学术培训等行业。数字健康产业数据的来源则是涵盖了整个产业的上、中、下游中的各个企业和组织。具体从数据内容来看,数字健康产业上游数据内容主要包含了医疗设备制造商、器械研发机构提供的医疗设备数据和生产数据等可公开数据,以及医药研发机构提供的临床试验数据、药物相互作用数据等可公开数据;数字健康产业中游数据内容主要包含了医疗信息系统、数字医疗服务商提供的电子病历、医疗影像、就诊记录等部分可公开数据,以及医药电商提供的消费者健康产品购买记录、用户健康档案等可公开数据。数字健康产业下游数据内容主要为医院、药店等医疗健康服务机构提供的患者就诊、用药记录等可公开的医疗服务数据以及消费者提供的个人健康档案、健康管理数据、健康需求数据等可公开数据。

2.2 数字健康产业数据特征

随着技术的进步和政策的推动,数字健康产业

不断发展和完善,也更好地服务于社会大众的健康需求。伴随产生的则是海量的数字健康产业数据,依据数字健康行业的特性,数字健康产业的数据特征可以从数据体量庞大且不断增长、数据的多源性和异构性、数据的敏感性和保密性、数据的流动性4个维度进行阐述。这些数据特征共同定义了数字健康产业的复杂性和潜力,为提高数据治理、医疗质量、促进健康管理提供了强有力的支持。

1) 数据体量庞大且不断增长

随着智能设备和健康监测工具的普及,数字健康产业数据呈现指数级增长。例如,在健康大数据分析场景中,通过处理海量的个人健康数据,可以发现疾病发生的模式和趋势,为公共卫生决策提供科学依据。

2) 数据的多源性和异构性

数字健康产业领域涉及的数据类型广泛,数据来自各种来源,包括医疗机构、设备制造商、器械研发商、医药研发商和患者等。数据格式和内容可能有所不同,需要整合和标准化。例如,在远程医疗场景中,医生需要综合患者的医疗历史、实时健康监测数据和生活习惯等多源数据,以做出更全面的诊断。

3) 数据的敏感性和保密性

由于健康涉及个人隐私,因此数据安全和隐私保护是行业的关键要求。例如,在健康信息管理系统中,必须实施严格的数据加密和安全措施,确保数据安全和隐私安全。

4) 数据的流动性

现代数字健康数据需要实现跨平台、跨系统的流动与共享。例如,在电子健康记录的应用场景中,通过电子病历系统的互操作性,可以实现患者信息的无缝转移。

2.3 数字健康产业数据治理实践现状

在政策立法实践方面,2016年,国务院办公厅发布了《关于促进和规范健康医疗大数据应用发展的指导意见》,其中明确指出健康医疗大数据是我国重要的基础性战略资源,应依据新兴信息技术的发展趋势,规范推动健康医疗大数据融合共享、开放应用。2018年,国务院办公厅发布《关于促进“互联网+医疗健康”发展的意见》中提到,要加快建设基础资源信息数据库,完善全员人口、电子健

康档案、电子病历等数据库,推动建设统一权威、互联互通的全民健康信息平台,健全全国医疗健康数据资源的统一规范目录和标准体系。2018年,国家卫生健康委发布的《国家健康医疗大数据标准安全和服务管理办法(试行)》中对数据的标准管理、安全管理、服务管理、管理监督等提出了详细的管控办法。除了健康医疗大数据相关的专门性行政法规外,我国在此期间也颁布了《中华人民共和国网络安全法》《中华人民共和国数据安全法》以及《中华人民共和国个人信息保护法》3部法律,其中包含了对于数字健康产业数据的分类分级、安全合规评估、个人健康医疗数据保护等制度,为医疗健康行业提供了配套规则体系,初步构成了我国数据治理的基础性法律体系。

在数字健康产业数据治理实例方面,CHIMA发布的《2021—2022年度中国医院信息化状况调查报告》中提到,医院的电子病历数据库逐渐形成,逐渐重视医院信息互联互通标准化成熟度测评,医院的信息化建设投入金额也普遍增加。魏玖长等^[18]从个性化健康管理服务、医疗卫生服务、公共卫生服务、医药服务方面描述了健康医疗大数据的治理如何赋能大健康产业升级。张振等^[19]从健康医疗大数据治理的内容、安全、质量、共享等方面提出了相应的对策和建议。

综上所述,目前我国虽然提供了相关的技术文件、部门规范性文件指引数字健康产业数据治理,但是数据立法的整体框架的呈现比较模糊和抽象,产业数据权属、评估、共享、权利保护以及交易规则等制度供给也不够全面、均衡、有效。在数字健康产业数据治理实例方面,也较多集中在健康医疗大数据治理方面,对于整个数字健康产业数据治理的实践较为匮乏。

数字健康产业是典型的数据集中、以数据为核心资产和业务的大数据行业,随着我国产业数字化的不断转型升级,数字健康产业数据资源建设已经发展到了一定阶段,数据积累也达到了相当规模,产业数据具有体量大、结构散等特点。因此,数字健康产业亟需打造一个新型、有效的产业数据治理体系,通过整合与管理产业链上下游的资源与数据,推动数字健康产业的可持续发展,充分发挥产业数据的强大力量,实现数字健康产业主体的协同发展。

3 数字健康产业数据治理的必要性与治理过程

3.1 数字健康产业数据治理的必要性

随着信息技术的飞速发展和人们对健康的日益关注,数字健康产业迅速兴起。在这一背景下,产生了大量的数字健康产业数据,这些数据不仅包括医疗记录、患者信息等传统医疗数据,还包括基因组学数据、健康管理数据、移动健康应用数据等新兴数据类型。如何促进对这类数据的有序、规范、安全和高效地使用,保证数据在数源单位、平台中心和和使用单位之间的高效、优质的流通,是充分发挥数据要素的作用,持续推进数字健康产业数字化发展的关键。本文将从数据安全、数据质量、数据共享、数据隐私保护和合规性、数据互操作性和集成需求5个方面对数字健康产业数据治理的必要性进行详细分析。

1) 数据安全需求

从数字健康产业数据来源分析,可知数字健康产业数据的来源主体多,数据多源且繁杂,其数据主权、数据边界界定较为困难,但数据的挖掘价值较大^[20];从数字健康产业数据内容分析,数字健康产业数据中涵盖生产、临床试验以及大量敏感的个人健康数据,数字健康产业数据具有更高的隐私性与保密性^[19]。因此,数字健康产业数据的安全治理显得尤为重要。在数据治理过程中,一方面要对数据从数据来源到数据应用的全生命周期进行管理,保障数据的可用性;另一方面需要全面对数据从来源到应用和共享等各个环节进行数据安全管控,从管理、技术等多个方面实行分级分类的数据安全防控策略,为数字健康相关企业提供有效的数据保护,以保障其数据权利和商业机密不受侵犯,切实保护国家数据资源安全。

2) 数据质量需求

随着互联网、人工智能等科学技术的进一步发展,数字健康逐渐成为主流,数字健康产业也逐步繁荣,对于主要依托于互联网技术的数字健康产业来说,其在发展、经营过程中产生了海量的数据资源,数据已成为数字健康企业和政府的一项重要生产要素。然而,当前数字健康产业包含主体众多,来源广且繁杂,由于录入错误、数据不一致、缺乏标准化等问题,数据质量参差不齐^[21]。因此,数字健康产业需要建立数据质量监控机制,规范数据

录入和整合流程,提升数据的准确性和一致性。

3) 数据共享需求

数字健康产业数据范围、标准等方面的不统一,使政府部门之间、政府与企业之间、企业与企业之间的数据的多向流动和融合存在较多问题,没有形成统一联动的共享机制,数据的归集、整合、清洗、比对等普遍滞后,在一定程度上阻碍了海量数据资源的共享^[2]。数字健康产业的数据治理目标之一是要克服制度、标准和技术等多方面的瓶颈,以有效推动政企间的数据共享,从而激活数据资源的价值。在数据治理过程中,结合数据共享与交换需求,健全数据的定义、范围、质量管理、安全管理等标准规范,建立健全数据分级分类制度,推进数据的标准化建设,促进数据的高效互通。因此,数字健康产业需要建立数据共享的标准和政策,推动数据共享的开放和安全性。

4) 数据隐私保护和合规性需求

数字健康产业涉及大量的个人健康数据,因此隐私保护和合规性要求是至关重要的。随着数据治理法规的不断完善,数字健康产业需要确保数据处理的合规性,保护用户数据隐私和数据所有权^[22]。因此,数字健康产业需要建立严格的数据合规性审核机制,确保数据使用符合法律法规和伦理规范,维护数据主体权益。例如,在云健康管理平台中,医疗机构需要确保患者的敏感信息不会被未经授权的人员访问。此外,根据不同的法规,医疗机构也需要遵守相应的数据保护法规。

5) 数据互操作性和集成需求

数字健康产业需要实现不同数据源之间的互操作性和集成。例如,在电子病历系统中,通过采用标准化的数据格式和接口,不同医疗机构之间可以共享患者的健康数据,从而实现更全面的医疗服务和协同工作。

3.2 数字健康产业数据治理过程

通过分析数字健康产业数据现状与数据治理的必要性可知,伴随着科技与信息技术的不断进步,数据已经变成了日益重要的生产要素和战略性资产,如何让数据创造更多的价值,发挥数据要素作用,推动数字健康产业的建设和发展,为政府产业数字化治理与企业精准决策提供数据辅助,产业数据治理已成为当前迫切需要解决的问题。数据治理是组织中涉及数据使用的一整套管理行为,数字健康产

业数据治理是指各级政府对于产业数据价值的挖掘与利用，主要是通过利用云计算、人工智能、物联网等智能技术，对数据使用的过程进行规划、指导、执行、监督和评估等，不断挖掘数据潜力，使得数据不断地重新赋能数字健康产业^[23]。PDCA(Plan-Do-Check-Action)是全面质量管理理论的思想基础和管理依据，其基本原理是通过反复地循环执行直到问题解决。从数据治理成熟度模型^[24]可知，数

据治理是一个包含管理建构和技术建构双重属性的复杂体系，数字健康产业数据治理作为数据治理的一个特殊应用场景，其自身也是一项具有双重属性的复杂任务，借鉴PDCA循环理论的方法与迭代思路对其进行持续性改进具有重要意义。因此，本文结合PDCA循环理论设计了数字健康产业数据治理过程，具体数据治理过程如图1所示。

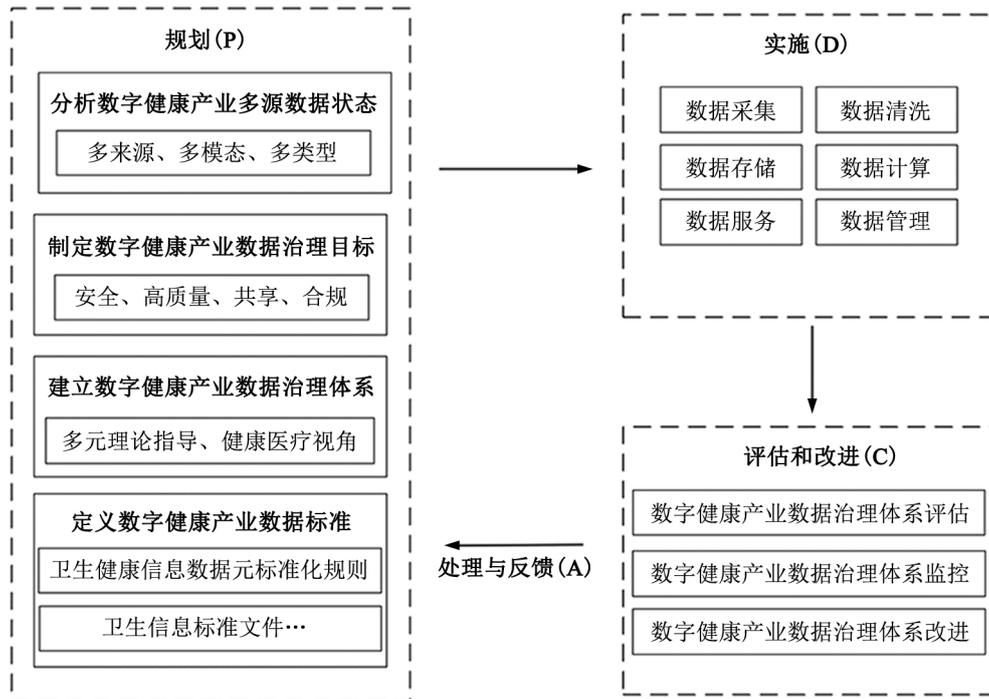


图1 数字健康产业数据治理过程

Fig. 1 Digital Health Industry Data Governance Process

规划阶段主要是针对数据治理实施工作进行的前期规划和设计，主要包括分析数字健康产业多源数据状态、制定数字健康产业数据治理目标、建立数字健康产业数据治理体系、定义数字健康产业数据标准等事项；实施阶段是数据治理的具体实践的操作阶段，其主要是在依据规划阶段的基础上进行技术层面的研发和实施，主要包含数字健康产业数据的采集、清洗、存储、计算、服务等操作；评估和改进阶段主要是通过制定相应的考核标准和制度评价体系对其数据治理效果进行评估，并在之后进行实时的监控，当数据治理效果不理想或者达不到相应的评价标准时，即可考虑对数据治理体系进行改进。

4 数字健康产业数据治理体系

4.1 数字健康产业数据治理体系逻辑框架

数字健康产业数据治理体系的逻辑框架是解决

数据治理相关问题的理论研究和 技术实践的基础，揭示了数据治理问题的分析框架和研究逻辑。自数据治理被提出以来，国外相继形成了 DAMA 框架、DGI、IBM 等框架，这些框架在开发的过程中主要是依据西方特有的场景和标准，在一定程度上并不适合我国的国情。《信息技术服务_治理第5部分：数据治理规范》(GB/T 34960.5—2018，简称《数据治理规范》)是由我国颁布的一个数据治理规范，从信息技术服务视角提供了数据治理通用框架^[25]，但是其缺乏从多维度对数据治理进行梳理。在数据治理理论基础的选择上，当前研究多聚焦于治理主体、客体、活动和目标等关乎体系构成内容的多元理论选择^[26]，目前基于五要素集成论的理论视角的方法论是构建数据治理体系逻辑框架的主流，适用于大数据的治理分析^[27-28]。因此，本文以《数据治理规范》和五要素集成论为理论基础，从治理

主体、治理客体、治理活动、治理工具、治理目标5个维度去分析和构建数字健康产业数据治理体系逻辑框架,如图2所示,其逻辑流程主要是治理主体围绕治理目标,通过对治理客体利用相应的治理

工具,开展一系列的诸如数据采集、存储、计算等治理活动,进而实现一系列的数字健康产业数据治理目标。

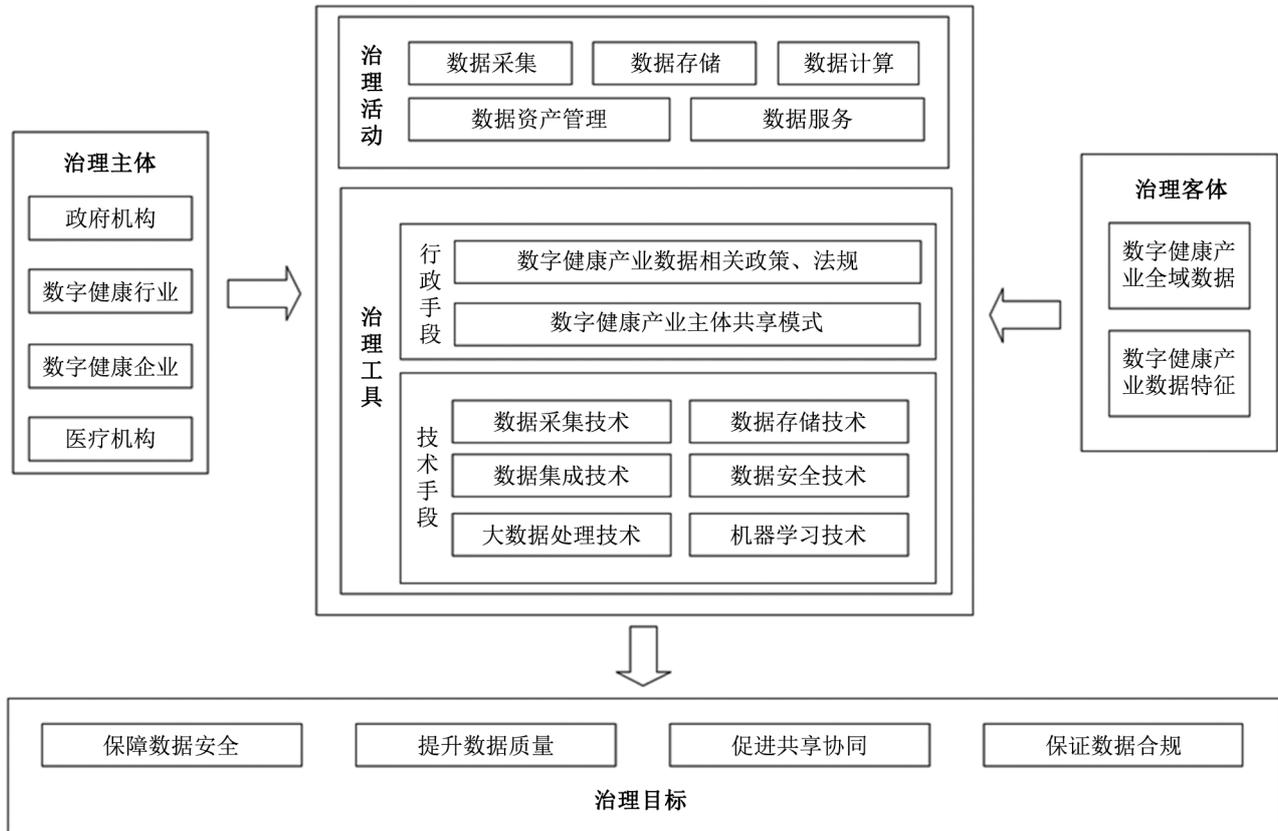


图2 数字健康产业数据治理体系逻辑框架

Fig. 2 Digital Health Industry Data Governance System Logical Framework

1) 治理主体

数字健康产业数据的治理主体由政府机构与数字健康相关行业、企业以及医疗机构等组成,它是一个以政府机构为主导的多层次的多元治理主体。政府主体主要包含数字健康产业领域的职能管理部门、数据管理机构等。数字健康行业主体是指旨在统筹、引领和推动整个数字健康产业发展的组织。这些组织通常由政府机构、行业协会、专业团体或跨国组织等建立和运营。其职责包括制定行业标准、促进技术创新、倡导政策制定、协调各方利益关系、推动行业合作与发展等。数字健康企业主体则主要包含数字医疗、数字医检、数字健保、数字康养、医药电商、器械研发、医药研发、信息化厂商、医疗设备经营商等企业。医疗机构主体则是以医院为首,具备医疗服务资质和条件,能够提供医疗诊疗、治疗、护理等医疗服务的实体或组织,其

在数字健康产业中扮演着至关重要的角色,是医疗服务提供的主体单位。

在进行数字健康产业数据治理时,要充分发挥政府主体内部之间、政府与其他治理主体之间的数据共享与协同,使得产业数据支撑起整个数字健康产业战略部署,共同营造内外共治共享的服务状态。治理主体的核心内涵即政府主导下的多元主体合作,因而在治理过程中要明确数据权属关系和各组织在治理过程中的角色与责任,精确部署机构数据治理行动,促进数字健康产业数据的协同共享。

2) 治理客体

数字健康产业数据治理客体是指在数字化健康医疗领域中产生的各类信息和统计资料,涵盖了医疗、健康管理、医学研究等方面的数据。这些数据来自数字医疗、医药电商、数字医检、数字健保、数字康养、器械研发、医药研发、信息化厂商、医

疗设备运营商、医疗机构等多个信息源，数据资源呈现出数据来源多样性、数据体量巨大、数据产生速度快等特征。从数字健康产业数据结构来说，其主要划分为结构化数据、半结构化数据以及非结构化数据。从其数据内容来说，主要包含行业统计数据、工商注册数据、政策法规数据、企业内部数据、产品数据、外部市场数据、健康医疗数据等。由于数字健康产业自身的特征等原因，其数据涉及治疗、研发、生产、管理、运维、服务等多个环节，尤其是涉及的海量健康医疗数据相较于其他类型的数据，其商业价值更高，但同时也导致了网络勒索、个人健康数据非法交易、侵犯公民隐私等数据安全问题的出现。

3) 治理活动

数字健康产业数据治理活动覆盖了数据全生命周期、数据处理和管理的关键业务节点，主要包含数据采集、数据存储、数据计算、数据资产管理、数据服务等环节，使得数字健康产业数据在相关技术的支持下服务于数字健康产业的发展以及政府和企业的战略决策中。其中，数据资产管理是数字健康产业数据治理活动中的核心环节，主要有数据标准管理、元数据管理、主数据管理、数据质量管理、数据安全生命周期等内容，通过此范围进行管理，可以对数字健康产业要利用或产生的业务数据进行准确性、及时性保障。

4) 治理工具

治理工具主要包含行政手段与技术手段，即管理性工具与技术性工具。管理性工具主要包括数据治理策略、政策、流程和共享模式等。数据治理策略是制定和规划数据治理目标、范围、优先级和方法的指导性文件，包括数据治理的愿景、使命、目标、原则和战略规划；政策是针对数据管理和使用所制定的规则和标准，涉及数据安全、隐私保护、数据共享和合规性等方面，数字健康产业由于涵盖了医院等医疗机构，因而其在一定程度上要遵循我国发布的卫生健康等政策；流程则包括数据采集、存储、处理、分析、共享和监管等环节的具体操作流程；数字健康产业主体的共享模式旨在通过共享资源、数据和服务，实现产业链上各个主体之间的协同合作，从而促进数字健康产业的发展 and 提升整

体效益，是一种政府主导下的一种行政调节手段。管理性工具的作用在于为数字健康产业数据治理提供框架和指导，确保数据管理和使用符合规范与战略目标。

技术工具主要是指从信息技术方面提升数据汇聚与处理的能力，常用的技术主要有数据采集技术、数据存储技术、数据集成技术、数据安全技术、大数据处理技术、机器学习技术等，例如将区块链技术应用与数字健康产业，在一定程度上可以很好地解决数据孤岛现象，提升产业数据要素的数量和质量。技术性工具的作用在于帮助数字健康产业实现数据安全、高质量和便捷共享，提升数据管理效率和价值。通过管理性工具与技术性工具的融合使用，来实现数字健康产业数据治理目标，进而深度挖掘数据价值。

5) 治理目标

数字健康产业数据治理是一个长期动态变化的数据治理过程，鉴于医疗健康数据资源的特殊性，其数据治理是在确保数据安全和隐私保护的前提下，促进数字医疗技术的创新和发展，优化医疗资源配置，促进跨界的合作与互联互通，强调数字化、信息化和跨界合作的特征。结合数字健康产业数据治理的必要性，可知其治理目标要围绕保障数据安全、提升数据质量、促进共享协同、保证数据合规4个方面。从数据角度来看，通过数据采集、数据集成、数据计算等一系列数据治理活动的开展以及相应的数据治理工具的使用，在一定程度上提升了数据的质量，保障了数据安全。从组织层面来看，开展数据治理活动可以降低各个机构对于数据资源的重复建设，在此基础上，可以在一定程度上减少数据的处理费用，从而达到收集、共享、协作和合规的目标。

4.2 数字健康产业数据治理体系技术架构

4.2.1 设计思路

随着大数据、人工智能技术的发展，数据需求也随之不断增大，同时数据资源在不同系统、不同部门之间的数据孤岛问题也日益突出。在此背景下，数据中台的概念被提出并用来解决政府和企业的数据治理与协同方面的问题，提升数据资源的管理效率和价值。数据中台是一种数据应用机制，它将数

据持续转化为资产，为企业提供服务，它拥有聚集整合、提取处理、可视化服务和价值转化等核心功能^[29]。尤其具有处理多种类数据、架构更具模块化和开放、数据安全和隐私保护性较高等优势。

目前数据中台已成为各大企业进行数智化转型的关键路径和数据治理方面的重要概念。Gartner^[30]在2016年的“Pace-layered Application Strategy”报告中指出，应该将企业的商业系统分为前台、中台、后端3个层级，并指出中台的核心角色是以灵活的方式对前台的应用需求做出反应。2015年，阿里巴巴首次在IT行业启动中台战略，其主要核心为构建企业核心资源和共性技术的可重复使用能力，避免重复构建，提高开发效率，争取将数据中台打造成融合多条业务系统数据的信息共享和服务发布平台^[31]。并且在数据治理的各个环节中，Hadoop大数据技术多用于治理活动过程中的标准化工具组件和功能开发^[32]。因此，基于“复用”“共享”的数据中台概念内涵，本文设计了基于数据中台的数据治理体系技术架构，为数字健康产业数据治理指明了技术路线。其基本思路为通过对全域的产业数据进行汇集，构建数据中台，然后建立可复用的数字健康产业应用服务，从而提高业务需求的响应速度和应用开发的迭代速度。

4.2.2 技术架构搭建

本文在数字健康产业数据治理体系逻辑框架的基础上，并基于数字健康产业数据治理技术架构设计思路，采用Hadoop大数据技术，搭建了数字健康产业数据治理技术架构。数字健康产业数据治理技术架构整体上包括基础设施、数据源、数据中台、应用服务四大模块建设内容，具体如图3所示。

1) 基础设施

基础设施主要是由网络、存储、计算、安全等软硬件设施构成，为数字健康产业数据中台提供资源与技术支持，保障整个技术架构的平稳运行。

2) 数据源

数据源是指数字健康产业链上、中、下游的链条组织机构中涉及的可公开和可利用的数据。其中，产业上游数据主要包括医疗设备制造商端数据、器械研发机构端数据、医药研发企业端数据等，产业中游数据主要包括医疗信息系统数据、数字医疗服务商数据、医疗电商数据，产业下游数据主要包括

医院等医疗机构端数据。

3) 数据中台

数据中台是指对数字健康产业中的数据资源进行整合和共享，构建一个数据生态系统，提供数据服务和数据产品，提高数据质量和效率。其数据治理流程分为数据采集、数据存储、数据计算、数据资产管理、数据服务5个模块。

数据采集模块是指对数字健康产业全域相关数据资源进行汇聚融合，通过爬虫、填报、对接等操作采集数字健康产业上、中、下游的数据资源，实现数据采集、转换、清洗、脱敏、加载等功能，打破数据孤岛现象，形成共享数据中心，为更高层业务提供数据支撑。其涉及的技术主要包含DataX、Sqoop、Kafka、Storm、Flink、Flume等。

数据存储模块的作用是存储各种采集到的数据，并将这些数据经过一定的处理解析成文件数据、结构化数据和半结构化数据等不同类型，然后分别存储到相应类型的数据库中。例如，文件数据通常被存储在HDFS中，半结构化数据则被存入HBase等数据库，而结构化数据则存储在MySQL等数据库中。本模块设计的数据库技术主要有HDFS、Hive、HBase、Impala、Redis、MySQL等。

数据计算模块的目标是通过相关技术对接入的数据进行清洗、挖掘、分析等处理，主要包括离线计算(MapReduce)、实时计算(SparkStreaming)、算法计算(TensorFlow)；离线计算适用于对大规模数据进行批处理分析。在数字健康产业中，可以利用离线计算技术对大量的医疗记录、生物监测等数据进行清洗和分析。实时计算适用于对数据流进行实时处理和分析，能够及时发现并处理数据中的异常情况。在数字健康产业中，可以利用SparkStreaming对医疗监测设备产生的数据流进行实时监控和分析。算法计算在数字健康产业中也扮演着重要角色，主要是利用机器学习和深度学习算法对医疗数据进行挖掘和分析。

数据资产管理模块贯穿于整个数据治理的全生命周期，其为数据采集、数据存储、数据计算、数据服务提供统一的制度规范和标准，主要包含元数据管理、主数据管理、数据标准、数据质量、数据血缘、数据生命周期、数据安全。

数据服务模块是对数据、模型和算法进行逻辑

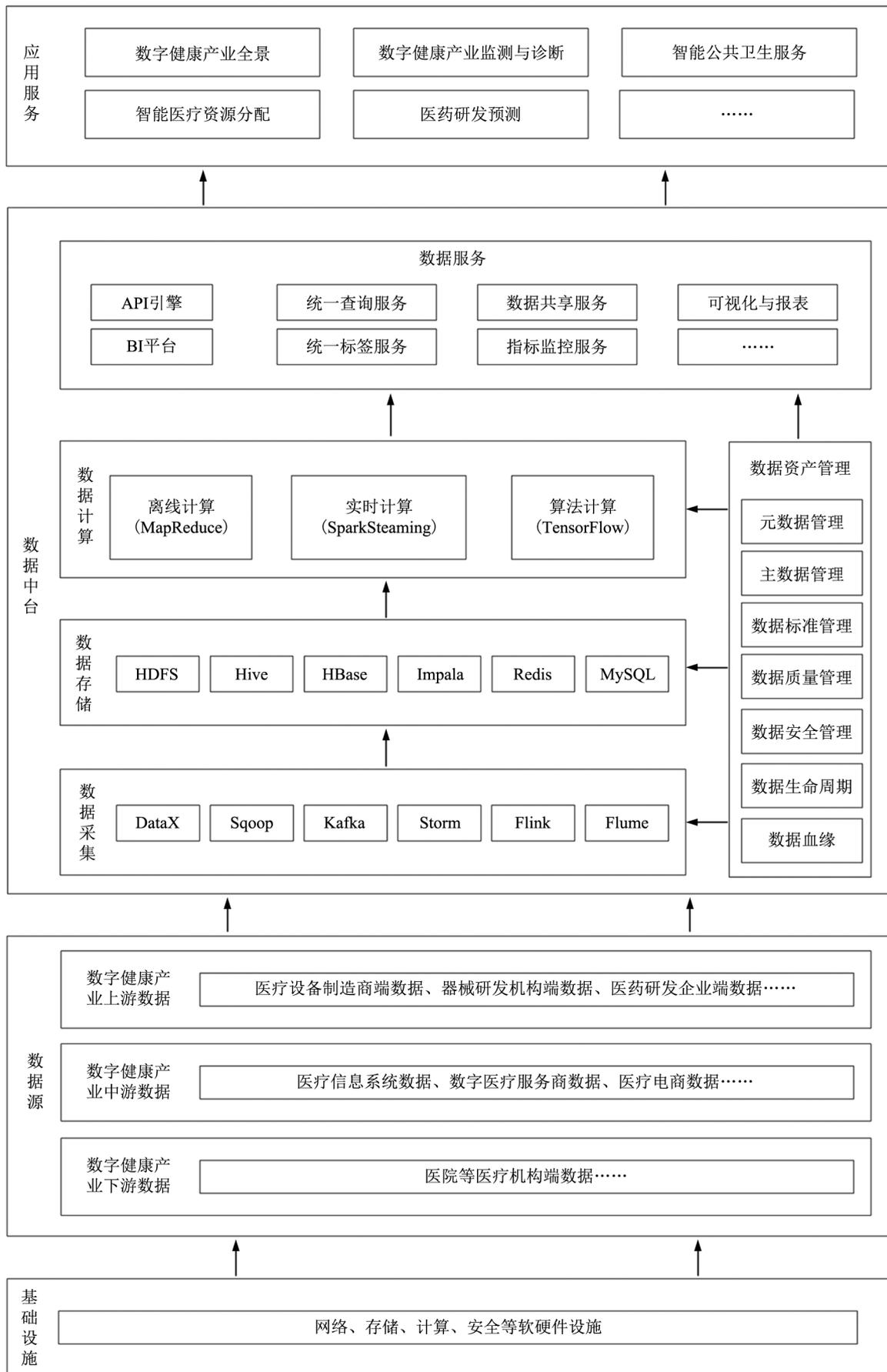


图3 数字健康产业数据治理体系技术架构

Fig. 3 Digital Health Industry Data Governance System Technical Architecture

封装,生成相应的API服务,供数字健康产业应用快速调用,数字健康产业的数据服务类型主要包含API引擎、BI平台、统一查询服务、统一标签服务、数据共享服务、指标监控服务、可视化与报表。

4) 应用服务

应用服务是依据数字健康产业中的实际业务需求而建立的,旨在发挥数字健康产业数据的效用价值,提升政府的数字化产业治理能力和企业的数字化服务质量。其涉及的应用服务主要包括数字健康产业全景、数字健康产业监测与诊断、智能公共卫生服务、智能医疗资源分配、医药研发预测等。在不同的应用服务中,不同的服务功能可以通过调用数据中台的API服务接口来实现敏捷开发和迭代。

5 结论与展望

构建数字健康产业数据治理体系逻辑框架与技术架构是部署数据治理活动、提升政府数字化精准产业治理与提高企业数字化能力的基础。本文在数字健康产业数据现状背景下,分析了数字健康产业数据治理的必要性,提出了数字健康产业数据治理过程,构建了数字健康产业数据治理体系逻辑框架和技术架构,其中得出的主要结论如下:①结合数字健康相关场景,从数字健康产业数据来源、数据特征和数据治理实践现状对数字健康产业数据现状进行了详细的分析和概括;②结合PDCA循环理论设计了包含规划、实施、评估和改进、处理与反馈的数字健康产业数据治理流程;③从治理主体、治理客体、治理活动、治理工具、治理目标5个维度出发,构建了数字健康产业数据治理体系逻辑框架,其主要逻辑为政府机构与数字健康相关行业、企业以及医疗机构围绕保障数据安全、提升数据质量、促进共享协同、保证数据合规等治理目标,采用管理性工具与技术性工具对数字健康产业中的海量多源异构数据开展一系列诸如数据采集、数据存储、数据计算、数据资产管理、数据服务等治理活动;④以Hadoop大数据技术生态和数据中台为基础,搭建了数字健康产业数据治理技术架构,其整体上包括基础设施、数据源、数据中台、应用服务四大模块建设内容。

本文丰富了数字健康产业数据治理理论体系研究和技术实践的发展,为数字健康产业数据治理提供了借鉴,从而充分挖掘产业数据价值,推进了数

字健康产业的数字化治理。此外,本文构建的数字健康产业数据治理体系逻辑框架与技术架构,主要是从相关的理论出发进行搭建,在实际应用场景中的兼容性与可扩展性仍需进一步检验。因此,后续的研究中需聚焦于数字健康产业具体应用情境开展数据治理实践,让数字健康产业数据治理体系逻辑框架与技术架构在具体的实践中不断地细化和完善,为我国数字健康产业数据治理打下坚实基础。

参 考 文 献

- [1] Quaglio G, Dario C, Stafylas P, et al. E-Health in Europe: Current Situation and Challenges Ahead [J]. Health Policy and Technology, 2016, 5 (4): 314-317.
- [2] 黄如意, 井洪. 数字化时代的数字健康: 内涵、特征、挑战与治理路径 [J]. 卫生经济研究, 2022, 39 (6): 60-63, 66.
- [3] 秦建友. 共同富裕背景下数字经济对健康产业发展的影响研究 [J]. 中国卫生经济, 2023, 42 (5): 72-76.
- [4] Maier E, Reimer U, Wickramasinghe N. Digital Healthcare Services [J]. Electronic Markets, 2021, 31 (4): 743-746.
- [5] Zheng Y L, Ding X R, Poon C C Y, et al. Unobtrusive Sensing and Wearable Devices for Health Informatics [J]. IEEE Transactions on Bio-Medical Engineering, 2014, 61 (5): 1538-1554.
- [6] Cross S P, Nicholas J, Bell I H, et al. Integrating Digital Interventions with Clinical Practice in Youth Mental Health Services [J]. Australasian Psychiatry, 2023, 31 (3): 302-305.
- [7] 梅宏. 数据治理之论 [M]. 北京: 中国人民大学出版社, 2020.
- [8] 夏义堃. 试论数据开放环境下的政府数据治理: 概念框架与主要问题 [J]. 图书情报知识, 2018, (1): 95-104.
- [9] 樊振佳. 企业科研数据治理与管护主体: NSR的实践与剖析 [J]. 图书情报工作, 2017, 61 (1): 56-63.
- [10] DAMA International. The DAMA Guide to the Data Management Body of Knowledge [M]. New York: Technics Publications, 2009: 37.
- [11] The Data Governance Institute. Defining Data Governance [EB/OL]. <https://datagovernance.com/definingdata-governance/>, 2024-03-28.
- [12] Seiner R S. Non-invasive Data Governance: The Path of Least Resistance and Greatest Success [M]. Technics Publications, 2014.
- [13] 蒋翼洋. 招商银行数据治理研究 [D]. 上海: 上海财经大学, 2020.
- [14] ISO. ISO/IEC 38505-1: 2017 [EB/OL]. <https://www.iso.org/standard/56639.html>, 2024-03-28.
- [15] 马奔, 毛庆锋. 大数据在应急管理中的应用 [J]. 中国行政管理, 2015, (3): 136-141, 151.
- [16] 何振, 周芳检, 杨文. 大数据时代城市应急管理行业协作体制机制创新研究 [J]. 湘潭大学学报 (哲学社会科学版), 2016, 40 (6): 26-31.

(下转第153页)

