

医学科学数据共享的理论逻辑与实践探索^{*}

杨琰平¹ 刘泽坤¹ 杨俊涛² 李 青¹

(¹ 中国医学科学院/北京协和医学院卫生健康管理政策学院 北京 100730

² 中国医学科学院/北京协和医学院基础医学研究所 北京 100005)

〔摘要〕 目的/意义 探讨医学科学数据共享的理论逻辑,分析当前实践中面临的挑战,为促进医学科学数据有效共享提供政策建议。方法/过程 首先构建医学科学数据共享的理论逻辑框架,分析其概念、特征、动因、发展规律和价值意义。其次分析我国医学科学数据共享的实践现状。最后从数据孤岛、隐私保护、责任归属和产权认定等角度分析当前医学科学数据共享所面临的挑战。结果/结论 针对性地提出建设统一数据共享平台、制定规范共享政策、构建合理激励机制和加强共享人员培训等建议,以推动医学科学数据有效共享。

〔关键词〕 医学科学数据;数据共享;数据孤岛

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2025.06.006

Theoretical Logic and Practical Exploration of Medical Scientific Data Sharing

YANG Longping¹, LIU Zekun¹, YANG Juntao², LI Qing¹

¹School of Healthy Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China; ²Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China

〔Abstract〕 **Purpose/Significance** To discuss the theoretical framework of medical scientific data sharing, and to analyze the current challenges faced in practice, so as to provide policy recommendations for promoting the effective sharing of medical scientific data. **Method/Process** Firstly, the theoretical framework of medical scientific data sharing is constructed, and the concept, characteristic, motivation, development principle, and value significance are analyzed. Secondly, the current practice status of medical scientific data sharing in china is analyzed. Finally, the challenges currently faced by the practice of medical scientific data sharing are analyzed from the perspectives of data silos, privacy protection, responsibility attribution, and property rights identification. **Result/Conclusion** Some targeted policy recommendations are proposed, including the establishment of a unified data-sharing platform, the formulation of standardized sharing policies, the construction of a reasonable incentive mechanism, and the strengthening of training for the personnel who work for data sharing, to promote the effective sharing of medical scientific data.

〔Keywords〕 medical scientific data; data sharing; data silo

〔修回日期〕 2024-11-21

〔作者简介〕 杨琰平,硕士研究生,发表论文2篇;通信作者:李青。

〔基金项目〕 中国医学科学院医学与健康科技创新工程(项目编号:2021-I2M-1-046)。

1 引言

随着人工智能和大数据技术的发展，医学研究逐渐步入大数据时代。截至目前，全球医疗数据规模约为数百 EB^[1]。但是海量数据分散在各机构和实验室中，利用效率低下，形成数据孤岛现象^[2]。医学科学数据共享对打破数据孤岛、提升疾病诊疗效率、推动医学进步至关重要^[3]。当前医学科学数据共享正向跨学科、跨机构和跨区域发展，如肿瘤数据共享等倡议已在全球范围内得到响应^[4]。多个国家和地区已制定相关法律法规和技术标准，以促进医学科学数据的开放和共享，例如欧盟《通用数据保护条例》^[5]为个人健康数据共享提供法律依据，明确了数据公权优先于数据私权。目前已有相关研究探索医学科学数据共享路径，但尚缺乏对医学科学数据共享理论逻辑的深入分析和对我国医学科学数据共享实践探索的系统梳理。因此，本研究旨在探讨医学科学数据共享的理论逻辑，回顾我国医学科学数据共享的实践探索历程，分析当前面临的挑战，以期为推动医学科学数据有效共享提供理论支持和实践指导。

2 医学科学数据共享的理论逻辑

医学科学数据共享以提升医学科学研究和临床实践效率、促进健康医疗数据建设和科学数据再利用为目标，具有重要的应用价值和战略意义。本研究从横向和纵向两个维度构建医学科学数据共享的理论框架。横向维度包括概念界定、内涵特征和跨学科比较，旨在从宏观层面厘清其核心概念、关键属性，为理论体系构建奠定基础。纵向维度深入探讨医学科学数据共享的动因、发展规律和价值意义，有助于揭示医学科学数据共享的内在机制和发展路径。双维度分析框架有助于系统阐释医学科学数据共享的理论逻辑，并为后续政策制定与实践探索提供科学依据。

2.1 横向维度

2.1.1 医学科学数据共享的概念定义 梳理科学数据和医学科学数据的概念界定，见表 1。医学科学数据共享是指在遵循伦理原则和法律要求的前提下，研究人员相互提供、获取医学科学数据的过程。该过程旨在促进医学科学研究进步、提升临床实践水平、加速药物研发和临床试验进程、促进健康管理和公共卫生发展。

表 1 科学数据、医学科学数据的释义及限定范围

数据类别	释义	限定范围
科学数据	在自然科学等领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据 ^[6]	通用科学领域所有用于研究活动的的数据
医学科学数据	在医学研究及其与多学科交叉融合的过程中，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动或解决医学科学问题的原始数据及其衍生数据 ^[7]	专指产生于医学研究的数据，或者非研究产生但被用于健康相关研究或解决健康问题的数据

2.1.2 医学科学数据共享的内涵特征 医学科学数据具有来源广泛、种类繁多、结构复杂的特征。既包括电子病历、影像数据等临床研究数据，也包括基因组学或代谢组学数据集等实验人群队列数据。本研究依据数据来源，将医学科学数据进行 3 级分类，见表 2。在海量医学数据中，大量数据存储于纸质媒介，超过 80% 的数据为非结构化数

据^[1]。因此医学科学数据共享具有多样性和以非结构化数据为主的特征。

表 2 医学科学数据 3 级分类

一级分类	二级分类	三级分类
临床研究数据	研究来源数据	临床研究数据、临床试验数据等
	业务来源数据	真实世界数据、公共卫生监测数据等

续表 2

一级分类	二级分类	三级分类
基础研究数据	分子生物学数据	基因组学数据、蛋白组学数据等
	细胞生物学数据	细胞培养数据、细胞增殖数据等
	病理学数据	分子病理学数据、病理切片等
公共卫生数据	疾病监测数据	传染病报告、慢性病登记等
	健康调查数据	营养调查、健康行为调查等
	环境健康数据	空气质量、水质、土壤污染等
药学数据	药物研发数据	药物设计、药物合成等
	药物使用数据	副作用报告等

2.1.3 跨学科比较下的医学科学数据共享 相较于其他科学领域，医学科学数据因其与个人健康、生活以及遗传物质等高度敏感信息紧密相关而具有显著独特性，其数据共享除了要遵循科学数据共享的一般原则，还必须遵守相关法律法规，执行严格的伦理审查。因此医学科学数据共享面临更多的伦理和隐私保护挑战。

2.2 纵向维度

2.2.1 医学科学数据共享的动因探析 医学科学数据共享的动因源自医学研究和临床实践对数据高效利用的需求。随着医学的迅猛发展和数字化医疗的普及，医学领域产生了海量数据资源。根据有关测算，未来 5 年医学领域数据的年复合增长率为 36%^[1]。这些数据涉及临床、生物信息、健康监测等，为医学研究和临床实践提供了丰富的素材，同时也带来了存储、管理和分析方面的挑战，亟需更为高效、灵活的数据存储和管理机制。

2.2.2 医学科学数据共享的发展规律 医学研究的核心任务是揭示健康与疾病的复杂机制，其研究方法与其他科学领域存在显著差异。出于伦理和实际操作的限制，医学研究往往依赖观察性研究而非干预性手段。这种研究方法要求研究人员从大样本

中收集大量数据，以确保结论的科学性。因此，医学研究天然依赖大量数据。2007 年吉姆·格雷（Jim Gray）提出数据密集型科研第四范式概念^[8]，指出科学研究正从数据贫乏时代进入数据充裕时代，数据将成为科研的新驱动力。在医学领域，随着高通量测序、大规模临床试验、医学影像技术等快速发展，医学科学数据呈现指数级增长态势，医学研究正迈入以数据为核心驱动的新阶段。

2.2.3 医学科学数据共享的价值意义 医学科学数据共享的价值和意义在于其能够显著提升医学研究的质量和效率。通过整合分散在不同机构和实验室中的碎片化数据资源，能够加强研究团队之间的协作，提高数据利用效率，促进知识的积累和创新。这种整合不仅有助于避免重复研究，节省成本，还可加速新疗法和新药的研发进程，对于提高疾病治疗效率和改善患者预后具有重大意义。

3 我国医学科学数据共享的实践探索

3.1 制度背景和政策支撑

我国高度重视医学科学数据共享工作，自 2013 年起陆续围绕科学数据平台建设、技术标准、管理规范出台一系列政策、规定，见表 3。特别是 2018 年实施的《科学数据管理办法》，从国家层面对科学数据的汇交、保存、共享与使用等进行统筹规范，明确由政府财政资金资助的各级科技计划所产生的科学数据应汇交至相应的科学数据中心。总的来看，我国科学数据相关政策逐渐完善，呈现出专业化、系统化、国际化趋势，但医学领域的数据政策距离完善和落实仍有进一步优化空间。应继续加强医学领域数据政策的研究和制定工作，不断完善相关政策体系，进一步提升医学科学数据的管理水平和治理效能，着力打造数据大国、数据强国。

表 3 我国促进医学科学数据共享的制度框架

序号	发布时间	制定/发布方	名称
1	2013 年	工业和信息化部	《关于数据中心建设布局的指导意见》
2	2016 年	国务院办公厅	《关于促进和规范健康医疗大数据应用发展的指导意见》
3	2016 年	原国家卫生和计划生育委员会	《涉及人的生物医学研究伦理审查办法》

续表 3

序号	发布时间	制定/发布方	名称
4	2018 年	科学技术部、财政部	《国家科技资源共享服务平台管理办法》
5	2018 年	国务院办公厅	《科学数据管理办法》
6	2018 年	国家卫生健康委员会	《国家健康医疗大数据标准、安全和服务管理办法（试行）》
7	2019 年	中国科学院	《中国科学院科学数据管理与开放共享办法（试行）》
8	2019 年	国务院	《中华人民共和国人类遗传资源管理条例》
9	2019 年	国家科技基础条件平台中心	《关于组织编制国家科技资源共享服务平台五年建设运行实施方案的函》
10	2020 年	全国人民代表大会	《中华人民共和国生物安全法》
11	2021 年	全国人民代表大会	《中华人民共和国数据安全法》
12	2021 年	科学技术部	《科技计划形成的科学数据汇交 技术与管理规范》
13	2023 年	国家卫生健康委员会、教育部、科学技术部等	《涉及人的生命科学和医学研究伦理审查办法》
14	2024 年	国家数据局等 17 部门	《“数据要素×”三年行动计划（2024—2026 年）》
15	2024 年	国家互联网信息办公室	《促进和规范数据跨境流动规定》

3.2 我国医学科学数据共享建设历程

3.2.1 医学科学数据共享平台建设情况 2004 年科学技术部牵头启动科学数据共享工程，陆续推动建设了多个领域的科学数据共享平台^[9]。经过多年的平台建设和数据汇交工作实践，已依托国家科技基础条件平台建成 20 个国家科学数据中心。其中在生物医学领域已建成国家人口健康科学数据中心和国家基因组科学数据中心两个国家级科学数据中心。国家人口健康科学数据中心作为医学科学数据共享的核心枢纽，承担着整合和管理全国医学科学数据的重要任务。该中心不仅为数据的集中存储和管理提供了基础设施，而且通过制定统一数据标准和共享协议，提升了数据的标准化和互操作性。这种开放式的医学科学数据中心平台为医学科研合作提供了新的交流模式，推动了数据资源的高效利用，极大地促进了医学研究的创新与合作。

3.2.2 医疗机构数据资源共享情况 根据国家卫生健康委员会数据^[10]，我国已基本建成国家全民健康信息平台，其中省级全民健康信息平台不断完善，基本已实现国家、省、市、县 4 级全覆盖；已有超过 8 000 家二级以上公立医院接入区域全民健康信息平台，20 个省份有超过 80% 的三级医院接入省级全民健康信息平台；25 个省份启动省内电子健康档案共享调阅工作，204 个地级市实现检查检验结果互通共享。医

疗数据共享不仅减少了重复检查、降低了医疗成本，也提升了患者就医的便利性。

3.3 现阶段面临的挑战

3.3.1 数据孤岛现象 数据孤岛是指因技术、管理或政策限制，不同数据库间的数据集无法进行有效整合或共享的现象^[2]。这些孤立的数据集可能包含宝贵的医学信息，但由于缺乏跨系统的数据交流和协作而不能被充分利用。这种现象阻碍了对医学信息的充分挖掘与利用，限制了研究者对现有医学数据的深入分析和二次开发，削弱了医学研究的创新性并制约了个性化医疗的有效实施。

3.3.2 个人健康隐私保护与数据共享存在矛盾 《信息安全技术 个人信息安全规范》(GB/T 35273—2020)^[11]将健康生理信息归为个人敏感信息。《科学数据管理办法》第 25 条规定“涉及国家秘密、国家安全、社会公共利益、商业秘密和个人隐私的科学数据，不得对外开放共享；确需对外开放的，要对利用目的、用户资质、保密条件等进行审查，并严格控制知悉范围”。为保护个人参与者的隐私，共享医学科学数据时必须进行去隐私化处理，删除姓名、出生日期、性别、种族等^[12]。当个人健康信息遭受侵害时，现有法律保护机制常存在救济路径不清、维权困难的问题^[13]。一方面是因为现行法律法规滞后于科技发展，另一方面是由于隐私保护的

利益相关者众多，责任归属判定难度较高。因此迫切需要明确界定各方在隐私保护中的责任，确保相应参与方切实承担保护个人健康信息的责任^[14]。

3.3.3 医学科学数据共享动力不足 当前医学科学数据共享较多依赖行政命令和期刊要求，缺乏内生动力，主要原因在于激励机制不完善。现有激励措施主要包括知识产权保护和提升学术影响力等方面。然而医学科学数据的知识产权问题极为复杂，不仅涉及著作权、专利权、商业秘密及科研信息，还要考虑个人隐私、社会公益和数据安全等伦理问题^[15]，其产权认定尚未形成专家共识，有待进一步探讨和明确。

3.3.4 利益相关者众多引发的诉求冲突 医学科学数据共享涉及多方利益主体，包括数据产生方、数据持有方、平台建设方、数据使用方以及监管治理方等，给治理体系和政策制定带来考验^[16]。亟待梳理分析各利益相关者的诉求，力争实现多方共赢。

4 推动医学科学数据共享的建议

4.1 建设统一医学科学数据共享平台

科学数据共享平台是支撑数据共享的核心基础设施，其在数据存储与传输共享过程中发挥着关键作用。建立标准化、功能全面、操作简易且用户友好的共享平台对于数据整合与共享至关重要。建议汇集卫生健康部门、医药企业、科研院所、医学院校、医疗机构以及科研人员力量，共同建设标准化、互联互通、易于访问的医学科学数据共享平台。

4.2 进一步完善法律法规与政策体系

一是明确数据权属与权益分配规则。明确界定数据采集、汇交、存储、使用、衍生等环节的持有者、使用权、管理权和收益权等，特别是要厘清原始数据提供方、数据持有方、平台建设方和数据使用方之间的权利义务边界。二是建立强制性与引导性相结合的汇交政策。对于接受国家财政资金资助的科研项目，应强制要求其在规定时限内汇交符合

标准的医学科学数据至指定或认可的数据中心。对于非财政资助项目和临床常规诊疗活动中产生的高价值数据，应通过政策引导或其他适当激励方式，鼓励自愿汇交。

4.3 构建合理的激励机制

首先，将医学科学数据视为资产，建立完善的价值评估和交易体系，促进数据的流通和利用。其次，确立严格的医学科学数据产权保护制度，发挥知识产权的规范与激励作用。最后，针对不同利益主体实施差异化激励策略。对于数据持有方，确保其产权得到保护，同时获得合理回报；对于数据使用方，鼓励其使用公开共享数据，同时配合完成数据用途调查、伦理审查等工作；对于平台建设方，激励其高效管理，确保平台顺畅运作。

4.4 加强医学科学数据共享人才培养和生态建设

一是加强数据处理人员专业培训，设立医学信息学、生物统计学、数据科学、法律伦理等相关专业或培训项目，培养既懂医学又懂数据管理、分析、安全、伦理的复合型人才，为医学科学数据共享提供人才支撑，确保其在数据收集与处理过程中既有效保护个人健康隐私，又能提升数据标准化和质量。二是面向广大医学科研人员，加强数据共享理念、政策、平台资源和技术方法的宣传、教育和培训，提升其数据共享意识和责任感，营造开放、合作、规范的共享文化。

5 结语

本研究通过构建医学科学数据共享的理论框架，从横向与纵向维度全面探讨其理论基础和实践应用机制。在此基础上系统梳理了我国医学科学数据共享的政策演进、平台建设现状及医疗机构间的数据共享进展，揭示了当前实践中面临的挑战，如数据孤岛现象、隐私保护问题及责任归属不明确等，提出建立统一医学科学数据共享平台、进一步完善法律法规与政策体系、构建合理的激励机制以及加强医学科学数据共享人才培养和生态建设等建

议。本研究不仅为理解医学科学数据共享的理论逻辑提供了新视角,也为解决当前实践中遇到的实际问题提出了切实可行的建议,期望能为进一步完善我国医学科学数据共享政策体系提供参考。但本研究仅聚焦理论框架构建和现有政策分析,对数据共享技术的具体应用和效果评估讨论不足。未来研究应更加关注相关技术解决方案在实践中的应用,进一步优化医学科学数据共享的技术支撑体系,并在此基础上探索在多方利益冲突中构建有效共享的路径和治理机制。

作者贡献: 杨琬平负责论文撰写与修订;刘泽坤负责论文修订;杨俊涛、李青负责研究设计、论文修订。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 清华大学统计学研究中心. 医疗大数据 [EB/OL]. [2024 - 09 - 15]. <https://tsing.v-dk.com/research/medical-big-data/>.
- 2 ALSHEIKH - ALI A A, QURESHI W, AL - MALLAH M H, et al. Public availability of published research data in high - impact journals [J]. Plos one, 2011, 6 (9): e24357.
- 3 TRIVEDI H, MESTERHAZY J, LAGUNA B, et al. Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson's natural language processing algorithm [J]. Journal of digital imaging, 2018, 31 (2): 245 - 251.
- 4 GREEN A K, REEDER - HAYES K E, CORTY R W, et al. The project data sphere initiative: accelerating cancer research by sharing data [J]. The oncologist, 2015, 20 (5): 464.
- 5 European Commission. General data protection regulation [EB/OL]. [2024 - 09 - 05]. <https://www.edps.europa.eu/>general - data - protection - regulation_en.
- 6 科学数据管理办法 [EB/OL]. [2024 - 09 - 10]. https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2018/201804/t20180404_139023.html.
- 7 关键. 医学科学数据共享与使用的伦理要求和管理规范 (一) 前言 [J]. 中国医学伦理学, 2020, 33 (3): 143 - 146.
- 8 HEY T, TANSLEY S, TOLLE K. 第四范式 [M]. 潘教峰, 张晓林, 译. 北京: 科学出版社, 2012.
- 9 国务院办公厅转发科技部等部门 2004—2010 年国家科技基础条件平台建设纲要的通知 [J]. 中华人民共和国国务院公报, 2004 (24): 16 - 19.
- 10 国家卫生健康委员会 2023 年 11 月 7 日新闻发布会文字实录 [EB/OL]. [2024 - 10 - 20]. <http://www.nhc.gov.cn/xcs/c100122/202312/d30f0fb6878649f68b7a052af351dc65.shtml>.
- 11 信息安全技术 个人信息安全规范: GB/T 35273—2020 [EB/OL]. [2024 - 10 - 20]. <https://std.samr.gov.cn/gb/search/gbDetailed?id=A0280129495AE05397BE0A0AB6FE>.
- 12 TUCKER K, BRANSON J, DILLEEN M, et al. Protecting patient privacy when sharing patient - level data from clinical trials [J]. BMC medical research methodology, 2016, 16 (S1): 77.
- 13 陈晓云, 张倩, 张成林, 等. 医疗机构泛知情同意和健康数据共享隐私保护的调查研究 [J]. 世界科学技术 - 中医药现代化, 2017, 19 (1): 118.
- 14 关键. 医学科学数据共享与使用的伦理要求和管理规范 (二) 隐私变迁与挑战 [J]. 中国医学伦理学, 2020, 33 (3): 288 - 293.
- 15 万佳林, 贾晓峰, 胡志民. 基于多案例研究的生物医学科学数据开放共享策略分析 [J]. 医学信息学杂志, 2024, 45 (2): 20 - 25.
- 16 关键. 法律框架下构建医学科学数据共享机制和保障体系 [J]. 中国科技资源导刊, 2020, 52 (2): 62 - 68.

欢迎订阅 欢迎赐稿