

跨机构临床数据共享平台构建与应用^{*}

王琳琳¹ 崔芳芳¹ 何贤英^{1,2}

(¹ 郑州大学第一附属医院 郑州 450052 ² 上海人工智能实验室 上海 200232)

〔摘要〕 **目的/意义** 设计共享服务平台，推动河南省临床数据跨机构共享应用。**方法/过程** 分析河南省跨机构临床数据共享面临的实际问题和需求，围绕数据采集、治理生产、交付共享 3 方面核心需求构建平台，开发标准化数据交互接口，搭建数据治理与开发一体化平台及共享服务平台。**结果/结论** 接入河南省 114 家医疗机构临床数据，建成含 2 589 个变量字段的多中心临床数据库及肺癌、女性两癌、糖尿病 3 个专病数据集，为医疗管理决策者、临床医务人员、医学研究人员提供高质量真实世界数据支持，并为我国区域临床数据共享应用提供参考。

〔关键词〕 临床数据；数据共享；数据治理；数据集

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2025.05.012

Construction and Application of a Cross-institutional Clinical Data Sharing Platform

WANG Linlin¹, CUI Fangfang¹, HE Xianying^{1,2}

¹The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China; ²Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China

〔Abstract〕 **Purpose/Significance** To develop a shared service platform, so as to promote the cross-institutional clinical data sharing and application in Henan province. **Method/Process** The practical problems and demands faced by cross-institutional clinical data sharing in Henan province are analyzed. A platform is constructed around 3 core requirements: data collection, governance and production, and delivery and sharing. Standardized data interaction interfaces are developed, and an integrated platform for data governance and development as well as a shared service platform are established. **Result/Conclusion** Clinical data from 114 medical institutions in Henan province are integrated into the platform. A multi-center clinical database comprising 2 589 variable fields is built, along with 3 disease-specific datasets for lung cancer, female cancers, and diabetes. These resources provide high-quality real-world data to support healthcare administrators, clinical practitioners, and medical researchers, and provide references for regional clinical data sharing and application initiatives in China.

〔Keywords〕 clinical data; data sharing; data governance; dataset

〔修回日期〕 2025-04-09

〔作者简介〕 王琳琳，高级统计师，发表论文 20 余篇；通信作者：何贤英。

〔基金项目〕 科技创新 2030——“新一代人工智能”重大项目（项目编号：2022ZD0160705）；河南省科技攻关项目（项目编号：242102311012）；河南省高等学校重点科研项目（项目编号：23A330007）。

1 引言

医疗信息化进程加速,促使医疗机构积累海量临床诊疗数据^[1]。这些数据不仅是医院运营、医疗管理的重要支撑,也是临床医学研究不可或缺的真实世界数据资源,已成为我国基础性战略资源的重要组成部分^[2]。真实世界临床数据(尤其是多中心数据)的共享应用价值巨大。目前,医疗机构内部临床数据共享相对成熟,2023 年度研究型医院(综合医院)排名前 30 位的医院中,16 家已建设院级医疗大数据平台。然而,跨机构临床数据共享应用仍较有限^[3-4]。河南省作为人口大省,临床数据资源丰富。据《中国统计年鉴》^[5]数据,2023 年河南省总诊疗人次 6.84 亿、出院人数 2 290.5 万,分别居全国第 4 位和第 2 位。有效汇聚并利用这些数据意义重大。2014 年河南省远程医学中心依托郑州大学第一附属医院,采用 10M/20M 光纤专网建设了覆盖全省 18 个市、108 个县、130 余家市县级医院的远程综合会诊系统^[6]。2021 年该中心升级搭建了基于临床数据交互共享的远程会商系统^[7]。本研究依托上述系统中的医疗专网和数据接口,基于大数据技术,通过数据采集汇聚、治理生产、应用平台开发等环节,构建跨机构临床数据共享平台,促进河南省临床数据的跨机构共享应用。

2 跨机构临床数据共享面临的问题及需求分析

2.1 制约跨机构临床数据共享的问题分析

多中心、高质量的临床数据是实现有效共享的坚实基础,然而河南省在跨机构共享方面仍面临以下挑战。一是不同医疗机构临床数据异构性显著。不同医疗机构之间,乃至同一医疗机构内部不同系统之间,在建设厂家、数据架构、数据库、术语体系及编码规则等方面存在显著差异^[8]。以医院信息系统(hospital information system, HIS)为例,我国三级医院信息系统平均部署 23 类厂商的异构系统^[9],而河南省情况更为复杂,涵盖 27 类不同厂商,且数据字段匹配率仅 38%,如“血压”字段存

在“BP”“血压值”“收缩压 & 舒张压”等 12 种不同表达形式。这种“数据烟囱”现象不仅增加数据预处理成本,还使跨机构数据难以直接满足共享需求。二是数据安全及隐私保护挑战。医疗数据蕴含患者个人信息和详细健康情况,在共享过程中面临多重风险^[10]:在公网环境下,数据可能遭受中间方攻击或被拦截,导致数据泄漏;敏感信息若未脱敏处理,患者隐私极易泄漏^[11];访问控制策略失效可能引发越权访问风险^[12]。以上均加剧了医疗机构的数据封闭倾向。

2.2 跨机构临床数据共享平台建设需求分析

跨机构临床数据共享平台建设对满足河南省医疗管理决策者、临床医务人员、医学研究人员等多方需求意义重大。一是医疗管理决策者亟须通过精准数据实现医疗资源动态调配和运营效率提升,特别是在分级诊疗背景下优化全省医疗资源配置时。二是临床医务人员依赖全面且准确的历次诊疗数据以更好地理解患者病情变化,作出更精准的诊断并制定个性化治疗方案,特别是在河南省远程医疗服务场景下进行复杂病例的连续管理中。三是医学研究人员需要多中心、多模态的真实世界数据,以突破单体机构样本量不足的瓶颈,尤其是在河南省开展大规模流行病学研究、临床转化医学研究等科研场景下。

3 跨机构临床数据共享平台构建

3.1 总体架构

结合河南省跨机构临床数据共享实际需求,综合运用大数据技术,设计涵盖数据采集、数据治理与生产、数据交付与共享 3 大核心模块的跨机构临床数据共享平台,平台架构,见图 1。一是数据采集体系。通过标准化接口和远程医疗专网将分散在不同医疗机构的数据汇聚至省中心。二是数据治理与生产体系。采用逻辑分层架构进行数据治理,将汇聚的各医疗机构数据分为贴源层、标准层和应用层;同时,构建数据治理与开发一体化平台,实现治理作业的流水线式调度和数据全生命周期管理,提高数据治理效率和质量。三是数据交付与共享体

系。通过开发共享服务平台及构建专病数据集，为医疗管理决策者、临床医务人员、医学研究人员等数据使用方提供多中心、高质量的临床数据服务。

数据治理方和数据使用方通过桌面云环境安全访问数据，确保全流程安全性与合规性。

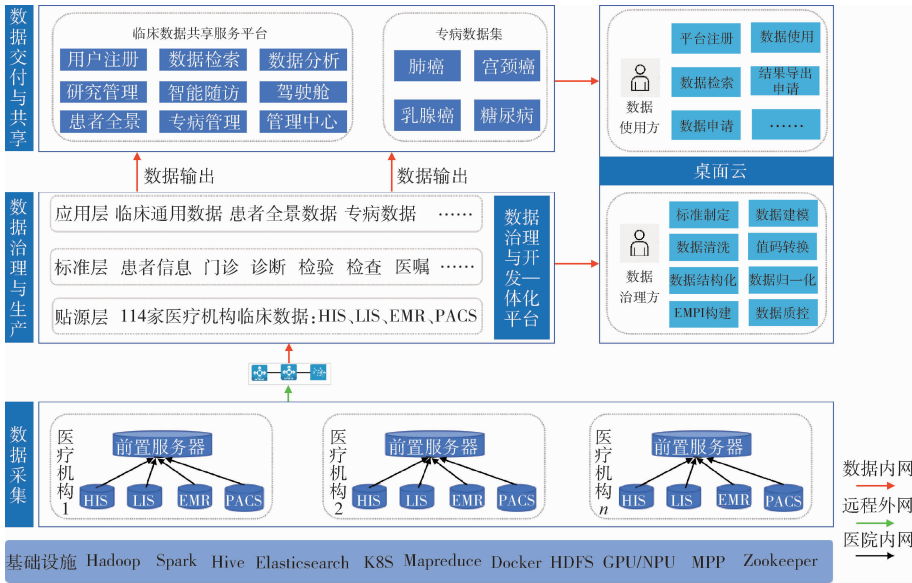


图 1 跨机构临床数据共享平台架构

3.2 跨机构临床数据采集体系建设

针对河南省不同医疗机构临床数据异构性问题，集成 SQL 视图与 Web 服务技术，开发标准化数据交互接口，实现高效协同的跨机构数据采集。一是数据采集标准制定。依据数据结构调研和行业专家咨询，制定跨机构数据采集标准，覆盖 HIS、电子病历（electronic medical record, EMR）、检验信息系统（laboratory information system, LIS）、影像存储与传输系统（picture archiving and communication system, PACS）等核心医疗系统，涵盖患者基本信息、诊断、医嘱、病历、检验、检查、影像、手术、治疗、随访等 16 类数据指标。二是数据采集技术。在河南省 114 家医疗机构部署前置机系统。该系统由前置机接口库和前置机程序两大核心组件构成，其中前置机程序调用接口库功能模块，按定义好的接收标准转换数据，完成数据采集与整理。

3.3 跨机构临床数据治理与生产体系建设

面向河南省跨机构临床数据存储、治理与管理综合需求，采用 Hadoop + MPP 湖仓一体架构，构建

具备高扩展性与灵活性的临床大数据中心。

3.3.1 数据治理与开发一体化平台建设 针对传统数据治理工具分散、流程割裂导致的效率低下问题，集成医疗行业专用数据治理工具和通用大数据组件，设计并开发多模态医疗数据治理与开发一体化平台，功能架构，见图 2。

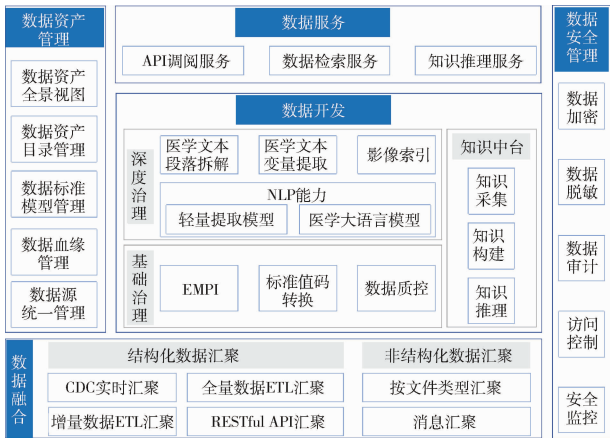


图 2 多模态医疗大数据治理与开发一体化平台架构

平台通过统一工作流调度，实现一站式数据融合、数据资产管理、数据开发、数据服务和数据安全。一是数据融合，通过变更数据捕获

(change data capture, CDC)、抽取 - 转换 - 加载 (extract - transform - load, ETL)、消息队列等技术, 满足结构化与非结构化、存量与增量、实时与流式等数据采集。二是数据资产管理, 通过数据标准管理、主数据管理、数据资源管理等子模块, 支撑构建统一数据标准与模型, 并生成数据治理全流程数据血缘关系图谱。三是数据开发模块, 涵盖基础治理、深度治理、知识中台等子模块。其中, 数据质控功能通过内置常用规则模板, 利用智能化校验机制, 从完整性、准确性等维度^[13-14]提高数据可信度和有效性。四是数据服务, 支持数据浏览、查询、下载, 关键词搜索、条件筛选, 智能化分析和推理等功能。五是数据安全

管理, 集成用户身份验证、数据加密及脱敏、数据审计、访问控制等安全机制, 并建立 5 级数据分类原则实现分级访问控制, 确保治理与共享全流程安全性。

3.3.2 跨机构临床数据治理与生产 依托数据治理与开发一体化平台, 对河南省 114 家医疗机构临床数据进行治理。先结合文献资料、标准指南及专家咨询, 构建临床数据模型, 再通过值码转换、数据映射、医学文本结构化、患者主索引 (enterprise master patient index, EMPI) 构建等, 完成跨机构临床数据的治理和标准化生产, 具体流程, 见图 3。同时, 为保证数据质量, 在数据治理与生产全流程中开展数据质控, 包括质控规则制定、数据治理人员生产自测及问题核查、质检人员核查等, 最终形成数据质量报告。

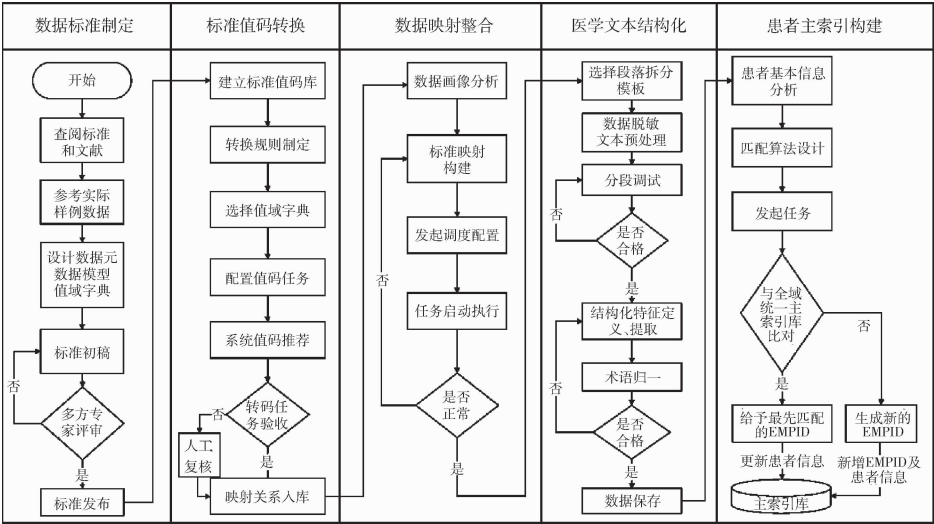


图 3 跨机构临床数据治理与生产流程

3.4 跨机构临床数据共享体系建设

3.4.1 临床数据共享服务平台建设 面向医疗管理决策者、临床医务人员、医学研究人员等多方用户需求, 采用微服务架构设计并开发临床数据共享服务平台。平台涵盖数据驾驶舱、智能检索、数据分析、患者全景、临床医学研究管理、智能随访、专病管理、管理中心 8 个功能模块, 并通过 Redis 与 Kafka 等中间件分布式部署, 见图 4, 实现平台化运营指标管理、患者管理、研究管理。其中, 数据驾驶舱可视化展示各层数据总体情况及平台数据

应用情况, 支撑医疗管理决策者精准决策; 患者全景基于 EMPI 构建患者历次就诊时间轴、诊疗详情及 3D 画像, 辅助临床医务人员精准诊疗。智能检索包括模糊检索、条件树检索、批量检索, 由 Elasticsearch 组件索引库提供高效搜索支持; 数据分析模块集成 40 余种常用数据预处理与分析算法, 并内嵌 R 语言、Python 语言等工具; 临床医学研究管理包括研究项目创建、研究人群管理、研究成果管理; 智能随访包括随访概览、随访项目、病例报告表 (case report form, CRF) 管理和任务管理; 专病管理支持单病种数据精细管理。这 5 个功能模块协

同支撑医学科研人员多中心研究。此外，管理中心支持数据申请、审批等流程。

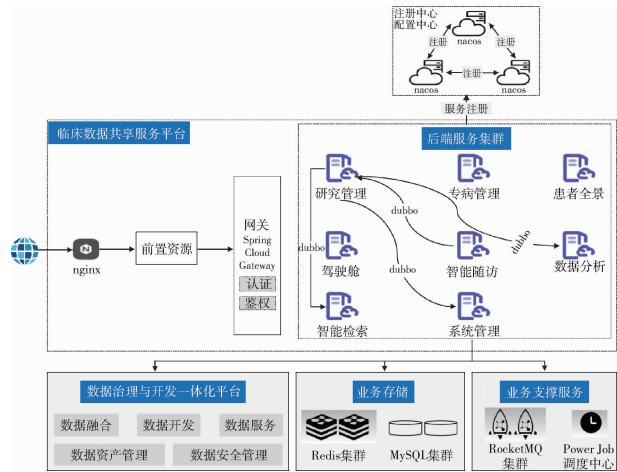


图 4 临床数据共享服务平台部署架构

3.4.2 专病数据集建设 面向特定病种或疾病，通过深度数据治理和复杂逻辑结构化处理，建设标准化数据集，为专病智能诊疗模型构建提供高质量数据支持。选取高发恶性肿瘤（肺癌、女性两癌）和常见慢性疾病（糖尿病），建设专病数据集。

3.4.3 数据共享服务管理 通过网络级逻辑隔离技术，在远程医疗专网基础上构建内部数据网络，并部署桌面云系统。数据使用者通过桌面云终端安全访问共享服务平台，按权限查看驾驶舱指标、检索数据或申请数据，其流程，见图 5。此外，在省中心部署融合 GPU 和国产化 NPU 算力的人工智能计算平台，峰值算力达 13.2PFlops，满足多模态、大规模数据分析计算需求。

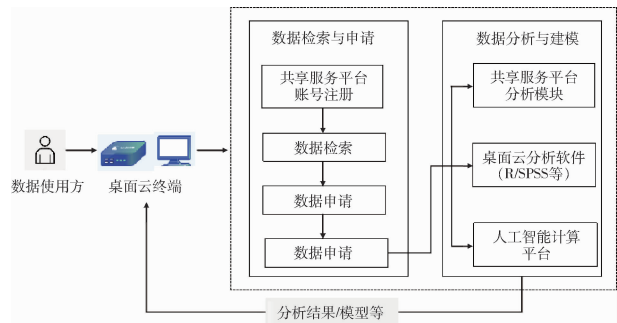


图 5 临床数据共享服务平台数据使用流程

4 应用效果及讨论

4.1 应用效果

经过近 4 年分期建设，河南省跨机构临床数据共享平台已上线应用。数据采集方面，平台已完成 114 家三级医疗机构 HIS、LIS、EMR、PACS 等核心医疗系统 16 类数据指标采集。通过标准化数据交互接口，实现患者跨机构诊疗数据的高效采集。在数据治理与生产方面，基于本研究构建的数据治理与开发一体化平台，结合临床诊疗流程，对 16 类数据指标进行模型重构及基础、深度治理，构建多中心临床数据库，涵盖 2 589 个变量字段、48 亿条数据，覆盖 2 163 万例患者、11 549 个病种、988 个术种，有效解决不同医疗机构数据异构问题。数据交付和共享方面，临床数据共享服务平台已上线运行。通过数据驾驶舱模块的数据资源、科研指标、运营指标等大屏，满足医疗管理决策者多维度运营指标分析需求。通过患者全景模块，助力临床医务人员全面了解跨机构就诊患者历次诊疗信息、充分掌握病情。通过智能检索、临床医学研究管理、数据分析、智能随访、专病管理等模块，满足医学研究人员多维科研数据收集与分析需求，并建成 3 个专病数据集，覆盖 9.9 万例肺癌患者、6.8 万例女性两癌患者、37.1 万例糖尿病患者。数据安全与患者隐私保护方面，构建多维度、多层次防护措施，通过远程医疗专网及内部数据网络、基于桌面云技术的安全可控数据访问环境、基于先进加密技术的患者隐私字段脱敏、依据 5 级数据分类原则的分级访问控制等，有效防范各类数据泄漏风险。

4.2 存在的不足和建议

真实世界临床数据共享对促进医院精准管理、临床精准决策、医疗创新和成果转化意义重大。本研究构建的平台为河南省临床数据跨机构共享提供安全可信的支撑，但当前应用主要聚焦运营管理、患者诊疗、临床医学研究等非营利性服务，在更广泛数据共享场景中仍显不足。一是数据接入范围有限，平台尚未接入患者健康监测等院外数据，而这

些数据对疾病早期预警、治疗效果评估等具有重要价值^[15-16]。对此后续将合理选择典型病种和监测指标,接入高危、患病人群连续监测数据。二是平台暂无法提供商保核保、临床辅助决策等场景的应用程序编程接口(application programming interface, API)调用服务。对此将持续优化平台支撑技术,在保障数据安全前提下,开发需求量较大的数据调用服务 API。三是因缺乏明确获益机制,平台暂无法向医药、保险、健康管理等第三方企业开放数据应用。对此后续将融合卫生经济学研究方法,从定价机制、收益分配机制等角度推进临床数据转化应用。

5 结语

本研究在深入分析河南省临床数据共享面临的问题与多方需求基础上,围绕数据采集、治理生产、交付共享 3 方面需求,构建跨机构临床数据共享平台,有效推动河南省临床数据共享应用。尽管本研究构建的平台具有一定创新性和可推广性,但仍存在数据接入范围有限、数据共享模式不丰富、获益机制未健全等不足。在未来的建设和应用中,一是接入健康监测等院外数据,丰富数据模态;二是开发疾病预警、早筛、精准诊疗、预后预测等模型及 API,为互联互通体系内的基层医疗机构提供辅助决策支持,助力河南省医疗服务同质化;三是健全获益机制,构建多方获益的数据价值体系,推进河南省临床数据的资产化。

作者贡献:王琳琳负责研究设计、论文撰写;崔芳芳负责文献调研、数据采集、论文修订;何贤英负责研究设计、论文审核。

利益声明:所有作者均声明不存在利益冲突。

参考文献

- 1 宋佳丽,涂志炜,黄潇.医疗大数据信息平台创新医院管理应用的探索与实践[J].现代医院,2024,24(10):1591-1594,1598.
- 2 关于促进和规范健康医疗大数据应用发展的指导意见[EB/OL]. [2024-06-10]. https://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm.

- 3 母晓莉,徐俊,鞠伟卿,等.基于医联大数据的实时数据平台的建设与应用[J].中国卫生信息管理杂志,2018,15(1):70-73.
- 4 郭一,胡盈盈,胡斐,等.区域健康医疗数据开放共享系统设计与应用[J].医学信息学杂志,2024,45(8):84-88,95.
- 5 国家统计局.中国统计年鉴(2024)[M].北京:中国统计出版社,2024.
- 6 赵杰,陈保站,李陈晨,等.面向区域医疗协同的远程医疗综合会诊系统设计与应用[J].医学信息学杂志,2016,37(5):20-23,45.
- 7 石金铭,卢耀恩,叶明,等.河南省疫情防控远程会商系统的建设与应用[J].中华医院管理杂志,2022,38(2):125-128.
- 8 吴燕秋,黎美秀,丁元杰,等.面向临床科研的全院级医疗大数据平台建设 with 数据治理实践探索[J].中华医学科研管理杂志,2021,34(2):81-86.
- 9 国家卫生健康委统计信息中心.全民健康信息化调查报告[M].北京:人民卫生出版社,2021.
- 10 陈怡.健康医疗数据共享与个人信息保护研究[J].情报杂志,2023,42(5):192-199.
- 11 QUAN G, YAO Z, CHEN L, et al. A trusted medical data sharing framework for edge computing leveraging blockchain and outsourced computation[J]. Heliyon, 2023, 9(12): 25.
- 12 WEI K, LI J, DING M, et al. Federated learning with differential privacy: algorithms and performance analysis[J]. IEEE transactions on information forensics and security, 2020, 15(4): 3454-3469.
- 13 吴骋,王志勇,徐蕾,等.基于人工智能的电子病历数据质量控制[J].解放军医院管理杂志,2021,28(2):134-135.
- 14 刘伟伟,王立军,庞娟,等.基于自然语言处理的肿瘤专科病历质控系统建设[J].医学信息学杂志,2024,45(2):77-81,91.
- 15 史森中,张和华,周超,等.基于 5G 智能终端的居家养老健康监测管理系统构建[J].医学信息学杂志,2023,44(9):74-79.
- 16 侯乐乐,邱山虎,孙子林.扫描式持续葡萄糖监测在糖尿病患者居家管理中的价值及展望[J].中华糖尿病杂志,2022,14(12):1363-1367.