

公立医院多院区协同背景下数据治理探索与实践^{*}

侯梦薇 兰欣 杨茜 卫荣

(西安交通大学第一附属医院 西安 710061)

〔摘要〕 目的/意义 针对日益增长的多院区数据协同管理需求,探索数据治理模式和方法路径。方法/过程 深入分析多院区医疗数据融合使用过程中存在的主要问题,根据各院区管理模式、信息化建设、技术架构等方面现状,针对回顾性数据和前瞻性数据采用不同模式与方法进行治理。结果/结论 医院多院区数据标准统一,数据共享率、使用效率明显提升,多院区业务系统数据质量不断改进,跨院区间数据安全保障机制得到完善。

〔关键词〕 数据治理;多院区管理;公立医院

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.08.005

Exploration and Practice of Data Governance in Multi-hospital Management Mode of Public Hospitals

HOU Mengwei, LAN Xin, YANG Xi, WEI Rong

The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

〔Abstract〕 **Purpose/Significance** Exploring data governance models and method paths in response to the growing demand for collaborative management of data across multiple institutions. **Method/Process** Thoroughly analyze the main problems in the integration and use of medical data in multiple hospital areas. Based on the current management mode, information construction, technical architecture, and other aspects of each hospital area, different models and methods are adopted for governance of retrospective and prospective data. **Result/Conclusion** The data standards for multiple hospital districts are unified, and the data sharing rate and usage efficiency are significantly improved. The data quality of the multi-hospital district business system is continuously improved, and the cross hospital interval data security guarantee mechanism is improved.

〔Keywords〕 data governance; multi-hospital management; public hospitals

1 引言

医院数据是医疗业务在信息系统中运行的产

物,也是大数据分析、人工智能等医院数据应用的原材料^[1-2]。健康医疗大数据在提升医疗服务质量、精准防治疾病、辅助管理决策、增强突发公共卫生事件处理能力等方面发挥较大作用^[3]。随着数字技术和信息技术在医疗领域的发展和应用,互联网医院和多中心区域协作系统已成为趋势^[4]。协同模式可实现医疗资源在参与医院之间的个性化配置,一方面使患者从中受益,另一方面使公立医院充分利用其资源,提高服务效率。良好的数据治理不仅是

〔修回日期〕 2023-05-17

〔作者简介〕 侯梦薇,中级工程师,发表论文 25 篇;通信作者:卫荣。

〔基金项目〕 陕西省社会发展科技攻关项目(项目编号:2020SF-285)。

医院多院区管理模式顺利执行的前提,也是医疗单位信息化及管理水平的侧面体现。

在多院区管理模式下,医院对数据的要求和管理手段开始转变^[5],主要体现在以院区间数据共享为目的,以支撑协同业务开展为基础,对数据采集与存储模式进行调整^[6]。由于多院区数据管理涉及数据共享,须构建与之相契合的多院区数据治理模式和数据标准化监管路径,解决数据标准不统一、各院区间数据无法共享、不同院区数据质量良莠不齐以及隐私安全风险等问题^[7]。因此,本文基于多院区管理模式,首先对医疗大数据治理及国内外研究现状进行阐述;其次总结并分析西安交通大学第一附属医院不同院区数据融合共享时存在的问题和实际情况,探讨问题产生的根本原因;进而提出解决方案,并从治理模式和方法策略两个方面进行系统性说明;最后,阐述解决方案实际应用效果。

2 医疗大数据治理概述

2.1 医疗大数据治理

医疗大数据治理^[8]是指依据相关政策,对医院、医疗机构、在线医疗机构和其他组织健康数据的收集、分析和使用进行监督和管理,通过不断创新管理方式提高健康数据质量,为医疗行业创造更大价值。数据治理具体包括医疗数据标准化、数据质量提升、数据管理和数据有效应用^[9]。医疗数据来源于患者诊疗过程,包含医学管理、医学影像、患者基本信息、医嘱、出入院数据、病历数据等。医疗大数据不仅具有大量性、高速性、多样性和价值性,还具有数据时间性(随着患者疾病发展周期而产生时序性结构数据)、非标准性(不同医生记录病历的习惯不同,没有统一标准)、不稳定性(医疗实验结果可能存在偏差)等特点,使构建数据治理体系的难度增大^[10]。因此,医疗大数据治理不仅要提升数据质量,更要建立系统性的治理框架,明确数据流程和权责分工。这标志着医院数据工作的全面进步,从过去单一的信息管理转变为如今面向治理的数据工作方法,实现从信息到数据、从管理到治理的跨越式提升。

2.2 医疗大数据治理研究现状

随着信息技术和生命科技的不断发展,医疗大数据在诸多领域得到广泛应用,成为融合最深、最具活力且辐射最广的新经济引擎。医疗大数据应用成为医疗健康行业发展的必要途径,世界各国深刻认识到其价值,对其应用及治理工作格外重视。美国建立医疗电子信息主动监测系统,利用大数据实现分散数据资源整合,现已建立覆盖本土完整的医疗健康大数据库。英国国家医疗服务系统收集全国医疗数据为本国医务人员提供帮助。加拿大颁布《个人健康信息保护法》,美国发布《大数据研究和发展计划》《美国联邦政府医疗信息化战略规划(2015—2020)》,以推动医疗数据治理发展^[11]。

我国医疗大数据产业起步较晚,在相关政策、技术、环境等方面影响下,正快速发展。2016年《全国医院信息化发展标准与规范(试行)》中指出,卫生信息化的发展过程应以相关数据标准和技术规范为基础。2018年《关于促进“互联网+医疗健康”发展的意见》中提出要提高卫生服务的效率和质量,《“健康中国2030”规划纲要》和《关于推进“互联网+医疗卫生”的意见》则是卫生信息管理的实施方案^[12]。近年来我国已先后建成多个医疗大数据中心和省级医疗数据平台,针对多源异构数据融合、治理进行探索。例如,2018年上海市成立数据中心,实现500家公立医院数据共享。医疗数据治理研究与实践方面,中山大学中山医学院基于通用数据模型相关理论^[13]、方法与技术开展健康医疗大数据治理并建立相关研究平台,为跨机构、跨部门工作提供高质量医疗数据。兰蓝等^[14]从医疗机构角度出发,提出健康医疗数据治理对策建议,为健康医疗数据在医疗机构内部和外部共享提供理论基础。四川省卫生健康委员会^[15]提出并探索“医疗机构信息化监管”,通过数据挖掘、人工智能等手段,为医院、医生、护士及医疗活动构建监管体系,最终实现全程监控和智能审核。

综上所述,数据治理理念已经涉及医疗领域,但在如何提高多院区管理模式数据质量方面的研究和应用较少,应列入研究重点。

3 多院区医疗数据存在的主要问题

3.1 不同院区数据标准不统一

数据业务表与主数据字典表无法一一对照，导致各业务系统数据割裂。在多院区不同信息系统中，经常存在同一字段具有不同编码、条目、参考标准、命名规则等现象。此外，部分数据由人工自定义生成，导致不同业务系统数据标准规范不统一。

3.2 院区之间数据无法共享

各院区业务系统建设时间不同且缺乏顶层设计，存储系统及数据类型多样，存在一定程度的信息孤岛问题，导致不同院区医疗数据之间没有关联关系，院区之间数据无法共享，多院区数据整合利用难度和成本较高。

3.3 多院区数据质量良莠不齐

医疗数据质量体现在完整性、规范性和整合性

方面。多院区医疗数据来源于不同业务系统，诊断、用药、手术名称等医学术语标准多、定义不统一，医疗数据质控体系不完整。因此，无法保证不同前端的数据质量。

3.4 跨院区数据共享存在较大安全风险

医疗数据包含大量个人隐私，跨院区医疗数据安全管理能力参差不齐，用户访问权限、数据脱敏处理水平不一，存在数据泄露和被窃取风险。

4 多院区数据治理模式与方法路径实践

4.1 数据治理模式的提出

4.1.1 回顾性与前瞻性数据治理模式对比 医院数据可划分为回顾性数据和前瞻性数据，前者指对历史数据进行治理，使其符合管理要求；后者指对业务流程、管理模式进行调整，确保系统未来产生的数据符合医院要求，见表 1。

表 1 回顾性与前瞻性数据治理模式对比

项目	回顾性治理	前瞻性治理
场景	针对历史数据进行治理，系统建设时数据利用要求相对较低，目标较模糊	针对未来产生的数据进行治理，系统建设时对数据质量要求较高，目标较清晰
数据治理时间	数据产生后	数据产生前
组织架构	主要是信息部门针对数据问题进行治理，组织相对薄弱	全院针对数据问题进行管理流程优化，组织相对完善
管理模式	由信息部门进行管理，主要局限在数据采集、汇聚、加工等操作	由医院从上层进行管理，针对数据产生的源头开始处理，数据全生命周期由多部门协同管理

4.1.2 回顾性数据治理 在多院区数据治理过程中，包含大量存量数据，其产生的系统、时间以及数据规范不同，为多院区数据治理工作带来挑战。通过建立医院临床数据中心并对不同院区的医疗数据进行汇集、清洗、集成、深度加工等操作，提升医疗数据质量，方便后期分析和利用。在数据汇集阶段，由于多院区信息系统建设时间、厂商和目标大多不一致，因此需要先明确数据汇集范围并确认采集方式，再执行数据采集工作。在数据清洗阶段，将不规范的编码和格式不符合要求的数据通过类型分析、清洗策略制定、实施清洗以及质量验证

4 个步骤进行处理。在深度加工阶段，针对医疗数据进行文本数据结构化处理和脱敏处理，使数据的可利用性和安全性达到医院要求。

4.1.3 前瞻性数据治理 医院既要针对历史数据进行回顾性治理，还要构建数据源头的前瞻性数据治理体系，从优化业务系统数据库表结构及流程出发，从源头生成高质量数据。前瞻性治理是一项全院工程，除了回顾性治理要素外，还应对医院各院区管理流程、组织架构设置、数据安全和隐私保护等问题进行统一管理。

4.2 数据治理方法路径实践

4.2.1 建立数据治理目标与标准 医疗数据是医院进行管理决策的重要依据，也是医院资产的重要组成部分。数据治理目标可以分解为以下 3 方面。一是对不同院区数据标准及指标进行整理，并形成统一的多院区主数据字典。二是对医疗数据进行摸底，明确各院区不同业务系统数据产生的源头，并明确部门责任，落实“谁产生数据谁对数据质量负责”的要求。三是构建包含数据规则制定、数据核减等操作在内的明确的数据闭环管理流程。多院区数据治理是一项长期、复杂、系统的活动，涉及大量数据、主题和指标的使用。治理工作应遵循“先计划、后行动”原则，建立一个自上而下、负责任、分层次的组织结构，选择适当的技术方案和工具，完善治理过程中的工作机制和内容，协调发展和实施各项任务，见图 1。

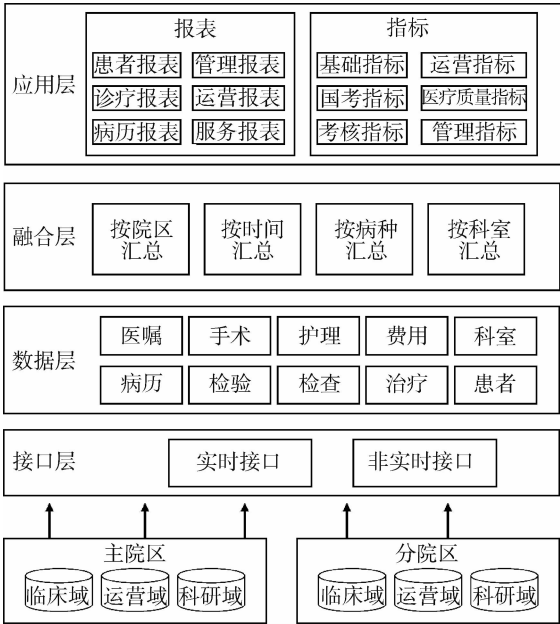


图 1 数据治理体系框架

4.2.2 实施流程 参照《卫生管理基本数据集》等医疗行业数据标准规范制定标准，探索多院区数据治理方法路径，并将数据治理实施步骤归结为 4 个方面。(1) 统一数据标准。建立统一数据标准是多院区数据治理体系的基础，也是数据治理成功的关键^[16]。数据标准的定义是制定一个共同的文件格式

式框架，支持不同信息系统之间的数据规范、分类代码、医学术语和质量参数的沟通和交换。在数据标准制定阶段，共将 76 个医院数据字典分为科室字典、员工字典和诊断领域等 8 大类，并根据行业标准 and 国际标准对数据字典进行标准化定义，以制定医院区域内的通用数据标准指标^[17]。(2) 数据格式校验及清洗。在规范标准后着手建立数据中心，所涉及内容包括临床数据中心、管理数据中心、影像数据中心等。对不同院区的主数据、元数据和所有医疗数据源头进行标准化管理，所有数据经过接口初步验证后开始进行错误处理，主要检查医疗数据格式是否满足相关标准、数据内容是否正确。由于医疗数据的多样化，要保持其唯一性和可靠性还需要进行数据清洗。根据预先设置好的数据清洗规则^[18]，通过数据提取 - 转换 - 加载 (extract - transform - load, ETL)、校验和清洗后才能进入数据中心，见图 2。(3) 信息系统架构统一。新系统通过整合医院内部数据和业务流程，实现架构的统一性。该系统合理组织数据交换、分类、传输等，使医院的信息架构、应用架构和技术架构保持高度一致，有效支撑各专业领域的医疗服务。(4) 建立保护信息安全机制。为了保护患者隐私、保障信息安全，医院利用信息安全管理工具和技术手段，为不同部分建立信息保护机制^[19]，包括但不限于：研究行业准则，对医院不同部门的数据进行风险分析；动态评估安全状况，对医院不同部门的数据进行分类分级；利用风险敏感技术，制定相应安全策略和制度；进行攻防演练的持续安全行动等。

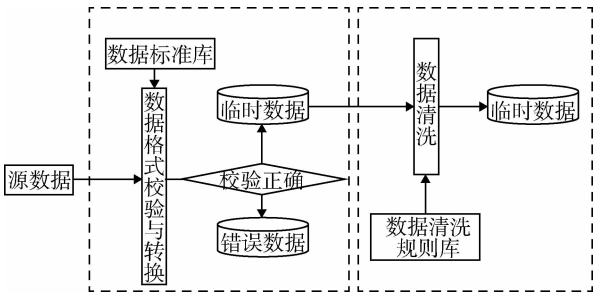


图 2 数据格式校验及清洗

通过“计划 - 执行 - 检查 - 处理” (plan - do - check - action, PDCA) 循环机制，结合数据实时

核查，持续优化各院区存在数据质量问题的业务系统，即发现问题、确认问题、解决问题、持续改进。通过持续的数据质量分析，不断发现问题并提出改进建议，推动业务系统不断优化，从源头严格控制数据质量；同时，做好院区人员培训和考核，提高职工的数据责任意识。

5 数据治理效果

5.1 总体效果

对多院区 58 个医疗业务系统 2006 年以来的所有数据进行汇集，总量超过 10 亿条，其中 80% 为非结构化数据。经过数据集标准化、共享文档标准化、电子病历改造、业务系统功能改造等，在多院区数据治理工作中重点管控数据质量，实现多院区业务系统数据质量不断改进，取得一定治理成效。

5.2 统一数据标准，提升医院综合管理水平

通过建立多院区统一规则及相关标准，完成多院区包括检验、检查、输血、治疗、用药在内的 12 个业务流程闭环，通过实时记录关键节点数据实现医疗过程全流程监控和质控管理，提升医院综合管理水平。

5.3 统一数据中心，支持多院区信息共享

为提高医院多院区间数据共享率和使用效率，

采用统一患者管理系统和数据中心收集各院区不同部门的患者数据。由于采用统一功能设计，患者的医疗记录、检查结果、医学影像等数据可以在不同院区间实现畅通共享。支持多院区远程会诊，与传统远程会诊不同，新系统利用跨区域数据管理技术，整合不同院区的医疗数据与医学影像，使会诊医生能够直接访问合作医院的电子病历与医学图像，有效解决传统跨院区会诊时信息不对称的问题。

5.4 提升数据质量，通过国家评审

《电子病历系统应用水平分级评价标准（试行）》将医疗数据质量作为主要评价方面，尤其关注电子病历数据质量。医疗数据治理水平不仅决定电子病历分级评价结果，更关乎疾病诊断相关分组（diagnosis related groups，DRG）入组精准性。从长远来看，数据质量是医疗质量监管和推进医疗质量安全的核心要素，是医疗机构实现高质量发展的有力工具，也是国家对医疗机构考核管理的重要依据。高质量的医疗数据对医疗行业发展具有重要意义。西安交通大学第一附属医院针对多院区医疗数据进行科学化和精细化治理工作后，数据质量得到大幅提升，在国家电子病历 5 级评审及互联互通评审中得到肯定，数据达标率超过 96%，见表 2。

表 2 数据治理前后符合率对比

医院项目	数据表字段名称	数据治理前符合率（%）	数据治理后符合率（%）
检查项目名称	EMAR_ PACS_ REPORT. ITEMNAME	100	100
检查项目编码	EMAR_ PACS_ REPORT. ITEMCODE	95	100
检查描述	EMAR_ PACS_ REPORT. DESCRIPTION	75	97
诊断	EMAR_ PACS_ REPORT. DIAGNOSE	75	97

5.5 建立多院区数据安全制度，完善跨院区信息安全保障

建立保护医院数据的机制和系统，通过实时监测、分析、响应、报告和其他安全功能，以及雇用人员管理安全、控制和操作，确保医院数据安全，

抵御各类攻击，确保医院数据安全事件数量逐年下降。

6 结语

医院多院区数据管理架构的优势在于：统一数

据标准,促进不同院区间的数据交换与利用;提高数据质量,发挥数据价值;加速院区间的数据共享;提升信息技术投资回报;优化医疗服务质量。实施数据治理模式,可确保数据仓库标准化管理、业务系统数据字典共同维护、经验教训共享,以提早解决数据质量问题,推动后续治理工作开展。

参考文献

- 1 张义丹,胡豫,许栋,等.三级公立医院高质量发展的内涵要义与实施重点探讨[J].中国医院管理,2021,41(10):1-3.
- 2 张岚.基层医院数据治理方法的探索与实践[J].信息系统工程,2020(9):87-88.
- 3 王志勇,吴聘,王立鹏,等.医疗大数据背景下的数据治理与质量监管[J].中国数字医学,2021,16(4):92-96.
- 4 李言生,韩士斌,苏明亮,等.医院真实世界数据治理平台建设探索[J].中国卫生信息管理杂志,2022,19(6):925-930.
- 5 俞鹏飞,罗颖文,刘建模,等.面向医院的大数据治理模型设计[J].医学信息,2021,34(10):18-20.
- 6 张琼瑶,黄基,李倩文,等.基于人工智能的大数据治理平台实践与探索[J].中国数字医学,2021,16(10):31-36.
- 7 张振,杨翠涓,徐静,等.健康医疗大数据应用发展现状与数据治理[J].医学信息学杂志,2022,43(7):2-8.
- 8 王静静,吴震,孙海萍,等.特色区域医疗信息平台构建及其在耳鼻喉患者管理中的应用研究[J].中国医学装备,2022,19(9):119-122.
- 9 姜永梅,赵宝春,陈强,等.多院区互联互通客户服务对“无障碍就医”的实践及效果[J].现代医院管理,2017,15(4):55-57.
- 10 吴龙,严晓明,陈秀娟,等.医疗数据湖建设及医疗数据治理探索[J].医学信息学杂志,2022,43(6):8-13.
- 11 叶琳,罗铁清.医疗数据治理综述[J].计算机时代,2021(5):10-12.
- 12 熊兴江.医疗大数据质量评价指标体系构建研究[D].武汉:华中科技大学,2019.
- 13 潘惊萍,谭坤,邓韧.省级医疗机构综合监管数据治理体系构建与应用[J].中国卫生信息管理杂志,2022,19(2):171-177.
- 14 兰蓝,李瑞,白波,等.医疗机构数据共享关键问题研究与数据治理对策[J].中国卫生信息管理杂志,2022,19(2):184-188.
- 15 张弘政,刘迷迷,李琳,等.基于通用数据模型的健康医疗大数据平台数据治理研究[J].医学信息学杂志,2022,43(6):2-7,13.
- 16 翟运开,路薇,张瑞霞,等.多维集成视角下精准医疗数据融合标准体系构建[J].中国卫生资源,2020,23(1):23-27.
- 17 陈敏,牟海燕,秦健.健康医疗大数据标准体系框架研究[J].中国数字医学,2018,13(4):14-16,33.
- 18 毛云鹏,龙虎,邓韧,等.数据清洗在医疗大数据分析中的应用[J].中国数字医学,2017,12(6):49-52.
- 19 王天罡,李晓亮,卫荣,等.医院数据安全综合防护体系建设十年实践[J].中国数字医学,2022,17(8):1-8.

关于《医学信息学杂志》启用

“科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力,端正学风,维护作者权益,《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”,对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库,可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为,本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度,自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部