中医临床数据资源管理与服务平台构建研究*

白文斌 肖 勇 田双桂 沈绍武

（湖北中医药大学 武汉 430065）

〔摘要〕 **目的/意义** 指导中医临床数据资源管理与服务技术平台建设，为释放中医临床诊疗数据活力提供参考。**方法/过程** 分析中医临床数据资源管理与服务平台建设需求，从总体架构、业务功能、数据安全等方面提出实现路径。**结果/结论** 中医临床数据资源管理与服务平台应用价值体现在中医临床多源异构数据集成、中医临床研究模式创新、中医药标准化建设、中医医院智慧化建设等方面。

〔关键词〕 中医临床；数据资源；管理服务；技术平台

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2023.08.015

Study on the Construction of the Traditional Chinese Medicine Clinical Data Resource Management and Service Platform

BAI Wenbin, XIAO Yong, TIAN Shuanggui, SHEN Shaowu

Hubei University of Chinese Medicine, Wuhan 430065, China

〔**Abstract**〕 **Purpose/Significance** To guide the construction of the traditional Chinese medicine (TCM) clinical data resource management and service technology platform, and to provide references for unleashing the vitality of TCM clinical diagnosis and treatment data. **Method/Process** The paper analyzes the requirements for the construction of the TCM clinical data resource management and service platform, and proposes implementation paths from the overall architecture, business functions, data security, and other aspects. **Result/Conclusion** The application value of the TCM clinical data resource management and service platform is reflected in the integration of multi－source data in TCM clinical practice, innovation of TCM clinical research modes, standardization construction of Chinese medicine, and intelligent construction of TCM hospitals, etc.

〔**Keywords**〕 TCM clinical; data resource; management service; technology platform

1 引言

以数据资源为关键要素，以现代信息网络为主要载体，以信息技术融合应用、全要素数字化转型为重要推动力的数字经济，推动着生产、生活和治理方式的深刻变革[1]。大数据时代[2]，云计算、大数据、移动互联网、人工智能等技术深入应用于中医药各领域，82.23%的中医医院已建立中医电子病历系统[3]，生产积累着海量复杂的中医临床诊疗数据，这些来源于中医临床诊疗过程的全量数据是传承创新中医药事业的重要资源。本文应用现代信息技术搭建中医临床数据资源管理与服务平台，聚

〔修回日期〕 2023－05－29

〔作者简介〕 白文斌，硕士研究生；通信作者：肖勇，高级实验师，硕士生导师。

〔基金项目〕 国家中医药管理局中医药项目（项目编号：GZY－GCS－2019－006）。

焦中医临床诊疗数据的采集标注、清洗脱敏、集成存储、交换共享、分析挖掘等，探索发现真实世界中医临床诊疗数据中的“事实与规律”，为中医药传承精华、守正创新提供高质量的数据供给，释放中医临床诊疗数据活力。

2 中医临床数据资源管理与服务平台建设需求

2.1 中医药传承创新发展需求

2021 年 1 月发布的《关于加快中医药特色发展的若干政策措施》明确提出要全方位提高中医医疗和中医药科研水平，加快推动解决中医药发展实践面临的突出问题[4]。传承创新中医药是打造中医药和西医药相互补充、协调发展中国特色卫生健康模式的必然要求[5]。数据利用有助于产品、流程和组织的改善和创新，促进企业劳动生产率增长 5% ~ 10%[6]，基于中医医院开展中医临床数据资源库建设，打破传统中医临床研究的惯性思维和路径依赖，以数据为驱动，规范整合中医医院临床数据资源，开展中医临床数据汇集、处理、转换、分析与挖掘，以数据开发利用为抓手，探索发现中医临床诊疗数据中的“事实与规律”，更好地发挥中医药原创优势。

2.2 智慧中医医院建设需求

数据已成为数字经济时代的新型生产要素。中医临床数据是中医医院智慧化建设的核心，在采集存储方面存在数据格式混乱、数据不完整、不一致、非结构化、存储分散等问题，在集成处理方面存在数据异构、质量参差不齐、数据融合难、有效分析工具缺乏、挖掘深度不足等难题，制约着在科研、管理和服务中的应用，需要建立中医临床数据资源管理与服务平台，聚焦抽取标注、清洗脱敏、集成存储、分析挖掘等关键环节，梳理数据标准制定与应用，研究数据抽取、数据清洗、数据结构化、数据转换、数据脱敏、数据汇交、数据关联、数据可视化、知识图谱、数据检索等关键技术和方法，构建多维度、全方位、多元化的中医临床数据处理技术方法库，以及中医临床数据挖掘模型，为构建智慧医疗、智慧服务、智慧管理一体化的智慧中医医院提供数据与方法支撑。

2.3 中医药标准化与信息化双向融合需求

中医临床数据资源管理与服务平台的构建是中医药标准化与信息化双向融合的创新实践。一方面需梳理中医临床科研一体化所需的数据标准，将其融入平台，强化数据标准在平台中的引领和基础支撑，提供标准关联、查询、下载、分发等功能，提升数据共享性和系统耦合度；另一方面通过平台实施，将数字技术融入中医药标准的制定、应用与管理中，建设一套涵盖术语、数据元、数据集、分类代码、主数据等的数据标准服务系统，构建中医临床数据抽取清洗、集成存储、交换共享、分析挖掘的标准体系。

3 中医临床数据资源管理与服务平台建设

3.1 总体架构

平台总体架构主要包括数据采集层、数据处理层、数据存储层、分析挖掘层、业务应用层，见图 1。数据采集层主要完成中医电子病历、中医病案首页、纸质病历等数据源中的患者基本信息、中医诊断信息、干预信息、疗效评价信息等数据采集与抽取，研究与应用多源异构数据汇交技术，完成中医临床诊疗研究所需的各类数据采集与集成；数据处理层基于中医临床数据标准库，采用自然语言处理、数据归一、属性编码等技术将采集抽取的中医临床数据转化为可分析数据，形成可用于分析挖掘的中医临床病例数据（case report form，CRF）库、中医临床科研数据资源（research data resource，RDR）库等；数据存储层集成 Hive、Hbase、MongoDB、SQL Server、MySQL 等技术，主要解决中医临床数据采集抽取、转换编码、装载归集形成的不同类型数据集中存储问题；分析挖掘层基于存储的 CRF 库与 RDR 库，集成关联分析、聚类分析、特征分析、预测分析、神经网络、复杂网络等挖掘算法和预测模型，从不同维度、不同主题分析处理；业务应用层负责面向不同用户可视化展示，构建精

准化中医临床诊断、干预与评价模型，绘制中医临床知识图谱。

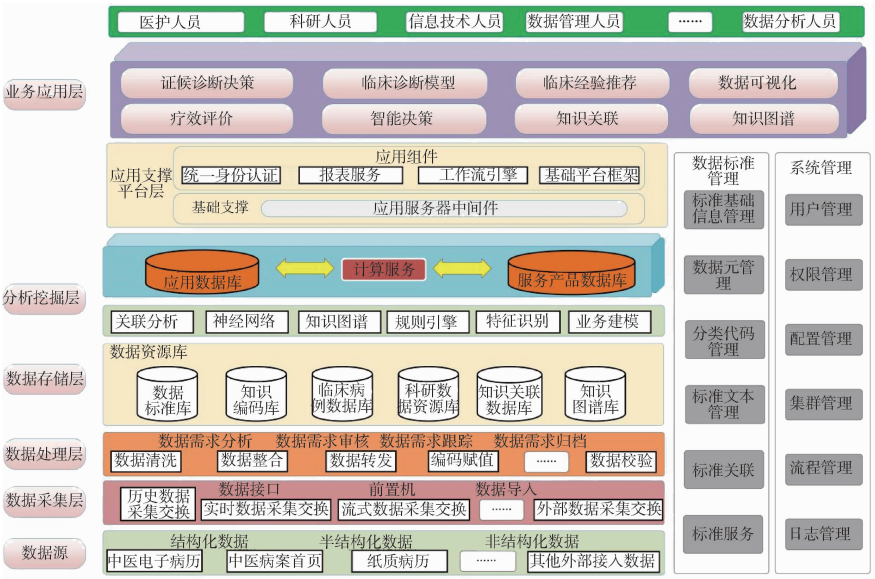

图 1 中医临床数据资源管理与服务平台总体架构

3.2 应用功能分析

3.2.1 *数据标准管理与服务* 中医临床数据标准是中医临床数据结构化、数据集成、转换编码与数据关联的基础支撑，主要包括中医名词术语、中医临床诊疗术语、中医病证分类与代码、临床中药编码、方剂编码及中医药信息数据元、中医电子病历基本数据集等标准。数据标准管理与服务模块主要包括数据元管理、分类代码管理、标准文本管理、标准关联、标准服务等子模块，可提供体现中医药特色的数据项名称、定义、表示格式、分类、允许值的管理功能，以及关联、查询、下载、分发等服务功能，为中医临床数据集成、数据规范化、数据编码化、数据知识化提供标准规范支撑。

3.2.2 *数据抽取集成* 数据抽取集成是中医临床数据资源库构建的基础和前提。中医临床数据内容主要涵盖患者基本信息、诊断信息、干预信息、疗效评价信息等，涉及医生工作站、护士工作站、中医电子病历系统、临床实验室系统、检查信息系统、医学影像系统、中医病案首页等多个业务系统或纸质病历等，具有多样性、多源性、异构性、真实性、复杂性等特性[7]。平台提供数据采集抽取范围确定、数据字段规范、接口对接、XML 解析、数据导入等功能，汇交融合多源异构的中医临床数据，形成中医临床原始数据库。若仅汇交单体医院的中医临床数据，可基于医院集成平台、临床数据中心采用直接自动抽取或通过 XML 转换器和抽取工具转换成 XML 文件，汇集到相应存储资源池。若汇交多家医院的中医临床数据，可通过 XML 文件、数据导入等功能和同步工具进行数据集成。同时支持前瞻性研究数据的集成，可建立实时数据采集接口或通过消息订阅方式动态采集数据。

3.2.3 *病例数据库生成* 平台提供 CRF 库结构（包括病例基本信息表、诊断信息表、干预信息表、疗效信息表等）设计、中医症状分词模型管理、质控规则管理、数据质量校验以及症状信息、中医病证、检查检验、中药处方、非药物疗法等数据处理与标引，采用人机结合模式对中医临床数据原始数据库的症状、病名证候、病因病机、辨证论治、方剂、中药、针灸推拿疗法等原始数据进行规范化、标准化与结构化处理，如对症状信息进行词性标注、分词、句子边界识别、句法分析、语义识别、指标识别等结构化处理，对中药药名进行规范化处理，将原始数据库转换成适于数据挖掘的 CRF 库。

3.2.4 *数据资源库生成* 平台提供专题模板管理、数据标准应用、编码规则管理、编码流程管

理、CRF 选取、编码赋值、数据质量校验、属性重构等功能，基于数据标准将 CRF 库数据赋予编码，形成具有多重属性的中医临床 RDR 库，将中医临床数据分析挖掘拓展到属性层，增加数据维度、信息量，以一种全新思维方式开展数据分析、探索中医临床客观规律。

3.2.5 数据分析与可视化 中医临床数据要素价值主要体现在数据可视化环节。平台集成报表开发工具、算法模型、多维分析工具等数据分析所需的通用工具包，中医临床术语处理、证候决策、方药决策等特殊工具，以及包括知识模型、自动推理、知识关联、知识识别、知识表达等关键技术，提供文本可视化、网络可视化、时空数据可视化、多维数据可视化等可视化方法和数据资源授权导出功能，导出 Excel、SPSS、SAS、MATLAB、R、Python 等统计分析软件兼容的文件格式，由中医临床医生、科研人员等在统一的环境下分析和挖掘中医临床潜在的隐性知识，帮助其直观找到数据资源间的关联性，绘制中医临床知识图谱。

3.3 数据安全防护体系

数据安全防护是中医临床大数据知识工程技术平台稳定运行的重要保障。全面贯彻《数据安全法》《网络安全法》《个人信息保护法》，落实医疗健康数据安全责任制度，实施中医临床诊疗数据的采集、标注、存储、传输、管理、应用等全生命周期安全管理，制定数据使用管理规定，建立数据提取、导入和导出的申请审批流程，严格管理患者信息、临床数据等，做好个人信息保护。在数据采集抽取集成时，采用去隐私、脱敏、匿名化、水印等方法对患者姓名、地址、电话等敏感信息进行转换处理，主动规避数据安全风险；存储与业务应用耦合度低的数据时，采用数据安全加密措施；分析与业务应用耦合度较高的数据时，实行分级分类管理和权限管控，逐级授权开放；数据备份可采用双机热备、分布式存储副本或纠删码方式、数据备份一体机等方式。同时，通过漏洞扫描、渗透测试、安全审计等技术摸排和评估安全风险，采用访问控制、数据库审计、行为审计等安全措施，解决数据使用场景的安全问题，定期开展实战演练，提升网络安全防护能力。

4 中医临床数据资源管理与服务平台价值体现

4.1 探索多源异构的中医临床数据集成

中医临床数据涉及患者入院记录、病程记录、会诊记录、医嘱信息、护理记录、手术记录、出院记录、病案首页等来源多样、结构不一、类型不同的信息[8]，实体数据汇集、交换、共享不畅，阻碍中医临床研究与发展。通过该平台构建，探索多源异构的中医临床数据抽取汇交策略、整合方案和质控机制，完成不同数据的采集、抽取、汇集、存储与关联，实现跨医院中医临床数据的交互共享，满足中医临床数据资源库构建及其管理服务的需要。

4.2 创新中医临床研究模式

相比传统的人工逐份查找病历或信息工程师协助查询获取整合中医临床诊疗数据，平台以真实世界数据为驱动，以中医临床数据标准为基础，通过语义识别、自然语言处理、数据归一、数据编码、关联分析、特征分析、神经网络等技术实现中医临床数据的抽取规范、整合集成与分析挖掘。数据工程师利用平台可进行中医临床数据汇集与知识编码，规范处理数据形成 CRF 库和 RDR 库；临床与科研人员利用平台可掌握研究数据指标与质量，自由选择研究指标或变量，激发科研灵感，使用提供的中医临床数据分析挖掘工具，构建精准化中医临床诊断、干预与评价模型，推进真实世界中医临床研究模式应用[9]。

4.3 推进中医药标准化建设

标准制定的目的在于应用。平台不仅管理中医临床数据标准，还将数据标准的应用融入信息系统研发、数据清洗、数据编码和分析挖掘，规范中医临床数据采集、处理与分析，建立中医临床数据结构化、业务流程规范和数据关联最佳秩序，是中医临床数据标准组织实施与推广应用的具体体现。在中医临床医生、科研人员、数据工程师、计算机工程师建设或使

用时，发现中医临床数据标准客观存在的问题，为改进和修订标准技术内容提供第一手资料。

4.4 促进中医医院智慧化建设

数据是智慧中医医院建设的核心，其质量是灵魂。通过中医临床数据资源管理与服务平台的构建与运行，梳理清楚中医临床数据资产状况，探索构建数据质量控制管理体系，为中医医院数据治理提供一套切实可行的方法库和技术平台，为数据资产管理、数据分类分级管理、数据共享与应用服务提供借鉴和经验，用数据促进中医医院智慧化建设。

5 结语

数据赋能中医药已成为中医药人共识，中医临床医生、科研人员、数据工程师、IT 人员、数据分析师等应主动拥抱数字技术，共同应用中医临床数据资源管理与服务平台汇集中医临床诊疗数据，构建 CRF 库、RDR 库等数据资源库，不断培育和挖掘面向中医临床诊疗服务、科学研究等场景的数据应用产品，创造数据价值，提供数据增值服务，构建适于发挥中医药特色优势的数字技术平台和方法体系。

参考文献

1 国务院．“十四五”数字经济发展规划［EB/OL］.［2022 - 12 - 12］. http://www.gov.cn/zhengce/zhengceku/2022 - 01/12/content_5667817.htm.

2 陈国青，吴刚，顾远东，等．管理决策情境下大数据驱动的研究和应用挑战——范式转变与研究方向［J］.管理科学学报，2018，21（7）：1 - 10.

3 肖勇，沈绍武，田双桂．我国中医医院信息化建设发展历程及展望［J］. 医学信息学杂志，2021，42（7）：2 - 6，11.

4 国务院办公厅．关于加快中医药特色发展的若干政策措施［EB/OL］.［2021 - 12 - 12］. http://www.gov.cn/zhengce/content/2021 - 02/09/content_5586278.htm.

5 中共中央，国务院．关于促进中医药传承创新发展的意见［EB/OL］.［2022 - 12 - 12］. http://www.gov.cn/zhengce/2019 - 10/26/content_5445336.htm.

6 宋灵恩．激发数据要素价值 赋能数字中国建设［EB/OL］.［2022 - 01 - 21］. http://www.cac.gov.cn/2022 - 01/21/c_1644368244622007.htm.

7 肖勇，陈伟，沈绍武．国家中医药数据中心构建方案探讨［J］. 时珍国医国药，2016，27（7）：1783 - 1785.

8 周敏，邓文萍，马红敏，等．基于 XML 的中医临床数据汇交方案研究［J］. 世界科学技术（中医药现代化），2014，16（8）：1805 - 1808.

9 刘保延．真实世界的中医临床科研范式［J］. 中医杂志，2013，54（6）：451 - 455.

2024 年《医学信息学杂志》征订启事

《医学信息学杂志》是国内医学信息领域创刊最早的医学信息学方面的国家级期刊。主管：国家卫生健康委员会；主办：中国医学科学院；承办：中国医学科学院医学信息研究所。中国科技核心期刊（中国科技论文统计源期刊），RCCSE 中国核心学术期刊（武汉大学中国科学评价研究中心，Research Center for Chinese Science Evaluation），美国《化学文摘》《乌利希期刊指南》及 WHO 西太区医学索引（WPRIM）收录，并收录于国内 3 大数据库。主要栏目：专论，医学信息技术，医学信息研究，医学信息资源管理与利用，医学信息教育等。读者对象：医学信息领域专家学者、管理者、实践者，高等院校相关专业的师生及广大医教研人员。

2024 年《医学信息学杂志》国内外公开发行，每册定价：15 元（月刊），全年 180 元。邮发代号：2 - 664，全国各地邮局均可订阅。也可到编辑部订购：北京市朝阳区雅宝路 3 号（100020）医科院信息所《医学信息学杂志》编辑部；电话：010 - 52328672，52328686，52328687。

《医学信息学杂志》编辑部