

中医药科学数据管理平台设计与实现^{*}

徐飞龙^{1,2} 吕佳^{1,2} 吴佳荣^{1,2} 曾宇平^{1,2} 傅昊阳^{1,2}

(¹ 广东省中医院 广州 510120 ² 广州中医药大学第二附属医院 广州 510120)

〔摘要〕 目的/意义 解决中医药行业科学数据所有者共享意愿较低的难题,推动中医药行业科学数据标准化汇聚、共享及应用。方法/过程 遵循国家科学数据汇交与管理相关标准,融合区块链、数字水印等技术,打造中医药科学数据汇交可信流程,构建中医药科学数据管理平台。结果/结论 中医药科学数据管理平台为区域中医药科学数据汇交提供信息化支撑,有效提升区域内科学数据汇交效率与汇交意愿。

〔关键词〕 中医药;科学数据;管理平台

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.12.013

Design and Implementation of the Scientific Data Management Platform of Traditional Chinese Medicine

XU Feilong^{1,2}, LYU Jia^{1,2}, WU Jiarong^{1,2}, ZENG Yuping^{1,2}, FU Haoyang^{1,2}

¹ Guangdong Provincial Hospital of Chinese Medicine, Guangzhou 510120, China; ² The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510120, China

〔Abstract〕 **Purpose/Significance** To address the challenge of low willingness to share scientific data among stakeholders in the traditional Chinese medicine (TCM) industry, and to promote standardization, aggregation, sharing, and application of scientific data in the field of TCM. **Method/Process** By adhering to national standards for scientific data submission and management, integrating technologies such as blockchain and digital watermarking, the study aims to establish a trusted process for the exchange of scientific data in the field of TCM and build a TCM scientific data management platform. **Result/Conclusion** This platform will provide information technology support for regional TCM scientific data exchange, and effectively improve the efficiency and willingness of scientific data exchange within the region.

〔Keywords〕 traditional Chinese medicine; scientific data; management platform

1 引言

数据被提升为新型生产要素,科学数据既是开展科学研究的基石,也是政府部门制订政策、分析决策的重要依据。美国国家科学理事会发布《长生

命周期的数字化数据库:21 世纪科研与教育的必备基础》,指出科学数据成为未来科研、教育模式创新的变革中心,必须予以长期保存和维护。我国 2018 年印发《科学数据管理方法》,为科学数据共享提供了制度保障。

科学数据共享中存在的数据安全、隐私泄漏、知识产权风险、数据权益保障缺失等问题导致科学数据共享动力不足、共享技术规范缺乏,以及权威科学数据服务平台数据覆盖面不全等缺陷,进而导致中国数据共享程度与开放科学先行国家相比存在一定差距,中医药领域面临同样困境。如何在政策

〔修回日期〕 2023-07-26

〔作者简介〕 徐飞龙,高级工程师,发表论文 5 篇;通信作者:曾宇平,高级工程师。

〔基金项目〕 广东省科技计划项目(项目编号:2021B121200007)。

驱动基础上，借助技术手段保障中医药科学数据共享过程中的数据防篡改、合规使用、隐私保护、溯源确权，推动中医药行业科学数据标准化汇聚、共享及应用，成为行业科学数据共享亟待解决的重要问题。广东省中医院将区块链信息溯源、确权、防篡改、隐私数据保护技术能力及数字水印技术融入中医药科学数据管理信息化业务流程，构建广东省中医药科学数据管理平台，有效支撑区域中医药科学数据汇交，提升区域中医药行业科研机构及人员数据共享意愿。本文从平台架构设计、技术选型、功能实现等方面详细阐述中医药科学数据管理平台建设经验，为同类研究提供参考。

2 相关概念

2.1 中医药科学数据

借鉴《科学数据管理办法》中对科学数据的定义，中医药科学数据是指在中医药领域，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中医药科学研究活动的原始数据及其衍生数据。中医药科学数据资源涉及中医、中药、中医适宜技术（针灸、推拿等）、中医药事业管理等内容^[1]。

2.2 科学数据共享

科学数据共享对于减少数据重新收集成本、提高科学研究效率具有重要意义^[2]。科学数据具有长期保存、完整性、安全性需求，信息技术在科学数据共享中影响深远，结合已有理论与实践，科学数据共享主流实践模式是综合运用现代信息技术将数据资源集中、电子化、权威存储，并遵循安全可控、分级共享以及国际上流行的可发现、可访问、可互操作、可重用（findable、accessible、interoperable、reusable，FAIR）原则推动数据合理流动、配置。

2.3 中医药科学数据管理平台

中医药科学数据管理平台是指支持中医药科学数据汇交与采集、保存、共享与利用、更新与撤销等功能的计算机程序。

3 中医药科学数据管理平台设计与实现

3.1 中医药科学数据管理平台总体技术架构设计

遵循页面、数据、服务、应用分离的架构设计原则，采用面向服务的体系架构保持系统灵活性和扩展性，总体架构，见图1。

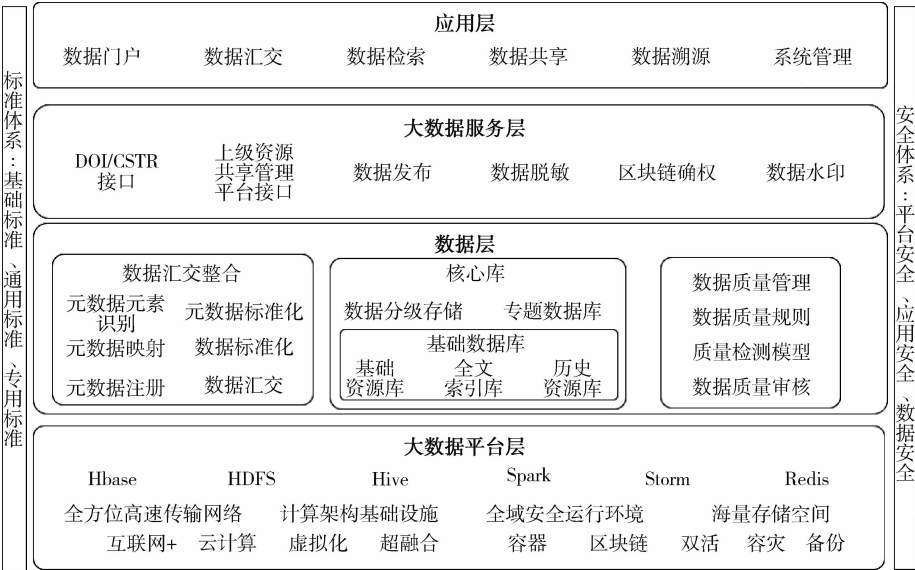
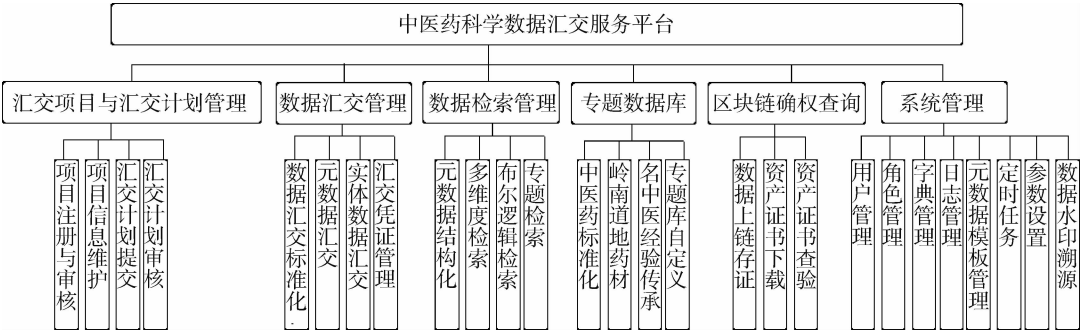


图1 中医药科学数据管理平台总体技术架构

前端技术采用 Vue、Element，后端采用 Java 语言，以 Spring Boot 结合 Spring Security 为基础的应用总体框架。底层数据存储采用 MySQL，Redis 作为数据缓存，Elasticsearch 作为平台快速搜索引擎，采用容器微服务架构设计，支持应用快速部署及启动。数据安全层面融合应用区块链、数据水印、数据加密等技术提升系统数据确权、数据溯源能力。平台应用消息总线集成汇交项目管理、汇交计划管理、数据汇交、质量审核、区块链确权、后台管理等流程。

3.2 中医药科学数据管理平台功能实现

面向国家、广东省大力发展基础科学研究战略部署，以中医、中药学科领域为重点，汇聚区域中医药基础科学数据资源，突出数据标准化汇聚和专题化服务。因此，平台核心功能包括数据汇交项目与汇交计划管理、数据汇交管理、数据检索、专题数据库服务、区块链确权查询、系统管理等。中医药科学数据管理平台功能架构，见图 2。



3.2.1 汇交项目与汇交计划管理 提供汇交项目注册、审核及信息维护等功能。创建汇交项目：平台建立科学数据汇交多级质控机制，用户在平台注册汇交项目，项目负责人提交项目相关证明信息。提供汇交计划提交及审核功能。有支撑项目可按照实际情况进行汇交计划提交，并按照计划提交科学数据。

3.2.2 数据汇交管理 数据汇交平台标准化程度、系统原型直接影响科学数据的可获取性。元数据作为关于数据的数据，用于说明数据的内容、质量、生成、加工、访问和获取方式以及其他有关特征。平台按照不同数据汇交专题，制订不同数据汇交标准，数据汇交人员上传科学数据元数据及实体数据，同时上传数据质量保证书，平台按照不同专题将汇交数据审核任务自动或手动分配给对应数据质控专家，数据质控专家登录平台即可按照数据质控规则进行数据审核，审核通过则生成数据汇交凭证和数据资产证书。

数据质量是发挥数据价值的前提。平台依据各

专题数据在数据库落地时的质量情况及数据质量规则类型，制订各专题数据质量的规范和标准。以名老中医经验传承专题为例，医案数据拥有者上传数据时，按照平台数据汇交标准汇交实体医案以及医案摘要数据。其中实体医案数据包括基本资料、发病过程、首诊证候、辨证论治、随访过程、按语等过程数据。医案摘要数据部分包括是否有项目来源、项目信息、数据名称、科技资源标识符、数字对象唯一标识符（digital object unique identifier, DOI）、学科分类、科学数据分类、数据集摘要描述、关键词、共享范围、共享途径、数据生成日期、数据发布日期、数据保护期、资源类型、数据格式、产生方式、是否包含隐私信息、是否涉及伦理问题、上传数据类型、数据质量描述、数据使用说明、所属专科、所属流派、主要临床表现、西医诊断、中医诊断、中医证型、治则治法、方剂、医师姓名、名医类型、中医医疗技术、中医药调护、综合疗效等字段，平台遵循有标采标、无标根据实际情况制订原则，完成上述字段基础数据标准化。

医案数据标准化保障了汇交数据质量,也为后续医案数据挖掘与利用奠定基础。为提升数据汇交效率,平台开发设计数据批量汇交功能,支持网页填报和批量上传两种方式。用户可按照模板批量填报,避免页面填报烦琐。

3.2.3 数据检索管理 参照符合国际标准的开放档案信息系统、PERMIS、ISO16363 以及《GB/T 30522—2014 科技平台 元数据标准化基本原则与方法》等,遵循真实性、完整性、可靠性、长期可解释性原则^[3],基于 Elasticsearch 对元数据进行标准化、结构化存储,支持用户按照数据集名称、数据集描述、数据集关键词、数据资源创建者、数据集英文名称、DOI 标识、主题、学科分类、发表年度、汇交单位、科学数据名称、科学数据关键字、科学数据摘要等关键信息检索。平台采用分布式全文搜索引擎架构,支持实时分析,提升用户数据检索效能并保障平台性能可扩展性。

3.2.4 专题数据库 遵循基础优先、体现中医特色原则,设名老中医经验传承、岭南道地药材、中医药标准化 3 大专题库,按照规范对汇交数据进行加工、整理、入库,同时构建必要的数​​据应用环境,对外提供统一数据访问服务。为保证科学数据可解释性,中医药科学数据汇交须同时汇交元数据和科学数据实体,二者通过唯一标识符逻辑关联。考虑到元数据与数据文件在数据类型、文件格式、编码方式等方面的区别,元数据存储的平台全文索引库,实体数据文件存储在平台关系型专题数据库。

3.2.5 区块链确权查询 平台对接第三方数据区块链可信系统,通过令牌方式交互,支持用数据资源存证、数据资源交易存证、数据资源编号查询存证资源,以及用链上交易号查询链上信息。目前平台实现科学数据存证用户、科学数据名称、实体数据文件名称、实体数据文件哈希值等确权必要数据上链,并生成汇交凭证和数据资产证书,其中数据实体文件不上链,数据汇交人员可基于资产证书编码查询数据真实性。数据共享、数据下载记录信息也进行链上存储,基于区块链可追溯、防篡改特性实现科学数据汇交、共享下载等关键业务节点留

痕、追溯。同时,平台预留数据交易功能接口,后续将探索科学数据线上、线下交易联动服务。

3.2.6 系统管理 平台建立基于角色访问控制权的控制体系。提供用户管理、角色管理、菜单管理、单位管理、字典管理、参数设置、日志管理、基础数据管理、标准数据元管理等功能。后台提供项目管理配置、数据质量检验、元数据模板配置、数据脱敏、数据水印、数据溯源、数据共享权限设置、接口管理等个性化功能。

4 平台创新应用

目前,数据安全风险、隐私泄漏、知识产权风险、数据权益保障缺失等问题严重影响了科学数据的共享动力。平台引入定制化开发开源区块链底链+国密算法的数据可信方案,为科学数据上链、确权并颁发基于区块链的数据资产证书,在政策驱动基础上通过技术功能创新探索科学数据确权和共享的新模式,有力提升了数据生产者数据汇交动力。

数据安全是科学数据共享的前提。考虑到科学数据汇交场景中多为静态数据,平台在不改变业务数据有效性的基础上,通过对原数据添加伪行、伪列,对原始敏感数据脱敏并植入标记等方式进行水印处理,可识别数据分发者、分发对象、分发时间、分发目的等元素,同时保留目标环境业务所需的数据特性,支撑科学数据规范分发,降低数据共享风险责任,提供数据版权归属保护。

平台建设过程从元数据、文件命名、可信赖存储、开放共享、格式管理、可持续保存 6 个方面^[4]进行需求分析及平台功能设计,为中医药科学数据汇交、共享提供技术支撑与平台保障。目前平台 3 个专题库已实现对应科学数据汇交及共享。名老中医经验传承专题库,实现了国医大师、全国名中医、省名中医等各类名中医 2 000 余份典型医案、7 000 余份脱敏病历、传承谱系数据标准化汇交、共享;岭南道地药材专题库,收录化橘红、广陈皮、阳春砂、广藿香、巴戟天、沉香、广佛手、何首乌、肉桂、穿心莲、高良姜、毛冬青 12 种岭南道地中药材及特色品种药材多维度科学数据,包括

药材基源数据、产地信息数据库、加工信息数据、质量信息数据、基因数据、化合物数据、药材评价数据。中医药标准化专题库, 汇聚国内外中医药标准数据 500 余条。

目前平台已正式上线运行 (<https://service.gdcmsdata.cn/mall/home>), 成为中医药领域权威、公开、可持续运营的科学数据服务平台, 数据链接有效性超过 99%。平台自上线试运行以来, 注册用户超过 1 000 名, 汇交数据用户超过 50 名, 涉及区域内多家中医药科研、医疗机构。

5 结语

数字化时代, 科学数据纳入信息化科研环境的基础设施范畴已成共识。中医药科学数据成为中医药科学研究、科技创新和国家发展的重要战略性资源。中医药科学数据汇交服务平台后续将进一步完善运行管理机制, 不断完善数据服务功能, 联合学术搜索引擎^[5], 探索线上线下联动科学数据交易,

建立为科研服务的数据管理价值链^[6]。

参考文献

- 1 中国国家标准化管理委员会. GB/T 38327—2019 健康信息学 中医药数据集分类 [S]. 北京: 中国质检出版社, 2019.
- 2 刘莉, 刘文云. 基于解释结构模型的科研数据共享影响因素分析 [J]. 情报科学, 2020, 38 (5): 27–33.
- 3 杨晨柳, 方安, 钱庆, 等. 面向人口健康科学数据长期保存的元数据模型构建 [J]. 中华医学图书情报杂志, 2020, 29 (9): 22–28.
- 4 杨晨柳, 方安, 胡佳慧, 等. 人口健康科学数据长期保存系统建设路径探析 [J]. 医学信息学杂志, 2021, 42 (9): 11–16, 51.
- 5 袁红卫, 黄松, 刘嫣. 麻省理工学院科学数据管理与共享平台调研及启示 [J]. 图书馆学研究, 2019 (13): 95–101.
- 6 崔旭, 赵希梅, 王铮, 等. 我国科学数据管理平台建设成就、缺失、对策及趋势分析——基于国内外比较视角 [J]. 图书情报工作, 2019, 63 (9): 21–30.
- 20 JAVANMARDI M, NOROOZI M, MOSTAFAVI F, et al. Exploring the motivations of pregnant women to seek health information; a qualitative study [J]. Iranian journal of nursing and midwifery research, 2022, 27 (5): 446–451.
- 21 SHARIFI M, AMIRI – FARAHANI L, HAGHANI S, et al. Information needs during pregnancy and its associated factors in afghan pregnant migrant women in Iran [J]. Journal of primary care & community health, 2020, 11 (1): 921915683.
- 22 BALASOIU A M, POMANA C, SIMA R, et al. The influence of educational, environmental and social factors in the decision to attend prenatal lectures among pregnant women in romania [J]. Maedica, 2021, 16 (4): 574–579.
- 23 贺丹, 朱焱, 卢芸, 等. 贵安新区农村孕产妇母婴健康素养及健康教育需求调查 [J]. 预防医学, 2022, 34 (12): 1283–1288.
- 24 安晶晶, 徐丽. 高危孕妇孕期保健知识认知度及健康教育需求情况的调查 [J]. 中国妇幼保健, 2018, 33 (4): 741–744.
- 25 LIU Y Q, YU Y, BAI J B, et al. Development and psychometric properties of the maternal health needs scale in Chinese maternal women [EB/OL]. [2022–12–18]. <https://pubmed.ncbi.nlm.nih.gov/31805429/>.
- 26 洪梅, 姚引娣, 田潇, 等. 孕妇健康需求水平及影响因素研究 [J]. 湖北医药学院学报, 2021, 40 (4): 399–403.
- 27 吴晓兵, 程晓芸, 裴素云, 等. 北京市东城区 976 例孕产妇健康信息需求调查 [J]. 护理研究, 2020, 34 (20): 3715–3718.
- 28 LU Y, BARRETT L A, LIN R Z, et al. Understanding information needs and barriers to accessing health information across all stages of pregnancy: systematic review [J]. JMIR pediatrics and parenting, 2022, 5 (1): e32235.
- 29 杨姮, 郭红, 徐杨, 等. 职业女性母乳喂养体验质性研究的 Meta 整合 [J]. 中华现代护理杂志, 2021, 27 (28): 3797–3803.
- 30 KASSIM M. A qualitative study of the maternal health information – seeking behaviour of women of reproductive age in Mpwapwa district, Tanzania [J]. Health information and libraries journal, 2021, 38 (3): 182–193.
- 31 HSIAO S, HSU H, ERICKSON S R, et al. Seeking health information online among U. S. pregnant women: findings from the 2009–2018 national health interview surveys [J]. Women health, 2023, 63 (1): 59–69.

(上接第 64 页)