

上海市创新药产业全链条数据共享平台架构设计与构建策略研究

冯瀛尹^{1,2} 唐红梅^{1,3} 张世翔^{1,3} 黄天翔⁴

(¹ 上海中医药大学研究生院 上海 201203 ² 上海健康医学院健康与公共卫生学院 上海 201318

³ 上海健康医学院护理与健康管理学院 上海 201318 ⁴ 上海市普陀区疾病预防控制中心应急办 上海 200333)

〔摘要〕 **目的/意义** 探讨上海市创新药产业全链条高质量发展的数据共享平台构建策略,以推动创新药产业的数字化转型和高质量发展。**方法/过程** 分析上海市创新药产业数据管理的现存问题,提出涵盖基础设施层、技术支撑层、应用模块层和用户交互层的全链条数据共享平台架构。**结果/结论** 该数据共享平台能够系统性解决数据脱节问题,实现从药物研发到生产再到临床使用的全生命周期数据闭环管理,提升创新药产业的效率和质量,为个性化医疗提供数据支撑。

〔关键词〕 创新药产业;全链条;数据共享平台

〔中图分类号〕 R-058 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2025.10.012

Study on the Architectural Design and Construction Strategies of a Full-chain Data-sharing Platform for Shanghai's Innovative Drug Industry

FENG Yingyin^{1,2}, TANG Hongmei^{1,3}, ZHANG Shixiang^{1,3}, HUANG Tianxiang^{1,2}

¹ Graduate School, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China; ² School of Health and Public Health, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China; ³ School of Nursing and Health Management, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China; ⁴ Emergency Office, Shanghai Putuo District Center for Disease Control and Prevention, Shanghai 200333, China

〔Abstract〕 **Purpose/Significance** To explore the strategies for constructing a data-sharing platform for the full-chain of high-quality development of Shanghai's innovative drug industry, so as to promote the digital transformation and high-quality development of the innovative drug industry. **Method/Process** The existing problems of data management in Shanghai's innovative drug industry are analyzed. A comprehensive full-chain data-sharing platform architecture is proposed, covering the infrastructure layer, technical support layer, application module layer, and user interaction layer. **Result/Conclusion** The construction of this data-sharing platform can systematically address the data disconnection issues, achieve closed-loop management of the entire lifecycle data from drug research and development to production and clinical use. It will enhance the efficiency and quality of the innovative drug industry and provide data support for personalized medicine.

〔Keywords〕 innovative drug industry; full-chain; data-sharing platform

〔修回日期〕 2025-09-06

〔作者简介〕 冯瀛尹,硕士研究生,发表论文10篇;通信作者:唐红梅,研究员。

〔基金项目〕 上海市卫生健康委员会卫生健康政策研究课题(项目编号:2025HP24)。

1 引言

全球正经历以大数据、人工智能 (artificial intelligence, AI)、区块链等技术为主要推动力量的第 4 次工业革命^[1], 为创新药产业发展带来了前所未有的机遇。研究^[2]显示, 开发一种药物大约需要 8~10 年、投资 30 亿美元。部分跨国制药企业已构建起从研发、生产到商业化应用的数字化平台^[3-4], 以提高临床试验数据的可及性。上海市近年来在创新药产业数据共享方面已有一定政策基础, 如 2024 年 7 月上海市人民政府办公厅发布的《关于支持生物医药产业全链条创新发展的若干意见》^[5], 指出要释放产业数据资源价值。上海市拥有张江药谷等世界级产业集群, 截至 2025 年, 人工智能产业规模达 4 000 亿元^[6]。然而, 现有平台建设多聚焦单一环节优化, 尚未形成贯穿“研-产-用”, 即“药物研发-生产工艺参数-临床使用”的全链条数据赋能体系。本文以上海市为例, 分析创新药产业全链条数据共享平台的总体架构、模块设计及协同机制, 并提出构建策略。

2 创新药产业全链条数据管理现存问题

2.1 研发数据的管理困境

在创新药研发过程中, 临床试验数据作为连接基础研究与临床应用的核心纽带^[7], 其整合效率直接影响药物研发周期与成功率。目前研发数据主要面临标准化缺失与动态管理机制薄弱的困境。临床研究机构使用电子病历系统, 合同研究组织 (contract research organization, CRO) 使用电子数据采集系统, 制药企业内部则搭建专有数据仓库, 这些系统在数据格式、元数据定义、编码规则等方面存在显著差异。国家药品监督管理局已发布《临床试验数据管理工作技术指南》《药物临床试验数据递交指导原则 (试行)》等文件鼓励使用临床数据交换标准协会 (Clinical Data Interchange Standards Consortium, CDISC) 标准, 但由于缺乏通用软件^[8], 且各单位现用标准有所不

同^[9], CDISC 标准落地困难。临床试验数据从方案设计、入组筛选到疗效评价, 是一个动态过程。现有系统普遍缺乏实时更新与版本控制能力^[10], 使临床试验数据无法及时反映实际情况, 增加药物研发风险。

2.2 生产数据的封闭性

在创新药产业中, 生产工艺参数与质量控制数据是连接研发设计与规模化生产的核心纽带, 其数据协同水平直接决定药品质量稳定性和产业转化效率。研发实验过程形成的各项参数条件与规模化生产线的生产数据并不匹配, 很难将这些制药技术和模型成功放大至生产线。生产工艺参数是企业的核心无形资产, 大部分制药企业采用封闭式数据分析利用流程^[11]。与此同时, 产业全链条协同对制药企业的数据共享提出了现实要求, “封闭”与“共享”在制药企业、CRO 及原料供应商等多主体间持续拉锯形成矛盾。

2.3 使用数据的滞后性

药品上市后的真实世界疗效数据, 是优化药物研发、改进生产工艺的核心依据。然而, 上海创新药产业中, 临床使用端与研发生产端的数据反馈链条存在割裂^[12]。临床使用端产生的电子病历、随访记录等海量数据堆积在医疗机构内部, 如同被堵塞的河道形成“数据湖”, 下游的研发和生产端因缺乏数据输入, 陷入“干旱”状态, 阻碍了产品迭代。

3 “研-产-用”全链条数据共享平台架构

3.1 总体架构

基于上述数据管理问题, 提出分层解耦式数据共享平台架构。核心设计原则为数据可信流通、价值闭环驱动、安全合规可控, 以推动全链条数据共享的正向循环。平台总体架构分为 4 个层级, 分别为基础设施层、技术支撑层、应用模块层、用户交互层, 体现“需求驱动-能力支撑-资源承载”的递进关系, 见图 1。

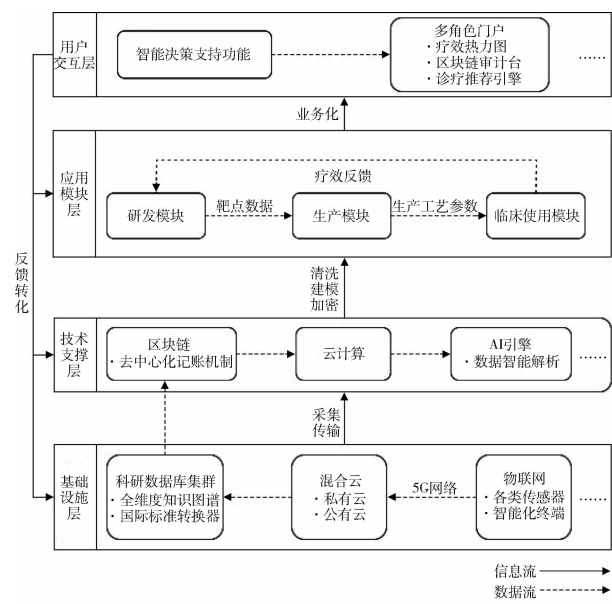


图1 “研-产-用”全链条数据共享平台总体架构

3.1.1 基础设施层 基础设施层作为数据共享平台的物理基座，整合医疗机构、制药企业、CRO、监管机构等多方数据源，收集、传输、储存临床数据、研发及生产数据、试验数据、审评数据等，依托分布式节点网络实现高效、安全的数据共享。首先，通过物联网设备矩阵构建数据采集的“神经末梢”，由各种传感器与智能产品收集不同类型生产、研究以及使用过程中的数据。利用 5G 网络作为数据流动的“超级动脉”，在临床试验、生产监控、远程医疗等场景中，保证数据的即时性和准确性。其次，利用由私有云和公有云组合而成的混合云进行数据储存，保障数据“安全-弹性”特性^[13]。其中，私有云用于存储如细胞株构建参数、患者基因序列等核心敏感数据；公有云用于处理药品说明书、通用生产工艺模板等非敏感数据，可灵活进行规模扩展或收缩，以应对不断变化的数据量并节省成本。最后，科研数据库汇聚全球权威生物医药数据库数据及本地数据，内置 CDISC 国际标准转换器，将异构数据转化为统一模型，通过统一的应用程序接口进行数据调用。

3.1.2 技术支撑层 技术支撑层作为数据共享平台的关键“智能工厂”，整合区块链、AI 引擎和云计算等创新科技，构建高效能信息处理与防护体

系，推动数据资源的充分开发和安全流转，将原始信息转化为智慧决策依据^[14]。区块链依靠去中心化记账机制，可细致追踪信息的采集、传输以及应用轨迹，构建完整的生命周期追溯体系。该特性可切实保护专利技术、临床研究资料以及个人隐私等敏感内容，又可契合监管审查和异常排查等要求，为全链条数据共享提供信任支撑。在此过程中，云计算可为 AI 引擎提供灵活且可扩展的运算能力。

3.1.3 应用模块层 应用模块层是平台实现业务价值转化的核心载体，将底层技术能力封装为面向研发、生产、临床使用场景的智能化工具，驱动全链条数据高效协同与闭环迭代。其核心包括研发模块、生产模块和临床使用模块，模块间的协同逻辑与数据闭环，见图 2。一是研发模块与生产模块间的正向数据流。研发阶段利用 AI 模型进行分子设计和虚拟试验，生成关键参数和试验结果；通过区块链技术加密数据；传输到生产模块的数字孪生系统，系统自动适配这些参数并优化生产工艺。二是生产模块与临床使用模块间的实时控制流。通过数字孪生系统实时监控生产批次质量属性，若发现偏差，系统能自动关联并及时追踪使用该批次药物的患者；临床使用模块根据实时数据触发监测方案，对患者的响应数据进行实时监测；将监测到的患者响应数据回流到生产模块，优化生产工艺，形成实时反馈闭环。三是临床使用模块与研发模块间的逆向反馈流。利用联邦学习技术分析临床使用模块的真实世界数据，保护患者隐私的同时提取有价值的信息^[15]；通过 AI 技术对联邦学习的结果进行关联分析，将分析结果反馈给研发模块，为新一代药物设计提供依据^[16]，形成从临床到研发的逆向反馈闭环。

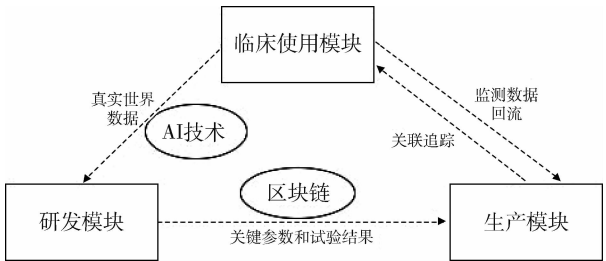


图2 应用模块层核心模块数据流

3.1.4 用户交互层 用户交互层作为数据共享平台的智能终端,通过智能决策支持功能和多角色门户,将专业技术转变为直观且便于使用的业务功能。智能决策支持功能通过数据驱动的决策工具,帮助用户从海量数据中提取有效洞见,建立从数据到执行的快速响应机制。通过多角色门户针对不同使用者给出差异化服务方案,保证医疗机构、制药企业、监管机构以及患者均可高效运用系统资源^[17],实现数据价值普惠与跨主体高效协同,进一步促进数据在不同主体间的流动。

3.2 模块设计 (图3)

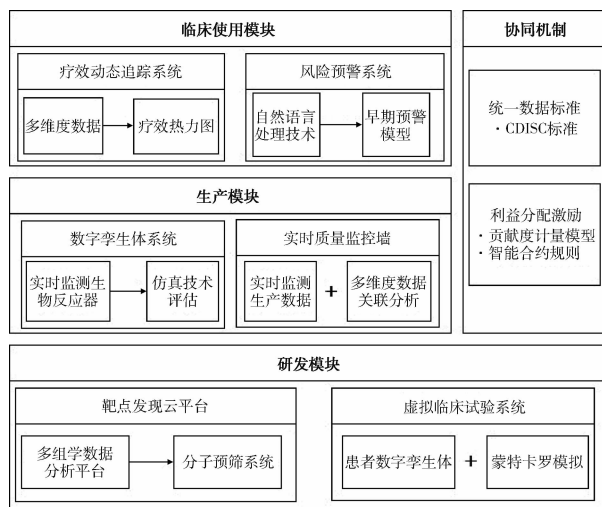


图3 “研-产-用”模块设计

3.2.1 研发模块 研发模块的主要目标在于加快药物研发速度,提升研发效率以及成功率。借助云计算的多组学数据分析平台整合基因组学、蛋白质组学以及代谢组学等多方面信息,帮助识别靶点;通过分子预筛系统从大量化合物中快速找出可能有生物活性的分子结构^[18-19]。在临床研究优化方面,模块中的虚拟临床试验系统依靠患者数字孪生体和蒙特卡罗模拟,为试验方案设计提供科学依据,避免无效尝试^[20]。其中,患者数字孪生体结合临床诊疗记录和遗传信息,可在试验规划阶段精准预测受试者应答情况;蒙特卡罗模拟能系统评估不同给药策略的疗效与风险特征,为确定最优治疗方案提供数据支持。

3.2.2 生产模块 生产模块主要包括数字孪生体系统和实时质量监控墙。借助高频率的数据采集,数字孪生体系统可实时监测生物反应器的各项运行参数,凭借仿真技术预估生产工艺参数对产品质量的影响,并自动生成优化建议,为生产操作提供指导。实时质量监控墙借助对生产数据的实时监测,依靠多维度数据关联分析,精准确定影响产品质量的关键因素,优化生产过程。

3.2.3 临床使用模块 临床使用模块主要由疗效动态追踪系统和风险预警系统构成。疗效动态追踪系统融合电子病历、可穿戴设备、医保支付等多维度数据,依据患者遗传特征、疾病分期等个体差异,生成不同群体的疗效热力图,并将实时临床观察结果同步反馈至药物开发系统。风险预警系统运用自然语言处理技术解析非结构化病历,识别关键临床指征,并依靠早期预警模型,及时发现潜在不良反应,保障临床用药安全。

3.3 协同机制

协同机制是“研-产-用”全链条数据共享平台实现跨环节数据闭环与多主体价值共生的核心规则体系。其一,统一的数据标准推动破解“语言不通”难题。例如,要求临床数据全部采用 CDISC 标准,标准化的数据格式和高质量的数据管理能够加速数据的共享和利用。同时,动态扩展机制使制药企业等相关单位能够根据自身需求自定义扩展字段。这种机制支持个性化需求,同时能够保持数据标准的统一性。其二,设置适当的利益分配激励,构建产业各方的共赢生态。通过贡献度计量模型,在数据维度,以病例数、完整性评分、更新频率等对贡献数据的数量、质量、时效性进行量化;在算法维度,通过衡量贡献的数据使 AI 模型预测精度提升的百分比,对相关单位的模型贡献度进行量化;在算力维度,通过衡量计算资源、存储资源的使用情况,对图形处理器使用小时数、存储空间占用率进行量化。通过量化手段激励各方提供高质量、高时效性的数据,提升数据共享的积极性。此外,通过智能合约规则,使医疗机构在提供数据的同时获得制药企业研发成果的优先使用权,制药企

业在开放数据的同时能够获得更多经济收益,患者在授权数据的同时能够获得个性化的用药建议和保险优惠,以此促进医疗机构与制药企业之间、各制药企业之间、患者与医疗机构之间的数据共享与合作,提升整个数据共享平台的协同效应。

4 平台构建策略

4.1 政策支持与保障机制

为促进数据共享的规范化与高效化,相关部门应加快制定支持政策。明确数据的所有权、使用权及数据共享授权机制,如数据所有权归属于数据产生方^[21],使用权可在合法合规的框架内授权给制药企业、科研机构等其他主体。同时,相关部门可采取多元化调控措施激发市场主体参与热情,如通过财政专项拨款支持数据共享基础设施的搭建与维护;构建激励机制,对主动参与数据协同的企业及机构给予一定程度的政策倾斜。

4.2 基础设施建设

通过构建先进的技术体系和基础设施,能够有效缩减中小型企业数据获取方面的投入,推动全行业数字化升级与创新效能增强。可由主管部门主导建立区域性医疗健康数据交换中心,汇聚药物研发记录、制造过程参数、诊疗实践信息等多元化数据资源,构建集中式数据存储库^[22]。进一步采用数据净化、格式规范等处理方法提升数据集质量,保障数据资源的实用价值与可信程度。

4.3 运维人才保障

全链条数据共享平台需要复合型运维人才团队,包括具有药学背景的数据专家、隐私计算工程师、区块链运维工程师等。其中,具有药学背景的数据专家的核心职能包括解析试验数据的临床意义、分析 AI 给出的预测结果及优化策略等;隐私计算工程师主要负责部署联邦学习系统及数据动态脱敏;区块链运维工程师主要负责联盟链节点维护及智能合约审计。同时,要建立可持续的人才保障机制,可通过产学研融合计划,如高校联合张江药

谷等产业基地,形成真实场景运维轮岗基地,缩短运维人才的适应时间。

4.4 多方协作与生态培育

联合高校、科研机构、医疗机构和制药企业等各方力量,共同构建“产学研医”联盟。设立专项合作项目,为联盟成员之间的联合研究和创新项目提供支持,促进“产学研医”深度融合。在此基础上,引入商业保险机制和金融资本,为数据共享平台提供风险保障和资金支持。同时,应加强跨部门协同与合作,成立协调小组负责统筹协调各部门之间的数据共享,及时解决数据共享过程中出现的问题。

5 结语

在全球医药行业数字化进程不断深化的背景下,数据共享已成为推动创新药产业高质量发展的关键要素。本文以上海市为例,深入分析了创新药产业全链条数据共享平台的架构设计与构建策略。该平台的建立将有效弥合各阶段数据断层,消除不同业务环节间的信息壁垒,实现从药物研发到生产再到临床使用的全生命周期数据闭环管理。不仅能够提升创新药产业的效率和质量,也能为个性化医疗提供坚实的数据支撑。

作者贡献: 冯瀛尹负责论文撰写与修订;唐红梅负责研究设计;张世翔负责提出研究问题、协调调研工作;黄天翔负责资料收集与分析。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 覃璇. 大数据视域下第四次工业革命的历史定位与实践启示 [J]. 社会科学辑刊, 2025 (3): 57-64.
- VIDIYALA N, SUNKISHALA P, PARUPATHI P, et al. The role of artificial intelligence in drug discovery and pharmaceutical development: a paradigm shift in the history of pharmaceutical industries [J]. AAPS pharmscitech, 2025, 26 (5): 133.
- 林少楠. 深度解析, 罗氏在 DCT 的落地情况 [EB/OL]. [2025-05-22]. <https://mp.weixin.qq.com/s?>

- _ _ biz = MzU4NDgyMzUxNQ = = &mid = 2247493493&idx = 1&sn = 8e9881f66f57dbc956d69a4accb37982.
- 4 高康平. 诺华数字化创新解析: 真正的“药神”, AI、大数据、数字药物的先行者 [EB/OL]. [2024 - 08 - 23]. https://www.sohu.com/a/249521630_133140.
- 5 上海市人民政府办公厅. 关于支持生物医药产业全链条创新发展的若干意见 [EB/OL]. [2025 - 06 - 04]. <https://www.shanghai.gov.cn/nw12344/20240730/0fe29fc2246e4b478757dee3a01ccd08.html>.
- 6 上海市经济和信息化委员会. 上海率先完成人工智能产业规模 2025 年目标 [EB/OL]. [2025 - 05 - 22]. <https://www.jfdai.ly.com/sgh/detail?id=1512168>.
- 7 彭朋, 杨妍, 元唯安, 等. 信息化对药物临床试验质量影响的实践及分析 [J]. 药物评价研究, 2023, 46 (3): 478 - 482.
- 8 王泽, 陈浩, 陈佳齐, 等. 临床数据交换标准协会体系及其应用现状 [J]. 中国公共卫生, 2024, 40 (2): 227 - 233.
- 9 马琳, 邓宏勇. 国内外临床研究元数据标准: 特点及现状 [J]. 中国循证医学杂志, 2023, 23 (4): 478 - 484.
- 10 张钟元, 赵志霞, 王欣, 等. 临床试验机构评价信息平台功能设计 [J]. 医学信息学杂志, 2025, 46 (2): 81 - 86.
- 11 丁静, 茅鸯对. 大数据战略下医药产业高质量发展协同共治路径研究——以浙江省医药产业数字化转型与监管为视角 [J]. 中国现代应用药学, 2022, 39 (21): 2732 - 2737.
- 12 沈丽宁, 满亚荣, 陈思思, 等. 健康数据空间框架构建与推进策略研究 [J]. 医学信息学杂志, 2025, 46 (3): 8 - 14.
- 13 林兵, 汪昕隆, 苏明辉, 等. 混合云中面向多目标的工作流数据放置策略 [J]. 计算机集成制造系统, 2025, 31 (1): 219 - 234.
- 14 ALSALEM M A, ALAMOUDI A H, ALBAHRI O S, et al. Evaluation of trustworthy artificial intelligent healthcare applications using multi - criteria decision - making approach [J]. Expert systems with applications, 2024, 246 (7): 1 - 15.
- 15 刘振涛, 李涵, 吴浪, 等. 基于联邦学习的医疗数据共享与隐私保护 [J]. 计算机工程与设计, 2024, 45 (9): 2577 - 2583.
- 16 ARMAIGNAC D L, HEAVNER S F, RAUSEN M, et al. Guiding principles for data sharing and harmonization: results of a systematic review and modified Delphi from the society of critical care medicine data science campaign [J]. Critical care medicine, 2025, 53 (3): e619 - e631.
- 17 YANG Y L, LIU M H, WANG L, et al. A ring signature scheme with linkability and traceability for blockchain - based medical data sharing system [J]. Peer - to - peer networking and applications, 2025, 18 (3): 103.
- 18 YILDIRIM Z, SWANSON K, WU X K, et al. Next - gen therapeutics: pioneering drug discovery with iPSCs, genomics, AI, and clinical trials in a dish [J]. Annual review of pharmacology and toxicology, 2025, 65 (1): 71 - 90.
- 19 王超, 肖辅, 李妙竹, 等. 人工智能在药物靶点的筛选及验证方面的应用进展 [J]. 中国药科大学学报, 2023, 54 (3): 269 - 281.
- 20 孟海华, 孙立, 陈明伟, 等. 虚拟临床试验的发展趋势及对策建议 [J]. 生命科学, 2024, 36 (11): 1315 - 1320.
- 21 唐荟清, 王高玲. 新质生产力视角下医疗数据共享的价值释放及创新模式构建 [J]. 卫生经济研究, 2025, 42 (2): 32 - 36.
- 22 张建楠, 王晓杰, 沙维淋, 等. 生物医药研发数字基础设施体系建设研究 [J]. 中国工程科学, 2023, 25 (5): 92 - 102.

《医学信息学杂志》开通微信公众号

《医学信息学杂志》微信公众号现已开通, 作者可通过该平台查阅稿件状态; 读者可阅览当期最新内容、过刊等; 同时提供国内外最新医学信息研究动态、发展前沿等, 搭建编者、作者、读者之间沟通、交流的平台。可在微信添加中找到公众号, 输入“医学信息学杂志”进行确认, 也可扫描右侧二维码添加, 敬请关注!



《医学信息学杂志》编辑部