• 医学信息研究 •

解析“数据域”：内涵界定、边界划分与可信数据空间实证应用

蔡煜锋[1,2] 胡德华[2] 和晓峰[1] 伍丽群[1] 赵成闻[1] 陈 芮[1] 吴旭生[1]

（[1]深圳市卫生健康发展研究和数据管理中心 深圳 518000
[2]中南大学生命科学学院生物医学信息学系 长沙 410013）

〔摘要〕 **目的/意义** 系统解析“数据域”内涵与边界，为“数据不出域”政策实践提供理论支撑。**方法/过程** 通过政策文献分析、概念界定与实证研究，明确“数据域”的内涵与外延。通过统一数据治理逻辑与安全责任边界，将深圳市医学人工智能创新平台与全民健康信息平台纳入同一逻辑域，运用隐私计算技术实现安全体系重构。**结果/结论** 在统一管理主体和安全责任边界下，可信数据空间可兼顾“数据不出域”与数据价值释放，为医疗数据要素流通提供应用范式。

〔关键词〕 数据域；数据要素；可信数据空间；数据安全
〔中图分类号〕R-058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673-6036.2026.02.004

Deciphering the “Data Domain”: Connotation Definition, Boundary Demarcation, and Empirical Application in Trusted Data Spaces

CAI Yufeng[1,2], HU Dehua[2], HE Xiaofeng[1], WU Liqun[1], ZHAO Chengwen[1], CHEN Rui[1], WU Xusheng[1]

[1]Shenzhen Health Development Research and Data Management Center, Shenzhen 518000, China; [2]Department of Biomedical Informatics, School of Life Sciences, Central South University, Changsha 410013, China

〔**Abstract**〕 **Purpose/Significance** To systematically analyze the connotation and boundary of the “data domain”, and to provide theoretical support for the policy practice of the “data remains within its domain”. **Method/Process** Through policy document analysis, conceptual definition, and empirical research, the connotation and extension of the “data domain” are clarified. By unifying the data governance logic and security responsibility boundaries, the Shenzhen medical artificial intelligence innovation platform and the regional health information platform are incorporated into the same logical domain, and privacy computing technologies are employed to reconstruct the security system. **Result/Conclusion** Under unified management entities and security responsibility boundaries, the trusted data space can balance the requirements of “data remains within its domain” and the release of data value, providing an application paradigm for the circulation of medical data elements.

〔**Keywords**〕 data domain; data element; trusted data space; data security

〔修回日期〕 2025-12-29

〔作者简介〕 蔡煜锋，博士研究生，发表论文16篇；通信作者：吴旭生，正高级工程师，博士生导师。

〔基金项目〕 国家科技重大专项（项目编号：2023ZD0509702）；深圳市“医疗卫生三名工程”项目（项目编号：SZSM202311031）。

1 引言

在数字经济高速发展的时代，数据已成为继土地、劳动力、资本、技术之后的第5大生产要素[1]。我国高度重视数据要素发展，将统筹数据发展与安全置于战略高位，发布了一系列法律法规、方案建议。《数据安全法》第六条提出，各地区、各部门对本地区、本部门工作中收集和产生的数据及其安全负责。《医疗卫生机构网络安全管理办法》对“网络安全管理”提出一个中心（安全管理中心）、三重防护（安全通信网络、安全区域边界、安全计算环境）要求，并提出要加强物理安全防护，完善机房、办公环境及运维现场等安全控制措施，防止非授权访问物理环境造成信息泄露。《浙江省公共数据资源授权运营管理办法》明确授权运营域的内涵，即由公共数据主管部门依托一体化智能化公共数据平台建设的，为运营机构提供加工处理授权运营公共数据服务的特定安全域，具备安全脱敏、访问控制、算法建模、监管溯源、接口生成、封存销毁等功能。综合来看，《数据安全法》虽未直接定义“域”，但强调了“谁收集谁负责、谁处理谁负责”的责任归属原则，暗示“责任域”的存在。《医疗卫生机构网络安全管理办法》将“医疗卫生机构”视为“数据管理域”。《浙江省公共数据资源授权运营管理办法》则细致要求了公共数据安全域应有的具体功能。

当前，在推进数据要素变革和释放数据要素活力的过程中，数据安全管理与高效配置面临严峻挑战[2]，需要进一步明确数据安全与流通的边界，为数据要素的安全、有序流动建立可靠的技术架构与实现基础[3]。数据管理边界，即“数据域”的相关概念与术语仍缺乏统一、明确的界定，极易导致管理权责不清、安全合规边界判断失据、跨主体流通协作受阻等问题，严重制约数据要素市场的建设和健康发展。

本研究系统解析“数据域”的内涵与边界，从“域”与“数据域”的基础定义出发，深入分析当前面临的问题与挑战，并概述“数据域”在可信数据空间建设中的实证应用，为数据要素的安全流通与高效配置提供理论支撑，推动数据要素市场与医疗卫生领域的高质量发展。

2 基本概念与政策

2.1 “域”与“数据域”

在计算机科学领域，“域”在不同系统与软件中的含义不同：“域”既是Windows网络操作系统的逻辑组织单元，也是internet的逻辑组织单元[4]。《牛津高阶英汉双解词典（第10版）》中将“域”解释为互联网上以同一组字母结尾的一组网站。而“数据域”区别于网络域之处在于，其核心并非指向网络的连接与寻址方式，而是指向围绕数据要素生成、存储、处理与流转全过程所划定的逻辑与物理边界。

虽然目前针对“数据域”尚未发布标准化术语定义，但综合政策法规施行及行业实践情况，可以认为“数据域”的实质并非物理边界，而是基于管理权属、合规责任与安全策略一致性定义的逻辑边界，具体而言就是由规则和机制构成的集合，决定特定网络环境中，符合管理权限、安全策略及合规要求的计算机或设备，能否接入并参与数据存储、处理与交互等数据活动。本研究提出的“数据域”概念与“责任域”“信任域”“安全域”“治理域”等概念既相关联又有区分：“责任域”聚焦数据权责边界，“数据域”则进一步按业务逻辑聚合数据，支撑协同决策；“信任域”强调数据共享的技术可信机制，“数据域”则将其嵌入业务场景，实现合规与技术的双重信任；“安全域”侧重数据防护标准，“数据域”则将安全要求融入划分逻辑，如高敏感域绑定严格管控策略；“治理域”强调数据管理责任框架，“数据域”则将其落地为可操作单元。“数据域”将权责划分、信任机制、安全管控等治理要素进行有机融合，是体现多维度治理的综合概念。综上，本研究将“数据域”的内涵界定为：在统一管理主体和安全责任边界下的数据治理与计算环境，在该环境内的数据存储、处理与访问均受同一安全策略与合规性约束，对外仅开放受控数据交

互接口。“数据域”概念的外延在不同语境下指向不同。在数据治理语境下，指数据控制权与责任边界；在网络安全语境下，指对应数据隔离、访问控制与可信边界；在隐私计算语境下，指用于界定数据存储和计算的逻辑范围。

2.2 “数据不出域”

将“数据域”概念关联并落地为具体管理实践，是推动数据要素安全流通的关键。近年来，国家层面和地方出台的多份重要政策文件均明确提出“原始数据不出域”要求，进一步强化了“数据域”在数据共享与安全控制中的核心地位。2022 年《关于构建数据基础制度更好发挥数据要素作用的意见》首次系统提出“原始数据不出域、数据可用不可见”。在随后发布的《“十四五”全民健康信息化规划》《关于加强数据资产管理的指导意见》等一系列文件中，相关内容持续演进和细化。至 2025 年已发展为涵盖“数据可用不可见、数据可控可计量”的完整要求体系，清晰地体现了从原则探索到精细化管理的发展脉络。然而，这些政策指引在具体的落地实践中，却因“域”的定义模糊而面临一系列问题与挑战。

3 当前面临的问题与挑战

3.1 相关概念模糊导致管理权责不清

医疗机构、监管部门和科技企业对“域”的理解存在显著差异，导致数据存储、处理与使用的物理边界与逻辑边界模糊。一方面，物理边界可能被理解为单一机构的服务器或本地数据中心；另一方面，逻辑边界可能涵盖医疗联合体、区域卫生平台甚至虚拟网络。这种多元解读使医疗数据管理的权责归属陷入混乱，难以建立清晰的安全责任体系，导致以下问题。一是权责混乱，当数据在模糊的“域”之间流动时，一旦泄露或被滥用，难以追溯责任主体[5]，形成安全责任真空。医疗数据共享环节链条长，隐私信息泄露风险随之增加[6]。二是监管重叠，不同监管部门基于各自对“域”的理解制定规则[7]，导致医疗机构面临多重标准。不同部门对“数据不出域”的解释存在差异，使医疗机构在建设区域医疗数据平台时难以适从，严重阻碍医疗数据互联互通的整体进程。

3.2 标准适配困难，数据互通受阻

技术适配成本高昂已经成为医疗机构落实“数据不出域”的主要障碍。医院在推进数据互通过程中，通常采用卫生信息交换标准（health level 7，HL7）[8]等国际通用标准，同时还要对历史系统进行升级和本地化适配。对于信息化建设水平不高的医疗机构来说，此类技术改造成本高昂且复杂。由于不同厂商开发的医疗信息系统采用各自的数据格式与接口标准，系统间数据交换困难[9]。如果大型综合医院无法与基层医疗系统共享患者病历，易导致慢性病随访数据断档、转诊流程烦琐、长期疗效跟踪困难等问题。

跨域协作困难，创新受限。当前各医院采用不同电子病历系统，数据难以汇总统计，数据管理和开发利用难度大[10]，严重制约了多中心临床研究等科研协作，以及对大量宝贵健康数据资源所蕴含的巨大科研价值的有效挖掘。

3.3 安全合规边界模糊，监管滞后

安全合规边界模糊，一定程度限制了医疗健康数据领域的创新活力。在医疗健康大数据安全管理实践中，“域”的概念界定不清使数据控制者难以判断在何种情况下构成数据“出域”，机构无法明确自身是否符合《个人信息保护法》《数据安全法》等法规要求，也无法实施精准有效的安全保护措施。这种边界模糊不仅增加了数据泄露风险，更导致医疗机构在数据利用方面过度谨慎，阻碍了医疗科研创新和公共卫生进步。随着联邦学习[11]、区块链[12]、安全多方计算[13]等新技术在医疗数据领域的应用，传统“域”边界概念面临挑战。监管框架更新速度远跟不上技术发展步伐，导致医疗机构和创新企业在开展数据合作时缺乏明确指引。

3.4 数据权益纠纷复杂，共享动力不足

数据权益矛盾主要体现在两个层面：一是患者

对自身健康数据拥有合法权益，但数据的采集与处理又离不开医疗机构的专业能力；二是医疗机构在数据加工中投入了资源，但数据的敏感性质又限制了其自主决策权。《个人信息保护法》将医疗健康信息规定为敏感个人信息，要求取得个人的“单独同意”。这种法律要求在实际操作中变得异常复杂，当数据需要在不同“域”之间流动时，知情同意的具体形式、范围与有效性均成为难以厘清的问题。

当数据共享边界模糊时，医疗机构缺乏动力推动数据跨“域”使用。在数据要素市场化的进程中，如果数据所有权和使用权的界定出现模糊，容易造成价值创造与回报分配难以对等，继而使医疗机构管理者倾向于采取保守策略，将数据限制在机构内部使用，但此类做法可能影响患者诊疗的连续性。

综上，清晰界定“数据域”有利于进一步明确数据管理的安全责任逻辑边界和管理主体，有助于消除权责混乱，减少监管重叠，降低技术适配成本，促进互联互通与跨域协作，进而激发数据创新活力与共享动力，解决数据权益纠纷。

4 可信数据空间实证应用

“域”界定模糊导致权责不清、标准缺失与创新受限等问题，根源在于常将“数据不出域”简单等同于数据部署的物理隔离。这种认知与实践加重了机构间的“数据孤岛”现象，一定程度上扼杀了数据要素流通所必需的系统性信任。应从执着于分散的物理边界，转向构建统一、逻辑上、可信的“数据域”。可信数据空间正是针对此困境提出的系统性解决方案，即在同一管理体系与安全体系下，将多方数据资源纳入一个逻辑统一但权责清晰的“数据域”，从而在保障“数据不出域”的同时，高效促进数据要素的价值流通与协同创新。

2024 年 11 月国家数据局印发《可信数据空间发展行动计划（2024—2028 年）》，主要包括 3 项行动：可信数据空间能力建设行动、可信数据空间培育推广行动、可信数据空间筑基行动。2025 年 11 月国家卫生健康委员会等多部门印发《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》，在总体目标中提出“到 2027 年，建立一批卫生健康行业高质量数据集和可信数据空间”，将“支持各地建立高质量医疗健康数据的可信数据空间”作为深化重点应用的重要内容之一，并将“开展高质量数据集建设和可信数据空间建设试点”作为加强组织保障的重点工作之一。

本研究积极响应国家级战略指引，通过明确“数据域”定义解决数据要素流通中的关键问题。以深圳市现有全民健康信息平台和医学人工智能创新平台为坚实底座，通过系统性功能整合、技术升级与规则创新，在同一“数据域”内构建符合国家规范、服务于医学人工智能创新的医疗健康可信数据空间。定义明确前，深圳市医学科学数据研究应用可信数据空间搭载于医学人工智能创新平台，而医学人工智能创新平台与深圳市全民健康信息平台的物理空间分属两地，各自物理隔离且安全体系不同，虽然同属一个管理主体，也有同源互认的安全体系，但存在数据流通障碍和壁垒，如系统兼容技术适配成本较高等。定义明确后，对医学人工智能创新平台与全民健康信息平台进行安全体系统一化，虽然物理存储位置不同，有物理隔离，但其数据管理主体和安全体系相同，可以明确认为是同一“数据域”，基于隐私计算（如联邦学习、安全多方计算）等安全技术，可允许数据要素进行安全的域内流通，充分利用算力、存储等资源，实现数据要素高效配置和深度应用。

本研究通过将“数据域”定义明确为统一管理权属与安全责任边界的逻辑治理域，解决了物理隔离导致的“数据孤岛”问题、标准碎片化引发的合规风险，以及数据流通的协同障碍。可信数据空间建设实践表明，逻辑边界清晰化可显著提升数据资源利用效率，为“人工智能+医疗卫生”深度融合提供关键基础设施支撑。

5 “数据域”定义的适用边界与跨域情形

在多元管理主体协同、跨区域平台共建或政企联合数据空间等复杂场景中，“数据域”定义外延的适用性体现为通过逻辑划分实现数据资源的高效

整合与安全治理。“域”的判定标准应围绕价值导向、技术适配、风险可控3个维度展开。

从价值导向看，“数据域”应服务于场景的核心目标。以广东省政务服务“一网通办”平台为例，该平台全面集成省、市、县、镇、村5级政务服务事项，将个人服务和法人服务按照办理热度、主题、部门、“最多跑一次”等分类进行导航，支持“社保服务跨省通”“跨域通办服务专区”“跨境通办服务专区”等业务场景。不仅解决了数据孤岛问题，更通过“数据域-业务场景”的映射关系，使数据流动直接服务于公共服务均等化、市场监管协同化等政策目标。

技术适配性是“数据域”落地的关键支撑。在区块链驱动的物联网数据跨域共享场景[14]中，“数据域”可定义为“设备类型-数据类型-权限等级”三维结构，即通过智能合约实现去中心化的访问控制，既保障了数据交换的可靠性，又满足了跨域场景下细粒度管控需求。

风险可控性要求“数据域”划分兼顾安全与合规。在跨区域数据共享中，“数据域”的判定可引入“影响对象-影响范围-影响深度”矩阵模型。涉及个人健康信息的“数据域”，其影响对象为自然人，影响范围超过500人则应标注为“较敏感数据”，当影响深度评估显示存在隐私泄露风险时，应强制采用加密传输、脱敏处理等技术措施。这种分类分级标准与《网络安全审查办法》《数据安全法》等法规衔接，确保“数据域”划分既符合业务需求又满足合规要求。在政企联合数据空间中，政府侧“数据域”（如人口基础信息）与企业侧“数据域”（如消费行为数据）的边界划分，应通过数据沙箱[15]、联邦学习[16]等技术实现“数据可用不可见”，在保障公共利益的同时维护企业商业秘密。

6 结语

本研究首次在学术与实践结合层面，对“数据不出域”这一关键政策原则中“域”的概念进行清晰界定，明确提出其本质是基于统一管理权属和安全责任边界的逻辑治理域，而非物理隔离域。通过重新界定深圳市全民健康信息平台与医学人工智能创新平台之间“域”的关系，在技术上精准施策、在规则上有效协同，最终打通数据要素安全流通与价值释放的“最后一公里”，提供了可复制的建设范例。本研究目前存在单区域案例普适性验证不足问题，量子加密、动态脱敏技术与现有框架适配性研究亦存在欠缺。未来可拓展至多区域、多场景对比实验，进一步验证“数据域”概念内涵和外延的普适性与可扩展性，深化人工智能驱动的动态边界管理机制研究，探索量子加密技术与隐私计算的融合应用，为全球数字安全治理贡献中国方案。

作者贡献：蔡煜锋负责文献调研、研究设计与实施、论文撰写与修订；胡德华、和晓峰负责文献调研、研究设计与实施；伍丽群、赵成闻、陈芮负责研究设计与实施；吴旭生负责研究设计、论文撰写与修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 张平文，邱泽奇. 数据要素五论[M]. 北京：北京大学出版社：2022.
2 王海杰，王开阳. 数据要素驱动新质生产力发展的机制、挑战与应对措施[J]. 中国流通经济，2025，39(1)：3-13.
3 王如玉，梁琦. 从“数字中国”迈向“智能中国”——由数据生产要素形成谈起[J]. 河南社会科学，2025，33(9)：13-23.
4 鞠光明. 微软网络操作系统[M]. 北京：机械工业出版社，2007.
5 范渊，刘博，段平霞. 数据安全与隐私计算[M]. 北京：电子工业出版社，2025.
6 贾末，计虹，李翠霞，等. 医疗数据共享过程中的信息安全与隐私保护[J]. 中国数字医学，2025，20(11)：19-24.
7 张威. 数据跨境流动的“安全港”规则重构——基于数字贸易与国家安全的平衡[J]. 南腔北调，2025(25)：61-64.
8 王宇，郭思捷，王亚宁，等. 公立医院医疗信息共享的挑战及应对策略研究[J]. 信息与电脑，2025，37(4)：128-131.
9 郁文谊，刘亚军，郜翀. 联邦治理×智能合约：长三角跨

（下转第36页）

ing, 2021.
4 曹树金, 吴育冰, 韦景竹, 等. 知识图谱研究的脉络、流派与趋势——基于SSCI与CSSCI期刊论文的计量与可视化[J]. 中国图书馆学报, 2015, 41(5): 16-34.
5 罗依宁, 崔雷. 医学相关信息学子学科研究重点与关系探究[J]. 医学信息学杂志, 2025, 46(5): 50-55, 66.
6 梁镇涛, 巴志超, 徐健. 基于引文的跨学科领域发展路径分析——以眼动追踪领域为例[J]. 图书情报工作, 2019, 63(23): 65-78.
7 李点, 文庭孝, 许林勇. 引文施引双角度医学信息学交叉测度分析[J]. 医学信息学杂志, 2024, 45(5): 46-52, 64.
8 TAMINE L, GOEURIOT L. Semantic information retrieval on medical texts[J]. ACM computing surveys, 2021, 54(7): 1-38.
9 WU X, NGUYEN T, LUU A T. A survey on neural topic models: methods, applications, and challenges[J]. Artificial intelligence review, 2024, 57(2): 1-30.
10 宋俊杰, 尹裴, 邓诗语, 等. BERTopic在医疗领域文章主题挖掘中的应用与分析[J]. 软件工程, 2025, 28(4): 62-66, 72.
11 YU D, XIANG B. Discovering topics and trends in the field of artificial intelligence: using LDA topic modeling[J]. Expert systems with applications, 2023, 225(9): 120114.
12 夏苏迪, 谢靖, 是沁, 等. 基于BERTopic模型的我国中医药老年康养政策量化研究[J]. 医学信息学杂志, 2025, 46(6): 43-49.
13 ROMAN E, JOANNE Y. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts[J]. Frontiers in sociology, 2022(7): 886498.
14 MAARTEN G. BERTopic: neural topic modeling with a class-based TF-IDF procedure[EB/OL]. [2025-09-29]. https://arxiv.org/abs/2203.05794.
15 YU G, ROBERT T, HAO C, et al. Domain-specific language model pretraining for biomedical natural language processing[J]. ACM transactions on computing for healthcare, 2022, 3(1): 1-23.
16 GARG M, WANG L, GHANCHI B, et al. Biomedical literature QA system using retrieval-augmented generation (RAG)[EB/OL]. [2025-09-29]. https://arxiv.org/abs/2509.05505.
17 PATRIK L S, FREDRIK S, SANJA V, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics[J]. Science, 2016, 353(6294): 78-82.
18 CHENGLONG X, JEAN F, GEORGE E, et al. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression[J]. Proceedings of the national academy of sciences of the United States of America, 2019, 116(39): 19490-19499.
19 JOHN J, RICHARD E, ALEXANDER P, et al. Highly accurate protein structure prediction with alphafold[J]. Nature, 2021, 596(7873): 583-589.
20 ZEYE L, HONG J, FENGWEN Z, et al. The examination of expression patterns, underlying mechanisms, diagnostic accuracy, and potential AI-driven drug development approaches for ferroptosis-related genes in heart failure via single-cell and bulk RNA sequencing analyses[EB/OL]. [2025-09-29]. https://doi.org/10.1161/circ.150.suppl_1.4140707.

（上接第29页）

层级医疗数据协同的创新探索[J]. 医学信息学杂志, 2025, 46(11): 28-34, 41.
10 王彦霞. 新时代医院电子病历档案共享机制构建研究[J]. 兰台内外, 2025(17): 1-4.
11 MAURER M M, PFITZNER B, VAN DE WATER R P, et al. Privacy preserving federated learning for 90-day mortality prediction in colorectal surgery: a multicenter retrospective development and comparison study[J]. International journal of surgery, 2025, 111(12): 9065-9074.
12 王钰涵, 孙燕杰. 区块链隐私计算赋能智慧医疗数据共享[J]. 中国科技信息, 2025(21): 131-134.
13 LI J, WANG D, QI G, et al. Alliance chain-based simulation on a new clinical research data pricing model[J]. Annals of translational medicine, 2022, 10(15): 836.
14 李哲成, 张波. 基于区块链的轻量级工业物联网跨域认证与数据共享方案[J]. 计算机研究与发展, 2025, 62(10): 2416-2427.
15 刘鲲, 王羽赫. 数字政府背景下基于数据可信计算沙箱技术的应用[J]. 网络安全和信息化, 2023, (11): 113-116.
16 胡业飞, 陈美欣, 张怡梦. 价值共创与数据安全的兼顾: 基于联邦学习的政府数据授权运营模式研究[J]. 电子政务, 2022(10): 2-19.