数据胶囊：面向数据要素化与交易流通的新范式

王锐璇[1,2,3]，刘哲理[1,2,3]
（1. 南开大学密码与网络空间安全学院，天津 300350；2. 天津市网络与数据安全技术重点实验室，天津 300350；
3. 数据与智能系统安全教育部重点实验室，天津 300350）

摘要：针对数据要素化过程中由数据易复制与非竞争性导致的权属难界定、流转易失控及价值难评估等现实困境，创新性地提出数据胶囊流通范式，以“环境受控、数据有形、人有身份”三角模型为设计逻辑，构建覆盖本体论、权属论、流通论和定价论的四层理论体系，通过数据本体与加工使用策略深度绑定、可信执行环境技术支撑，实现数据要素在安全受控前提下的自由流通与价值释放。多维度的对比分析结果显示，数据胶囊范式在数据要素化支撑能力与综合性能上整体优于传统技术方案。该范式适用于数据集交付、特殊文件受控流转、数据合规共享、个人隐私数据可控流转等现实应用场景，能够为可信数据空间提供权属流转与全流程可控的技术补充。

关键词：数据要素化；数据胶囊；数据流通；可信执行环境
中图分类号：TP309
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB2026205

Data capsule: a new paradigm for data elementization and transaction circulation

Wang Ruixuan[1,2,3], Liu Zheli[1,2,3]
1. College of Cryptology and Cyber Science, Nankai University, Tianjin 300350, China
2. Tianjin Key Laboratory of Network and Data Security Technology, Tianjin 300350, China
3. Key Laboratory of Data and Intelligent System Security, Ministry of Education, Tianjin 300350, China

Abstract: To address difficult ownership definition, easy circulation loss of control, and hard value evaluation challenges, caused by data’s high replicability and non-rivalry in data factorization, a new data capsule circulation paradigm was proposed. Based on the triangular model of “controlled environment, tangible data, and identified personnel”, a four-layer theoretical system covering ontology, ownership, circulation, and pricing was constructed. Through in-depth binding of data with processing and usage strategies and support of trusted execution environments, secure data circulation and value release were enabled. Multi-dimensional analysis shows that the data capsule circulation paradigm outperforms traditional solutions in data factorization support and comprehensive performance. The proposed paradigm is applicable to scenarios such as dataset delivery, controlled circulation of special files, compliant data sharing, and controllable circulation of personal privacy data, providing a vital technical supplement for trusted data space to ensure ownership circulation and full-process controllability.

Key words: data elementization, data capsule, data circulation, trusted execution environment

收稿日期：2026–04–15；**修回日期：**2026–04–29
通信作者：刘哲理，liuzheli@nankai.edu.cn
基金项目：国家自然科学基金联合基金资助项目（No.U25B2028）；国家自然科学基金重点基金资助项目（No.62432012）
Foundation Items: The Joint Fund of the National Natural Science Foundation of China (No.U25B2028), The Key Program of the National Natural Science Foundation of China (No.62432012)

0　引言

数据要素化是指将原始数据转化为具有经济价值和可利用性的生产要素的过程，是充分激发数据经济潜力，驱动数字经济高质量发展的核心引擎。这一过程通常被视为数据形态的动态演进，涵盖了从数据资源化、数据产品化到数据资产化的 3 个关键阶段[1]。其中，数据资源化通过采集、清洗、脱敏和标注等治理手段保障数据的可用性；数据产品化通过标准化封装赋予数据可交易性；数据资产化则通过流通机制实现数据的价值变现。

当前，数据要素化在理论与实践层面均面临深刻挑战，其根本原因在于数据的独特属性。相较于土地、劳动力等传统生产要素，数据具有无形性、可复制性、场景依赖性的本质特征，与传统生产要素的产权清晰、独占性、标准化流通形成鲜明差异，这种属性差异导致数据要素化在理论与实践层面均面临深层次挑战，成为制约全国统一数据大市场建设的关键障碍。

1　数据要素化挑战

1.1　权属界定困难

政策层面虽提出数据资源持有权、数据加工使用权、数据产品经营权“三权”分置框架，为多方权益平衡提供了制度指引，但该框架主要侧重于利益分配。从安全的视角来看，数据的可无限复制性使其所有权无法像物理资产那样形成排他性占有，一旦数据脱离原始控制域，所有权主体便丧失对数据的实际管控，导致数据源头方因担心个人资产流失产生“不敢供、不愿供”的心理。因此，数据主权主体仍需构建以所有权为基础的实质性控制权。同时，衍生数据、融合数据的权属判定缺乏统一标准，进一步加剧了数据确权的复杂性。

1.2　流通自由与安全合规矛盾

数据的非竞争性与易复制性使其在流通过程中面临管控困难的固有难题：数据在数字空间中可被无痕复制，且副本与原件具有同等效力，传统物理资产的保护手段完全失效。一方面，数据一旦完成初次交付，所有权方难以阻止接收方的非法复制、二次传播与倒卖行为，极易陷入交付后管控权丧失的困境，造成数据价值快速稀释，严重损害数据生产者的积极性；另一方面，现有技术手段多聚焦于数据不出域的共享场景，依托联邦学习、安全多方计算等实现数据可用不可见，但这类技术无法支撑数据出域的交易与交付需求。数据不出域便不能流通[2]，数据出域则难以管控，这一难题已成为数据要素市场化配置的瓶颈。

1.3　数据价值评估难题

一方面，数据的价值取决于具体应用场景、分析能力与使用主体，买卖方信息不对称导致数据价值具有双向不确定性[3]；另一方面，数据质量、稀缺性、时效性缺乏统一量化标准[4]，为数据要素的市场化定价带来了阻碍。

针对上述三大挑战，现有研究方案呈现出两条演进路径。一条是自顶向下的政策驱动路径，以“数据二十条”为核心，通过构建可信数据空间等标准体系，从制度层面规划宏观架构。然而，该路径目前仍侧重于顶层设计，针对底层技术落地、具体安全机制落地与交易能力支撑的实现路径仍显模糊。另一条是自底向上的技术演进路径，学界主要以隐私保护计算技术为切入点，针对性解决数据流通中特定阶段或单一场景的安全问题。在数据权属方面，区块链技术常被用于实现数据权属流转可追溯[5-6]；数字水印技术则为数字版权的保护与溯源确权提供了可行思路[7-8]；同时有大量研究围绕细分场景，探索差异化的数据所有权保护机制[9-10]。在数据安全流转与共享利用方面，主要依托联邦学习完成跨主体联合建模[11-12]，基于安全多方计算实现数据协同共享与联合计算[13-14]，利用可信执行环境（trusted execution environment，TEE）实现可验证的数据处理[15-16]，同时传统访问控制方法也被广泛应用与研究[17]。在数据定价方面，逐步衍生出适配隐私约束的差分隐私定价模式、基于多方贡献度量化的可解释定价方法等多元研究方向[18-19]。然而，这些技术往往面临性能瓶颈或场景适配性窄的问题，难以覆盖数据要素全链条、多主体、跨域流转的复杂安全需求。

2　数据胶囊理论体系

数据胶囊的核心要义在于通过将安全策略与数据对象进行全生命周期的强耦合绑定，实现数据在流转过程中的策略跟随与受控流通。本文提出的数据胶囊范式与 Wang 等[20-21]提出的“Data Capsule”有着本质区别：后者侧重于数据共享场景下的数据合规使用与访问控制，而本文则立足于数据要素化

背景，旨在解决数据跨域流通场景下的安全管控问题。然而，仅依靠这一内生安全架构仍不足以支撑数据要素化的全流程需求，因此本文进一步提出了数据胶囊的三角模型，并构建了涵盖本体论、权属论、流通论和定价论的四层理论体系。

2.1　三角模型

数据要素化的主要矛盾在于数据的无形态、权属易失控，而传统解决方案未能从根本上建立适配数据特征的管控逻辑。本文提出“环境受控、数据有形、人有身份”的数据胶囊三角模型，如图 1 所示，其作为数据胶囊的核心设计逻辑，实现数据要素自由流通与可控流通的统一。

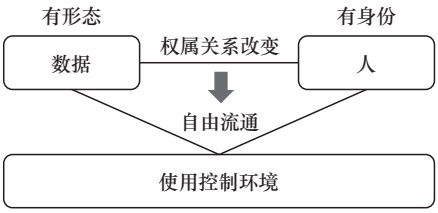

图 1　数据胶囊三角模型

（1）使用控制环境：统一受控的根基

使用控制环境是数据胶囊生成、使用的全流程载体，通过统一部署、标准化建设与严格的身份认证，赋予环境唯一数字身份，成为三角模型的根基与不变量。该环境以可信执行环境为基础，结合公钥基础设施、分布式账本等技术，实现数据胶囊操作的机密性、完整性与可追溯性，确保所有数据操作均在受控域内完成，从底层降低数据脱离管控的风险。使用控制环境的统一受控是数据胶囊实现安全流通的前提，其身份化设计使环境本身可被验证、可被监管，为数据与主体的交互提供可信基础。

（2）数据：以胶囊为载体的有形形态

数据胶囊为数据赋予可管控、可标识的逻辑有形形态，使数据从无序的比特流转化为标准化的数字资产单元。数据胶囊通过密文封装与策略封装，将数据与控制策略、元数据、数字签名深度耦合，成为数据传输、存储、使用的最小单元，赋予数据类物理资产稀缺性与可控性。数据胶囊的逻辑有形形态使数据可被确权、可被计量、可被交易，解决了数据因无形而难以管控的底层难题。

（3）人：身份化的主体

对数据要素的所有参与主体（数据提供方、数据使用方等）进行统一的身份管理与属性认证，赋予主体唯一的数字身份，从而便于将人和数据联系起来。

三角模型支撑数据自由流通的逻辑在于，将使用控制环境的统一性与数据胶囊形态的固定性作为不变量，而将人与数据的权属关系设为可变量。通过在标准化的使用控制环境中执行使用、加工等操作，数据胶囊的持有权、加工使用权、收益权得以在不同身份主体间进行动态流转，进而支持数据的可控交易流转。

2.2　四层理论体系

如图 2 所示，本文基于三角模型，构建本体论、权属论、流通论、定价论四层理论体系，从数据胶囊视角系统论证其对数据要素权属界定、合规流通、市场化定价的全方位支撑，形成较完整的理论框架。

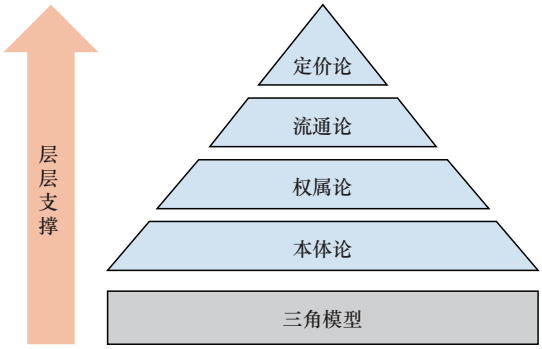

图 2　数据胶囊四层理论体系

2.2.1　本体论：数据胶囊的本质与生命周期构建

本体论旨在系统性地界定数据胶囊的本质范畴、核心组件与生命周期，解决数据要素化过程中载体标准化与全链路受控的底层逻辑难题。

数据胶囊由元数据、策略、加密数据 3 个部分组成，是一体化、自描述、可自治的数据基本单元。元数据为数据胶囊赋予唯一标识，包含胶囊标识符、生产者标识符、数据描述、生成时间等信息，是数据胶囊确权、追溯的基础。策略描述语言实现自然语言的用户偏好到机器可执行规则的转化，是数据胶囊细粒度管控的核心；加密数据即原始数据的密态形式。数据胶囊按照其粒度可分为原子胶囊、复合胶囊、聚合胶囊。原子胶囊对应单一维度的原始数据；复合胶囊由多个原子胶囊融合而成；聚合胶囊则通过对复合胶囊的加工分析形成具有更高价值的衍生数据，实现数据价值的分层释

放。此外，数据胶囊构建了涵盖“生成–封装–存储–流通–使用–更新–销毁”的全生命周期闭环模型，确保数据在全生命周期均以数据胶囊的形式存在。

2.2.2 权属论:“三权”分置的胶囊化重构与权属判定

数据胶囊将数据资源持有权、数据加工使用权、数据产品经营权“三权”分置框架进行技术化重构，实现权利的清晰分离与灵活流转。数据胶囊的所有权对应数据资源持有权，由原始数据拥有者持有，体现为对数据胶囊封装载体的占有与流转控制；数据胶囊的加工使用权对应数据加工使用权，通过策略授权机制实现，主体仅能在可控执行环境中，按照预设策略对数据胶囊进行编辑、运算或分析，实现数据可用不可见；数据胶囊的收益权对应数据产品经营权，由持有权方授权管理，通过收益分配机制实现数据要素的价值变现，收益分配规则可作为策略字段标注在数据胶囊中，成为交易完成时的利益分配凭证。

在权属认定方面，数据胶囊支持动态演进的判定逻辑。原始胶囊遵循“生成者持有”的原则，数据生产者为原始胶囊的天然持有方，享有绝对的持有权与收益权；衍生或融合胶囊则依据加工贡献度及事前约定的合约，实现多主体间的协同确权与收益共享。为保障权属的确定性，系统结合唯一标识、数字签名及分布式账本技术，实现了权属变更的全程存证与不可篡改，从而在技术层面厘清了数据资产的产权边界。

此外，数据胶囊范式严格区分了原始数据所有权与胶囊所有权。前者作为数据生产者的本源权利具有不可剥夺性，而后者则具备商品属性并可随交易流转。数据使用方获取胶囊所有权，并非获得原始数据的控制权与底层主权，而是取得在合规契约约束下，对数据胶囊进行二次加工、创造新价值并获取经济收益的技术凭证。

2.2.3 流通论:胶囊化的流通本质与交易本质

数据胶囊的流通并非原始数据比特流的简单复制，而是胶囊的持有权、加工使用权、收益权的合约化转移，即三角模型中人与数据的关系的转换。数据流通的交互逻辑体现为受控使用、受控加工、受控交易 3 个过程，所有过程均在可控执行环境中完成，且严格遵循策略约束，确保数据流转的合规性与可控性。

数据胶囊适配数据要素市场的多样化需求，构建了场内交易、场外交易、跨境交易、融合交易四大流通模式，实现数据要素的全场景流通。场内交易依托数据交易所进行标准化数据胶囊的挂牌、竞价、撮合、结算，交易所作为第三方监管主体，保障交易的公平性与合规性。场外交易实现数据买卖双方的点对点胶囊合约签署、授权、使用、计量，适用于个性化、定制化的数据需求场景。跨境交易通过数据胶囊，结合跨境合约与监管沙箱，实现数据胶囊的跨境安全流通，适配数字经济的全球化发展需求。融合交易支持多源数据胶囊的合规融合，实现多主体的价值共创与收益共享，推动数据要素的规模化配置。

数据胶囊构建了信任机制、合规机制和监管机制三位一体的交易机制，为数据流通提供全方位保障。信任机制通过胶囊受控封装、分布式账本存证等技术融合，实现数据“可用不可见、可控可追溯”，构建交易主体间的信任基础。合规机制通过将监管策略与合规要求硬编码于胶囊内部，实现合规策略的自动校验与跨域适配，以实现流通即合规。监管机制通过数据胶囊全生命周期可存证、细粒度监控与自动审计，实现对数据流通的事前预防、事中控制、事后追责，推动监管范式从外部被动审计向内生自动执行转变。

2.2.4 定价论:数据胶囊的价值评估机制

数据胶囊的定价论不探究数据价格与数据质量、应用场景之间的具体关联，但该关联一旦确定，数据胶囊就可全程保障其完整性与准确性，确保数据价格与数据质量、应用场景相互匹配。

此外，定价论构建了数据胶囊的三维价值构成模型，实现了对数据价值的分层量化与精准评估。

（1）基础价值：由数据的内在属性决定，包括数据质量、完整性、准确性、时效性、覆盖面等，是数据胶囊价值的基础。

（2）策略价值：由数据胶囊的管控策略决定，包括合规等级、隐私保护强度、使用灵活性、管控粒度等，策略要求越高代表数据胶囊的可用范围越小，相对应的价值相对越低。

（3）应用价值：由数据胶囊的场景适配性决定，包括场景复用性、融合能力、衍生价值、决策支撑能力等，场景适配性越强、能创造的经济价值越高，数据胶囊的应用价值越强。

3 数据胶囊体系架构

基于三角模型与四层理论体系，本文设计了数据胶囊的轻量化软硬件协同架构，实现数据胶囊从生成到流通的全流程支撑。该体系架构严格遵循三角模型“环境可控、数据有形、人有身份”的设计要求：通过构建以可信执行环境为基础的生产环境和消费环境支撑“环境可控”，通过数据以数据胶囊的形式传输实现“数据有形”，以公钥基础设施等支撑“人有身份”。在四层理论体系的落地层面，该架构依托数据胶囊标准化封装结构，落实本体论要求（详见第3.2节）；通过身份认证、权属可信流转和权属流转记录，满足权属论要求（详见第3.1节）；通过数据胶囊生成与消费模块协同设计落实流通论的设计思想（详见第3.3节）；通过基于可信执行环境的行为度量为定价论提供基础支撑。

3.1 系统架构

数据胶囊体系的参与实体基于四层理论体系的权属划分，结合三角模型的主体身份化要求，分为用户主体与监管主体两大类，各主体均具有唯一的数字身份与属性认证，实现权属关系的清晰界定与灵活流转。

（1）用户主体包括数据提供方和数据使用方，分别代表拥有并上传原始数据、生成数据胶囊的一方与购买并按策略使用、加工数据胶囊的一方，是数据胶囊权属流转的主体。

（2）监管主体负责对数据胶囊的生成、流通、使用进行全流程监管，可以通过对数据中心、分布式账本等基础设施的监管记录，确保数据要素化的合规性。

数据胶囊的基础设施是三角模型中使用控制环境的具体落地，为数据胶囊的全生命周期提供统一、受控、可信的技术支撑，主要包括生产环境、消费环境、公钥基础设施、分布式账本、云存储五大模块，各模块协同工作，实现数据胶囊的封装、存储、流转、验证与监管。

（1）生产环境负责数据胶囊的生成与封装，接收数据提供方的原始数据与策略规则，完成密文封装与策略封装，生成标准化的数据胶囊，并为胶囊赋予唯一ID与数字签名。

（2）消费环境负责数据胶囊的使用或加工，接收数据使用方的访问请求，完成策略验证、身份认证等操作后执行数据胶囊的使用与加工操作，并将操作结果反馈给数据使用方，同时更新胶囊状态。

（3）公钥基础设施为所有参与主体与基础设施提供身份认证、密钥管理与数字签名服务，是实现主体身份化、胶囊权属可追溯的基础。

（4）分布式账本负责数据胶囊状态、权属流转、操作记录的存证，实现数据胶囊的全生命周期可追溯，保障数据胶囊状态的新鲜性与权属唯一性。

（5）云存储负责数据胶囊的安全存储与检索，为数据胶囊的流通提供高效的存储载体，确保数据胶囊的可访问性。

3.2 数据胶囊结构

在微观层面，数据胶囊是一种保护隐私的数据封装结构，包含数据部分、元数据部分和策略部分。它可以被形式化定义为$\mathcal{C}=\{\mathcal{D}=\mathrm{Enc}(\mathrm{data}),\mathcal{M},\mathcal{P}\}$，其中，$\mathcal{D}$为敏感数据的加密表达，$\mathcal{M}$为敏感数据的元数据，$\mathcal{P}$为与敏感数据强绑定的一组策略规则，用以规范数据在何种条件下可被访问及具体的使用方式。

策略包括加工策略和使用策略两大类。加工策略规范数据胶囊在转移所有权时的条件和方式，使用策略规范数据胶囊在转移使用权时的访问权限和使用方式。无论是加工策略还是使用策略都可以划分为数据使用方身份（Who）、可访问时间（When）、可访问环境（Where）、可访问内容（What）和使用方式（How）5个维度。表1对数据胶囊策略维度及其具体类别进行了举例说明。就数据使用方身份而言，其可按粒度区分为身份标识符与身份属性；可访问时间不仅包括有效时间，还包含可使用次数，该次数会随着数据的使用而递减，降至0时该数据胶囊将不可再用。可访问环境通常限制数据仅能在特定地理区域或特定设备上使用。可访问内容与使用方式通常需针对目标编程语言设计，因此在不同编程语言（如Python、SQL）中会采用不同的表示形式和语法。

3.3 操作接口

数据胶囊系统架构如图3所示，包含以下功能函数，作为对用户提供的接口。

（1）初始化。首次接入系统的用户向数据中心发起请求，通过系统内各个参与方的交互，完成身份确认、密钥分配等工作。

（2）数据胶囊生成。以数据所有者的数据和策

表 1　　**数据胶囊策略维度及其具体类别**

分类		策略标识	策略值	含义
使用策略	Who	UserID	100001	标识谁可以使用该数据胶囊
		UserAttribute	Insurance Company	拥有特定属性的使用者可以使用
	When	TimeDDL	2027-01-01 00:00	仅在此时间之前，数据胶囊有效
		UseTimes	3	该数据胶囊可使用次数为 3 次
	Where	ProcessDevice	TEE	数据胶囊只能在可信执行环境中使用
	What	Project	年龄，地区	只能使用年龄和地区两个域的数据
		Filter	年龄>18	筛选年龄大于 18 岁的数据条目
	How	OutputEnc	是	只能输出密文
		DatabaseOperator	Sum:年龄	只能对年龄字段执行求和操作
加工策略	Who	BuyerID	200001	标识数据胶囊的数据使用方
	When	TimeDDL	2027-01-01 00:00	产权到期时间
	Where	ProcessDevice	TEE	数据胶囊只能在可信执行环境中编辑
	What/How	AllowedTransformation	Aggregation	数据使用方可以对数据胶囊进行一些编辑操作

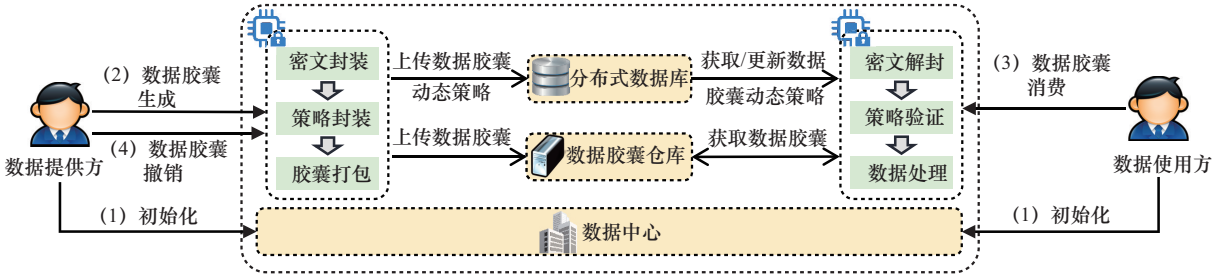

图 3　数据胶囊系统架构

略为输入，生产环境完成数据胶囊封装的工作，并将数据胶囊上传到云服务器存储。

（3）数据胶囊消费。任何用户都可以发起对某个数据胶囊的访问，访问请求参数包括数据胶囊标识符、要执行的程序等。但只有满足策略要求的访问才是被允许的，且能正确执行代码获得计算结果或输出一个新的数据胶囊。

（4）数据胶囊撤销。只有数据胶囊的所有者才有权发起撤销数据胶囊的请求。撤销之后，数据胶囊不能再流通。

3.4　数据胶囊模型分析

数据胶囊并非对现有技术的否定，而是基于数据要素化的底层需求，对现有技术进行融合重构与范式升级。为全面评估其技术效能，本节选取了当前学术界与产业界最具代表性的 3 类主流方案作为对比基准。

（1）单一隐私保护计算技术方案（后文用 P-A 表示）。以联邦学习、安全多方计算、差分隐私、同态加密等技术为基础，聚焦于数据“可用不可见”的隐私保护问题，通过技术手段实现数据在不泄露原始信息的前提下进行协同计算。典型方案包括蚂蚁集团 SecretFlow[14]、微众银行联邦学习开源框架 FATE[22]。

（2）区块链与隐私保护计算融合方案（后文用 P-B 表示）。将区块链的不可篡改与可追溯属性与隐私保护计算技术的隐私保护属性相结合，聚焦于数据流通中的可信追溯与隐私保护两个问题，通过技术组合实现数据操作的可追溯与数据内容的隐私保护。典型方案包括浙江大学团队的 PrivData[23]、基于区块链的计算网络 PlatON[24]。

（3）传统沙箱技术方案（后文用 P-C 表示）。以沙箱技术、可信执行环境为基础构建封闭的受控环境，聚焦于数据场景化受控使用的问题，通过技术隔离并融合特定的访问控制技术实现数据在特定环境内的安全使用，是数据流通的场景化管控解决方案。典型方案包括 PrivGuard[21]、DataSafe[25]、Operon[10]等。

数据胶囊范式与上述 3 类主流方案的详细对比

如表2所示。其中，√表示在该维度上表现良好，〇表示表现一般，×表示表现较差。

表2 数据胶囊范式与其他方案对比

对比维度	P-A	P-B	P-C	数据胶囊范式
权属支撑	×	〇	×	√
流通支撑	〇	〇	〇	√
定价支撑	×	×	×	〇
性能效率	〇	×	√	√
场景适配	×	〇	√	√
落地可行性	〇	×	〇	√

在权属与定价支撑上，P-A与P-C完全无权属界定、价值度量能力，P-B仅能实现权属简单追溯，均无法解决衍生/融合数据确权与定价问题，而数据胶囊通过权属论重构“三权”分置、明确判定规则，通过定价论构建综合定价框架，实现权属清晰流转与价值可量化评估。

在流通支撑上，P-A仅支持协同计算式共享、无交易能力，P-B流通模式有限且碎片化，P-C仅能实现封闭环境内的有限流通，三者合规均依赖外部审计，而数据胶囊支持场内交易、场外交易、跨境交易、融合交易四大流通模式，通过内生合规机制实现自动化校验，且能实现跨域无缝流通。

在性能效率与场景适配上，P-A和P-B存在计算交互或共识机制的性能瓶颈，且前者场景适配性差、需定制化开发，后者因标准不统一导致场景适配性一般，而数据胶囊基于可信执行环境规避了复杂交互损耗，能支撑大规模高并发场景，同时通过五维策略规则实现细粒度场景定制，可灵活覆盖个人、企业、政务等全场景。

在落地可行性上，3类主流方案均存在技术复杂度高、定制化以及集成成本高的问题，在中小企业落地门槛高或跨场景拓展难，而数据胶囊弱化底层技术细节，提供标准化接口，不需要定制化开发，落地门槛低，可适配不同技术能力的参与主体。整体而言，相较于上述方案，数据胶囊尝试从统一框架出发，对权属、流通与管控问题进行系统整合。

4 数据胶囊的技术实例化

前文提出的数据胶囊范式，构建了面向数据要素化的通用模型架构与理论体系，其权属管控、策略执行等核心模块具备技术解耦特性，可根据场景的安全需求、性能指标与部署条件，适配不同底层技术实现。本节聚焦数据胶囊性质的技术落地需求，选取可信执行环境、属性基加密、分布式账本、策略执行等技术，给出一套具体的技术实例化思路，通过构建可落地的协议，支撑数据胶囊的理论目标。

4.1 实现机制

为了支撑数据胶囊的模型，本文将需求解构为功能上权属可流转，安全性上数据胶囊防复制。数据胶囊需求与技术支撑如表3所示。权属可流转要求权属的变更过程是可信的，权属的记录是不会被回滚的。防复制体现在数据使用方不能获得原始数据明文，因此不仅需要执行过程是机密不可篡改的，还要保证数据使用方的操作是满足数据胶囊中的策略要求的。

构建的分布式账本需要支持以下基本操作。

Insert(id,s)：在账本中插入一个数据胶囊状态，其中，id是数据胶囊标识符，s是数据胶囊的状态信息。

Update(id,s')：更新数据胶囊状态，其中，id是数据胶囊标识符，s'是新的数据胶囊状态信息。

$s \leftarrow$ Query(id)：查询数据胶囊状态，其中，id是数据胶囊标识符，返回数据胶囊当前状态信息。

属性基加密的变体之一为密文策略属性基加密（ciphertext-policy attribute-based encryption，CP-ABE）。一个基本的CP-ABE方案可以表示为以下

表3 数据胶囊需求与技术支撑

目标	需求	对应技术
权属可流转	执行过程机密且不被篡改	可信执行环境
	防止状态策略（如权属和When策略）被回滚	分布式账本
防复制	执行过程机密且不被篡改	可信执行环境
	Who策略，只有授权使用方才能访问； Where策略，只有在TEE中才能解开数据胶囊	属性基加密
	What、How策略，使用者只能获得结果，无法获得原始数据	策略执行

4 个算法[26]。

mpk,msk ← Setup(λ)：私钥生成器（PKG）以安全参数λ为输入，生成主公钥和主私钥对。

sk ← KeyGen(msk,S)：PKG 使用主私钥 msk 和解密者的属性集合S生成解密者的私钥 sk。

C ← Encrypt(mpk,M,P)：加密方使用主公钥 mpk 和访问策略P加密消息M，生成密文C。

M/ ⊥← Decrypt(sk,C)：解密密文C，其中，sk 是解密者的私钥。如果属性满足访问策略，sk 可以正确解密出消息M，否则解密失败。

4.2　具体协议设计

系统支持以下基本操作。

初始化阶段。用户（如数据提供方、数据使用方）与可信执行环境（如生产环境、消费环境）在数据中心完成注册、证书与密钥的分发及必要的初始化工作；同时初始化系统组件，例如生成 ABE 的主公钥与主私钥、初始化分布式账本 DT 等。

数据胶囊生成阶段。首先，数据提供方与生产环境完成远程认证并建立安全通道，随后向生产环境提交原始数据、对应策略及数据所有者标识。生产环境负责解析数据文件、进行数据加密与策略封装，然后将形成的数据胶囊推送至数据胶囊仓库 storage，并将其动态策略写入分布式账本 DT。操作完成后，生产环境向数据提供方返回数据胶囊标识符。

数据胶囊消费阶段。首先，数据使用方与消费环境完成远程认证并建立安全通道，随后向消费环境提交 ABE 私钥 sk_u、目标数据胶囊标识符 dcID、对数据胶囊的操作类型 op 以及处理函数f，其中操作类型包括数据胶囊使用（op = p）与数据胶囊加工（op = t）两类。消费环境首先依据胶囊标识符查询并验证该胶囊的动态策略，策略验证通过后，从数据胶囊仓库 storage 获取完整数据胶囊并校验其完整性。接着验证处理函数f是否满足身份策略及相应的使用或加工策略，验证通过则在消费环境中执行f并将计算结果或生成的数据胶囊通过安全通道返回数据使用方，验证不通过则直接拒绝操作请求。数据胶囊的使用或加工操作完成后，消费环境即时更新胶囊的动态策略，并将更新内容同步写回分布式账本 DT。

数据胶囊撤销阶段。数据胶囊所有者发起撤销请求，提交需撤销的胶囊标识符 dcID 给生产环境。生产环境核验请求方与胶囊所有者身份，验证通过后将该胶囊的可用次数设置为 0，并删除或使相关密钥失效。

图 4 为一种数据胶囊协议设计示例，展示了数据胶囊生成阶段生产环境针对数据提供方的输入的处理流程、数据胶囊消费阶段消费环境的具体处理流程，以及数据胶囊撤销阶段生产环境执行操作的伪代码。

4.3　协议分析

为了评估数据胶囊模型在功能实现、威胁模型覆盖与适用场景等方面的优势，本文对近年来相关领域的代表性工作[27-37]的对比分析如表 4 所示。其中，●表示相关工作支持对应特性，◐表示相关工作部分支持对应特性，○表示相关工作不支持对应特性，F1 表示保护数据所有权，F2 表示支持数据所有权转移，F3 表示细粒度访问与使用控制，F4 表示新鲜性，F5 表示异步流转能力，T1 表示防范恶意数据所有者，T2 表示防范恶意数据使用者，S1 表示数据共享场景，S2 表示数据交易场景。

在权属保护与转移机制方面，文献[27]与文献[28]通过核查机制或非同质化代币（non-fungible token, NFT）记录所有权变更，但在防范副本流转上存在局限；文献[29]、文献[30]及文献[31]未能严密保护所有权的唯一性，存在原始数据暴露风险。相比之下，本文方案将权属转移映射为数据胶囊的处置权更迭，支持数据提供方在转让后对胶囊进行重编辑或再交易，在保护原始数据所有权的同时，实现了权利主体的灵活转移。

在细粒度访问与使用控制方面，文献[25]与文献[20-21]仅支持静态策略，文献[32]的身份约束能力较弱，文献[33]与文献[34]分别实现了 XACML 形式与隐私预算策略，文献[29]的策略仅面向数据库场景，文献[30-31]仅支持属性基加密的访问策略，文献[27-28,35-36]的策略类型与执行机制相对有限。

新鲜性指数据及元数据始终保持最新状态而不易受回滚攻击。该特性主要针对动态数据与动态元数据场景。文献[33]与文献[32]分别利用区块链追加状态和共识协议防范回滚攻击，文献[37]则嵌入时间戳并采用密码学技术实现目标，文献[29]依赖数据库机制防止数据回滚，文献[34]设计了额外的完整性验证防范回滚攻击。其他工作只考虑了静态的数据和元数据，因此默认不支持新鲜性。多数方案

数据胶囊生成阶段

1) 用户输入 data, $P=P_1,P_2,P_3$ 给生产环境
2) $m \leftarrow \mathrm{ExtractMeta}(\mathrm{data})$ // 提取元数据
3) $k \xleftarrow{\$} \mathbb{Z}$ // 随机选择一个加密密钥
4) $C_d \leftarrow \mathrm{Enc}(k,\mathrm{data})$
5) $C_1 \leftarrow \mathrm{ABE.Encrypt}(\mathrm{mpk},k,P_1)$ //指定数据使用方身份属性
6) $C_2 \leftarrow \mathrm{ABE.Encrypt}(\mathrm{mpk},C_1,\text{"TEE"})$ //指定使用环境身份
7) $\mathrm{dc} \leftarrow (\mathrm{dcID},m,P_3,C_2,C_d)$ // 封装元数据、策略和数据
8) $S \leftarrow \mathrm{Sign}(\mathrm{sk}_p,\mathrm{dc})$ // 对数据胶囊签名，sk_p 为生产环境私钥
9) $\mathrm{dc} \leftarrow (\mathrm{dc},S)$
10) $\mathrm{DT.Insert}(\mathrm{dcID},P_2)$
11) $\mathrm{Storage.Insert}(\mathrm{dc})$
12) 返回 dcID 给用户

数据胶囊撤销阶段

1) 用户输入 dcID,ownerID 给生产环境
2) $P_2 \leftarrow \mathrm{DT.Query}(\mathrm{dcID})$
3) if P_2.ownerID!= ownerID then
4) //验证是不是所有者发起的请求
5) return false
6) end if
7) P_2.UseTimes← 0
8) $\mathrm{DT.Update}(\mathrm{dcID},P_2)$

数据胶囊消费阶段

1) 用户输入 $\mathrm{op},\mathrm{sk}_u,\mathrm{dcID},f$ 给消费环境
2) $P_2 \leftarrow \mathrm{DT.Query}(\mathrm{dcID})$
3) if !$\mathrm{Verify}(P_2)$ then // 验证动态策略
4) return $\perp$
5) end if
6) $\mathrm{dc} \leftarrow \mathrm{Storage.Query}(\mathrm{dcID})$
7) $(\mathrm{dc},s) \leftarrow \mathrm{dc}$
8) if !$\mathrm{Vrfy}(\mathrm{pk}_p,\mathrm{dc},s)$ then // 验证签名
9) return $\perp$
10) end if
11) $(\mathrm{dcID},m,P_3,C_2,C_d) \leftarrow \mathrm{dc}$ // 拆分数据胶囊
12) $C_1 \leftarrow \mathrm{ABE.Decrypt}(\mathrm{sk}_u,C_2)$
13) $k \leftarrow \mathrm{ABE.Decrypt}(\mathrm{sk}_c,C_1)$ // sk_c 为消费环境私钥
14) $\mathrm{data} \leftarrow \mathrm{dec}(k,C_d)$
15) if data ==$\perp$ then
16) return false // 解密不成功
17) end if
18) if $\mathrm{VerifyPolicy}(P_3,f,\mathrm{data}) == \mathrm{false}$ then
19) return false
20) end if
21) if $\mathrm{op} == p$ then // 使用操作
22) 返回 $f(\mathrm{data})$ 给用户
23) else 返回 $f(\mathrm{dc})$ 给用户 // 加工操作
24) end if
25) $\mathrm{UpdateState}(P_2)$ // 更新可使用次数等信息
26) $\mathrm{DT.Update}(\mathrm{dcID},P_2)$

图4 一种数据胶囊协议设计示例

表4 相关领域的代表性工作的对比分析

文献	时间	功能特性					威胁模型		适用场景		核心技术支撑
		F1	F2	F3	F4	F5	T1	T2	S1	S2	
文献[25]	2012年	●	○	◐	○	●	○	◐	●	○	信息流跟踪
文献[27]	2017年	◐	●	○	○	●	●	●	◐	●	核查机制
文献[20-21]	2019年	●	○	◐	○	●	○	●	●	○	TEE
文献[33]	2020年	●	○	●	●	○	○	●	●	○	区块链
文献[35]	2020年	●	○	○	○	○	●	●	●	○	区块链、TEE
文献[32]	2021年	●	○	◐	●	●	○	◐	●	○	分布式账本、TEE
文献[30]	2021年	○	○	◐	○	●	○	○	○	○	区块链
文献[37]	2021年	●	○	◐	●	●	○	●	●	○	同态加密
文献[23]	2023年	●	○	◐	○	●	○	●	●	○	区块链、安全多方计算
文献[29]	2023年	◐	○	◐	●	●	○	●	●	○	数据库
文献[34]	2023年	●	○	◐	●	●	○	●	●	○	TEE、差分隐私
文献[31]	2024年	○	○	◐	○	●	○	○	○	○	属性基加密
文献[36]	2025年	●	○	○	○	○	●	●	○	○	差分隐私、激励机制
文献[28]	2025年	◐	●	○	○	○	●	◐	○	●	区块链、NFT
本文方案	2026年	●	●	●	●	●	●	●	●	●	TEE、属性基加密、分布式账本

支持所有者上传数据后离线，便于异步交易和流转。相比之下，文献[33]、文献[35]及文献[28]等因需实时执行协议或双方协商，无法实现完全的异步流转。

在威胁模型维度，大多数工作假设数据所有者是诚实的，仅文献[27]、文献[28]及文献[36]引入了恶意所有者假设；文献[35]通过设置预期价格阈值

及信用激励机制，专门防范卖家的造假行为。相比之下，本文提出的数据胶囊模型确保打包过程在可信执行环境中进行且过程可证，所有者无法篡改封装逻辑，从底层杜绝了源头造假的可能性。在防范恶意使用者方面，文献[30]与文献[31]由于允许授权者解密原始数据，存在明显的副本扩散风险。文献[25]与文献[32]虽然利用TEE防止窥探，但在限制非授权使用方面仍显不足。在交易安全上，文献[28]与文献[27]分别通过智能合约和再交易限制来防范违约，但仍难以阻断链下传播。本文方案数据消费必须在受控的TEE内按策略执行，使用者既无法获取原始数据隐私，也无法脱离策略管控进行非法传播，从而在复杂威胁环境下实现了对恶意行为的有效闭环防范。

在适用场景方面，基本上所有工作都能实现数据使用权共享，但是文献[28, 30-31, 36]缺少对原始数据的安全性保护。文献[36]通过差分隐私保护了原始数据，但同时限制了数据的可用性，因而不能支持数据共享场景。要支持数据自由交易场景通常需要支持数据所有权转移，因此只有文献[27]、文献[28]和本文工作支持数据交易场景。

5　模型应用实践

5.1　支撑可信数据空间

数据胶囊模型能够融入可信数据空间体系，为数据权属流转与全流程可控提供了有效支撑，同时与隐私保护计算技术形成优势互补、协同赋能的体系关系。具体而言，数据胶囊模型中的生产环境与消费环境可以分别作为连接器的核心构件，在使用控制环境与策略执行机制的支撑下负责数据的封装与受控使用；而数据胶囊所依赖的分布式账本、公钥基础设施和数据胶囊仓库可作为服务平台的基础组件，承担使用控制模块所需的数据存储、策略管理与日志记录等功能。

此外，数据胶囊通过其细粒度的封装策略机制与策略执行设计，能够在所有权与加工使用权管理方面为可信数据空间提供有力补充。

5.2　典型场景

5.2.1　数据集交付

医院A拥有大量患者病历数据，存在数据商业化变现与患者隐私保护的需求。数据服务商B可为医院提供数据托管、安全加工与分销服务，并将合规数据产品供给药企及科研机构，支撑医药研发与学术研究。原始病历明文流转易造成隐私泄露；脱敏与差分隐私技术会削弱数据可用性，安全多方计算则协议定制成本高、部署复杂度大。

医院病历数据包括患者基本信息、诊断信息、用药记录、检查检验结果等内容。依托数据胶囊机制，医院可设置细粒度管控规则：该数据胶囊授权给B公司；B公司可以在不影响正确性的前提下对数据进行整理归纳并生成数据胶囊；姓名、身份证号和住址不得被查询。医院随后将密文数据与对应策略封装为数据胶囊。该数据胶囊的所有权被转移给B公司，B公司可在受控环境中对数据胶囊进行整理与加工，生成符合下游需求的数据产品胶囊并出售给制药企业或科研机构等数据使用方。

该应用模式可扩展至气象监测、交通测绘等敏感数据交易场景。借助数据胶囊的权属绑定与策略强制执行能力，本文方案在保障数据可用性与灵活流转的同时，有效保护隐私安全，支持数据所有者实现交易后的全周期管控。

5.2.2　特殊文件流转

除数据表等结构化数据外，针对工业CAD图纸、word文档等依赖专用软件打开编辑的非结构化文件，数据胶囊模型同样能够支撑其受控加工与安全复用。以工业CAD制图文件为例：数据权属方通常需将图纸标注纠错、版式规整等轻量化修改工作外包给第三方机构协作处理，同时严防图纸核心结构与涉密原始信息泄露。基于数据胶囊机制，文件可全程以密文形态完成跨主体传输；模型通过定制化安全策略严格约束操作权限，限定第三方加工后仅能输出衍生数据胶囊，杜绝原始图纸外流。此外，文件全量编辑操作将自动留存溯源日志，可供数据提供方全程审计监管，实现非结构化文件外包协作的安全可控流转。

5.2.3　数据合规共享

政务涉密数据阅后即焚、跨境数据流转等高安全应用场景，对数据的全周期延伸管控提出了更为严苛的约束要求。

以机密文件阅后即焚场景为例，发送方通过数字平台将敏感文件（如财务报告、法律协议草案等）传递给接收方，接收方仅在受限条件下（如限定时间、限定阅读次数）查阅，随后文件应在接收方设备及中转服务器上被自动删除。鉴于终端级控

制在实际中难以完全保证，数据胶囊模型主要聚焦于通过密钥管理与销毁机制来降低风险。依托数据胶囊模型，发送方可将敏感文件与访问策略统一封装，精准限定访问主体、单次阅览权限及90 s查看时长。接收方发起访问后，系统在受控环境内展示内容；当阅览超时或权限耗尽时，平台将自动锁定访问状态并销毁解密密钥，实现文件永久失效。

该管控机制同样可迁移应用至跨境数据等高敏感业务场景。依托软硬件协同的架构设计，数据胶囊模型既能够借助可信硬件强化运算隔离与密钥防护能力，也可联动安全擦除、受控浏览器等软件技术灵活适配需求，从而兼顾不同高安全场景下的防护标准与业务可用性。

5.2.4　个人隐私数据可控流转

在应用即服务场景中，应用通常需要收集大量用户信息。例如，游戏会收集身份证号用于实名认证与基于年龄的服务分级；电商会收集手机号与地址以支持物流配送；社交平台会收集联系人信息用于个性化推荐等。若平台在未获用户同意的情况下将数据出售或用于其他用途，则可能导致隐私泄露并使用户丧失对个人隐私数据的控制权。

数据胶囊模型可为个人信息构建可携带的数字身份胶囊，以保障敏感信息的机密性与可控流转。该数字身份胶囊可在不同应用间自由流转，同时对敏感字段的访问与使用由策略严格约束。以电商场景为例，用户向电商平台提交包含姓名、手机号、地址等字段的数字身份胶囊，策略可设为“电商平台不可解密；商户不可解密；快递公司仅能获取地址信息”。在该约束下，胶囊可在平台的商户与快递服务之间流转，但仅允许快递公司在物流配送中使用地址信息，禁止用于其他用途；电商平台与商户仅能关联胶囊标识符与订单信息，无法获得敏感信息。为实现可信执行环境的灵活性，生产环境可部署于用户手机，由手机厂商提供可信执行支持；用户可在手机端将个人数据与策略打包为数据胶囊并上传至云端数据胶囊仓库。

6　结束语

本文立足数据要素化现实困境，以“环境受控、数据有形、人有身份”三角模型为设计内核，以本体、权属、流通、定价四层理论体系为支撑，提出面向数据要素化的数据胶囊安全流通新范式，针对数据确权、流转控制与价值定价提出新的解决思路。相较于主流方案，数据胶囊在理论框架、技术适配、应用效率与全流程支撑上表现出一定优势，作为可信数据空间的有益补充，可广泛适配数据集交付、特殊文件受控流转、数据合规共享、个人隐私数据可控流转等实际场景。

参考文献：

[1] 陈刚, 颜斌斌, 汤珂. 数据的要素化与资产化: 理论辨析与实践探索[J]. 国际经济评论, 2024(5): 153-176.
Chen G, Yan B B, Tang K. Data elementization and assetization: theoretical analysis and practical exploration[J]. International Economic Review, 2024(5): 153-176.

[2] 李凤华, 李晖, 牛犇, 等. 数据要素流通与安全的研究范畴与未来发展趋势[J]. 通信学报, 2024, 45(5): 1-11.
Li F H, Li H, Niu B, et al. Research category and future development trend of data elements circulation and security[J]. Journal on Communications, 2024, 45(5): 1-11.

[3] 刘朝阳. 大数据定价问题分析[J]. 图书情报知识, 2016, 33(1): 57-64.
Liu C Y. Analysis on pricing of big data[J]. Document, Information & Knowledge, 2016, 33(1): 57-64.

[4] 江东, 袁野, 张小伟, 等. 数据定价与交易研究综述[J]. 软件学报, 2023, 34(3): 1396-1424.
Jiang D, Yuan Y, Zhang X W, et al. Survey on data pricing and trading research[J]. Journal of Software, 2023, 34(3): 1396-1424.

[5] Ruan P C, Dinh T T A, Lin Q, et al. LineageChain: a fine-grained, secure and efficient data provenance system for blockchains[J]. The VLDB Journal, 2021, 30(1): 3-24.

[6] Al-Mamun A, Yan F, Zhao D F. SciChain: blockchain-enabled lightweight and efficient data provenance for reproducible scientific computing[C]//Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2021: 1853-1858.

[7] İşler D, Cabana E, Garcia-Recuero A, et al. FreqyWM: frequency watermarking for the new data economy[C]//Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2024: 4993-5007.

[8] Qiao T, Liu X, Huang W K, et al. DSSmoothing: toward certified dataset ownership verification for pre-trained language models via dual-space smoothing[C]//Proceedings of the ACM Web Conference 2026. New York: ACM Press, 2026: 3111-3122.

[9] Zhou E Y, Guo S, Hong Z C, et al. Pistis: a decentralized knowledge graph platform enabling ownership-preserving SPARQL querying[J]. Proceedings of the VLDB Endowment, 2025, 18(11): 4602-4615.

[10] Wang S, Li Y R, Li H R, et al. Operon: an encrypted database for ownership-preserving data management[J]. Proceedings of the VLDB Endowment, 2022, 15(12): 3332-3345.

[11] Liu J W, Li Y T, Zhao M J, et al. EPFFL: enhancing privacy and fairness in federated learning for distributed E-healthcare data sharing services[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22(2): 1239-1252.

[12] Guo S Y, Zhang K Q, Gong B, et al. Sandbox computing: a data privacy trusted sharing paradigm via blockchain and federated learn-

ing[J]. IEEE Transactions on Computers, 2023, 72(3): 800-810.
[13] Poddar R, Kalra S, Yanai A, et al. Senate: a maliciously-secure MPC platform for collaborative analytics[C]//Proceedings of the 30th USENIX Security Symposium. Berkeley: USENIX Association, 2021: 2129-2146.
[14] Fang W J, Cao S D, Hua G J, et al. SecretFlow-SCQL: a secure collaborative query platform[J]. Proceedings of the VLDB Endowment, 2024, 17(12): 3987-4000.
[15] Liu C C, Guo H C, Xu M H, et al. Extending on-chain trust to off-chain: trustworthy blockchain data collection using trusted execution environment (TEE) [J]. IEEE Transactions on Computers, 2022, 71 (12): 3268-3280.
[16] Dhasade A, Dresevic N, Kermarrec A M, et al. TEE-based decentralized recommender systems: the raw data sharing redemption[C]//Proceedings of the 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Piscataway: IEEE Press, 2022: 447-458.
[17] Xue T, Wen Y, Luo B, et al. SparkAC: fine-grained access control in spark for secure data sharing and analytics[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(2): 1104-1123.
[18] Lin Q Q, Zhang J Y, Liu J F, et al. Demonstration of dealer: an end-to-end model marketplace with differential privacy[J]. Proceedings of the VLDB Endowment, 2021, 14(12): 2747-2750.
[19] Si M, Pei J. Counterfactual explanation of shapley value in data coalitions[J]. Proceedings of the VLDB Endowment, 2024, 17(11): 3332-3345.
[20] Wang L, Near J P, Somani N, et al. Data capsule: a new paradigm for automatic compliance with data privacy regulations[C]//Proceedings of Heterogeneous Data Management, Polystores, and Analytics for Healthcare. Cham: Springer International Publishing, 2019, 11721: 3-23.
[21] Wang L, Khan U, Near J P, et al. Privguard: privacy regulation compliance made easier [C]//Proceedings of the 31st USENIX Security Symposium. Berkeley: USENIX Association, 2022: 3753-3770.
[22] FATE. FATE: an industrial grade federated learning framework[R]. [2026-04-02].
[23] Lu T P, Zhang B S, Ren K. PrivData network: a privacy-preserving on-chain data factory and trading market[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(3): 1424-1436.
[24] PlatON Network. PlatON: a next-generation hierarchical privacy-preserving computing architecture[R]. [2026-04-02].
[25] Chen Y Y, Jamkhedkar P A, Lee R B. A software-hardware architecture for self-protecting data[C]//Proceedings of the 2012 ACM Conference on Computer and Communications Security. New York: ACM Press, 2012: 14-27.
[26] Waters B. Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization[C]//Proceedings of the 14th International Conference on Practice and Theory in Public Key Cryptography. Berlin: Springer, 2011: 53-70.
[27] Jung T, Li X Y, Huang W C, et al. AccountTrade: accountable protocols for big data trading against dishonest consumers[C]//Proceedings of the IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017: 1-9.
[28] Zhang D M, Xie L, Tao Y, et al. NMFT: a copyrighted data trading protocol based on NFT and AI-powered merkle feature tree[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 13041-13054.
[29] Albab K D, Sharma I, Adam J, et al. K9db: privacy-compliant storage for Web applications by construction [C]//Proceedings of the 17th USENIX Symposium on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2023: 99-116.
[30] Soltani R, Nguyen U T, An A J. Data capsule: a self-contained data model as an access policy enforcement strategy[C]//Proceedings of the 2021 3rd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS). Piscataway: IEEE Press, 2021: 93-96.
[31] Lyu Q Y, Zhou Y L, Ren Y Z, et al. Toward personal data sharing autonomy: a task-driven data capsule sharing system[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 9760-9774.
[32] Gao M Y, Dang H, Chang E C. TEEKAP: self-expiring data capsule using trusted execution environment[C]//Proceedings of the Annual Computer Security Applications Conference. New York: ACM Press, 2021: 235-247.
[33] Truong N B, Sun K, Lee G M, et al. GDPR-compliant personal data management: a blockchain-based solution[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 1746-1761.
[34] Liu H, Luo H Z, Li S F, et al. Privacy computing with right to be forgotten in trusted execution environment[C]//Proceedings of the GLOBECOM 2023 - 2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 2566-2571.
[35] Dai W Q, Dai C K, Choo K R, et al. SDTE: a secure blockchain-based data trading ecosystem[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 725-737.
[36] Zhang M, Li X H, Luo B, et al. An incentive mechanism for privacy preserved data trading with verifiable data disturbance[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22(4): 3960-3976.
[37] Burkhalter L, Küchler N, Viand A, et al. Zeph: cryptographic enforcement of end-to-end data privacy[C]//Proceedings of the 15th USENIX Symposium on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2021: 387-404.

[作者简介]



王锐璇（2000– ），女，南开大学博士生，主要研究方向为数据安全和可信执行环境。



刘哲理（1978– ），男，博士，南开大学教授、博士生导师，主要研究方向为数据隐私保护和人工智能安全。