数据要素流通与安全的研究范畴与未来发展趋势

李凤华[1,2,3]，李晖[4]，牛犇[1,3]，邱卫东[5]

（1. 中国科学院信息工程研究所，北京 100085；2. 中国科学院大学网络空间安全学院，北京 100049；
3. 网络空间安全防御重点实验室，北京 100085；4. 西安电子科技大学网络与信息安全学院，陕西 西安 710126；
5. 上海交通大学网络空间安全学院，上海 200240）

摘　要：针对数据从信息技术时代的受控共享向数据技术时代的数据要素泛在流通演化的趋势，分析了数据共享与数据要素流通的差异，阐述了什么样的数据才能成为数据要素，提出了数据成为数据要素所必须具有的6个属性，定义了数据要素流通模型及主要环节，明确了数据要素流通的研究范畴，梳理了数据要素流通研究范畴的相关概念，厘清了研究范畴所涵盖的关键核心技术，并对未来需要突破的关键技术进行了展望。

关键词：数据要素流通；数据确权；数据安全计算；隐私计算

中图分类号：TP393

文献标志码：A

DOI：10.11959/j.issn.1000–436x.2024106

Research category and future development trend of data elements circulation and security

LI Fenghua[1,2,3], LI Hui[4], NIU Ben[1,3], QIU Weidong[5]

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China
3. Key Laboratory of Cyberspace Security Defense, Beijing 100085, China
4. School of Cyber Engineering, Xidian University, Xi'an 710126, China
5. School of Cyber Science and Engineering, Shanghai Jiaotong University, Shanghai 200240, China

Abstract: Towards the trend of evolution from controlled data sharing in the information technology era to ubiquitous circulation of data elements in the data technology era, the difference between data sharing and data element circulation was analyzed, what kind of data could become a data element was explained, and six attributes that data must have to become a data element were put forward. The data element circulation model and its primary links were defined, and the research scope of data element circulation was clarified. Furthermore, the relevant concepts of data element circulation were combed through, the important key technologies were analyzed, and the key technologies that need to be solved in the future were outlooked.

Keywords: data element circulation, data rights determination, secure data computation, privacy computing

0　引言

数据要素的有序流通是数字经济高质量发展的内在驱动力，国家及各地政府高度重视数据要素流通的政策及技术保障，正加快完善数据要素流通相关政策法规，构建制度框架下的数据要素流通行为范式，推动数据要素市场的高效运行。

收稿日期：2024–03–06；**修回日期：**2024–04–30

通信作者：李晖，lihui@mail.xidian.edu.cn

基金项目：国家重点研发计划基金资助项目（No.2023YFB3106500）；国家自然科学基金资助项目（No.61932015）

Foundation Items: The National Key Research and Development Program of China (No.2023YFB3106500), The National Natural Science Foundation of China (No.61932015)

2020年，中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，正式将数据与土地、劳动力、资本和技术并列为五大生产要素。2021年，《“十四五”数字经济发展规划》明确指出要充分发挥数据要素作用，强化高质量数据要素供给，加快数据要素市场化流通，创新数据要素开发利用机制，鼓励市场主体探索数据资产定价机制，培育规范的数据交易平台和市场主体，建立健全数据资产评估、登记结算、交易撮合、争议仲裁等市场运营体系。2022年，中共中央、国务院发布《关于构建数据基础制度更好发挥数据要素作用的意见》（又称“数据二十条”），提出要建立合规高效的数据要素流通和交易制度，规范数据流通规则，构建完善的数据流通基础设施和交易场所，培育数据要素流通和交易服务生态，有序发展数据跨境流通和交易，建立数据来源可确认、使用范围可界定、流通过程可追溯、安全风险可防范的数据可信流通体系。2023年，《数字中国建设整体布局规划》指出数字技术与经济、政治、文化、社会、生态文明建设的深度融合将带动数据要素在各场景发挥独特作用，充分激活数据要素内在价值。《党和国家机构改革方案》提出要推进多层次数据要素市场建设，促进数据要素、数字经济与实体经济的深度融合。国家数据局会同中央网信办等17部门联合印发了《“数据要素×”三年行动计划（2024—2026年）》，发挥数据要素的放大、叠加、倍增作用，构建以数据为关键要素的数字经济，推动高质量发展。选取工业制造、现代农业、商贸流通、交通运输、金融服务、科技创新、文化旅游、医疗健康、应急管理、气象服务、城市治理、绿色低碳等12个行业和领域，推动发挥数据要素乘数效应，释放数据要素价值。

紧跟国家的政策规划，全国各省市陆续出台政策法规，对数字经济发展、数据要素制度建立和数据要素市场化流通等展开了一系列规划，广东、贵州、上海、海南、北京等省市出台了相应的数据条例。北京、上海、深圳、广州、江苏、浙江、湖北、陕西、黑龙江等省市先后设立了本地的大数据交易所。

以美欧为代表的发达国家对数据贸易活动问题的探索起步较早，发布了一系列聚焦数据开放共享和有序流通的政策法规。2009年，美国联邦政府发布《政府数据开放倡议》，并建立政府数据产品平台Data.gov，旨在鼓励政府部门之间、政府与公民社会组织之间的数据共享与合作，促进更广泛的数据流通。2013年，美国发布的《开放数据政策》鼓励开发者对公开医疗、教育、经济、农业等七大领域的数据进行二次开发和加工，旨在促进数据的开放性和可利用性，以支持创新和经济发展。2020年，美国总务署发布了《数据伦理框架草案》[1]，详细说明美国联邦政府在数据使用方面的标准和措施。此外，美国先后颁布了《信息安全港框架协议》《个人信息保护法案》《澄清境外合法使用数据法案》等，规范企业对数据的处理和跨境传输。

2018年正式生效的欧盟《通用数据保护条例》[2]强调数据控制者必须确保对数据的合法处理，并承担起对数据安全和隐私保护的责任，规定了跨境数据传输的限制和要求，确保将个人数据传输至国外时也要受到充分的保护。2020年，欧盟委员会发布《欧洲数据战略》和《欧洲人工智能白皮书》，着眼于加强数字基础设施，确保数据的安全、互操作性和可持续性，促进数据合法、安全的跨境流通，以推动创新和经济增长。2022年，欧盟通过的《数据治理法案》[3]旨在建立健全公共数据共享机制，促进整个欧盟内部和跨部门之间的数据共享，增加企业和个人对数据中介服务的信任，并为主要技术平台的数据处理实践提供一种新的欧洲模式，建立关于数据市场中立性的新规则，促进公共数据（例如健康、农业或环境数据）的再利用，并在战略领域创建共同的欧洲数据空间。2023年，欧盟通过的《数据法案》[4]界定了公共部门使用数据主体的相关数据的权利和约束，明确了数据访问、共享和使用的规则，规定了获取数据的主体和条件，完善了企业对政府数据共享规则的结构和专用功能，推动了产业价值链上的企业数据流通与共享，使更多私营和公共实体能够共享数据。2023年7月10日欧盟委员会通过了《欧盟-美国数据隐私框架》的充分性决定，要求国家监控机构执行更严格的数据访问标准，确保从欧盟转移到美国公司的个人数据具有与欧盟相当的充分保护水平。这使得个人数据从欧盟传输至经过框架认证的美国公司时，从法律上约束并预防了跨国数据访问的个人数据保护水平下降。

当前我国数据要素流通发展存在许多亟待解决

的问题，面向数据资产易复制、权属属性易损毁等特点，缺乏可容毁的数据确权和仲裁机制；数据交易过程中需要解决动态多方可信磋商机制与契约可信签署、契约约束的交割协同执行、一致性核验与权属转移等问题，也缺乏数据交易回退时权属返还和可验证删除技术；同时，亟待解决全生命周期数据使用控制的难点问题，构建数据要素流通全过程的低开销抗泄露防篡改存证机制，以支撑争议仲裁。此外，数据要素流通过程中的数据保护、隐私保护需要探索以数据安全计算、隐私计算为支撑的解决方案。传统的网络安全等级保护、云计算服务安全评估、商用密码应用安全性评估等关注的主要内容不属于数据要素流通与安全的研究范畴，需要在现有基础上研究增量和支撑增量构建所需技术体系中必备的共性技术，促进数据要素的更广泛流通和利用，推动数字经济的发展。

本文首先阐述数据要素流通的内涵，包括数据与数据要素的定义、数据成为数据要素的必要属性、数据共享与数据要素流通的区别、数据确权、数据交易等；然后介绍数据要素流通模型及主要环节，数据要素流通的研究范畴及关键技术；最后对数据要素流通的技术进行展望。

1　数据要素流通的内涵

1.1　数据与数据要素的定义

随着时代的演化，数据的含义不断演化。在拉丁文中，“数据”（Datum）的含义是“已知”或“事实”，为现实世界的客观记录。在计算机科学中，数据指描述客观事物且能被计算机程序处理的符号集合。2021年6月颁发的《中华人民共和国数据安全法》将数据定义为“任何以电子或者非电子形式对信息的记录”。

数据要素是指将原始数据通过加工整理、确权，使其成为具备潜在利用价值的数据资产，并通过在市场上交易流通，让这些数据资产成为可用于社会生产经营活动，能够推动社会生产力发展，并为使用者带来经济效益的重要生产要素。数据要素已成为推动数字经济发展的核心引擎，赋能行业数字化转型和智能化升级，成为国家基础性战略资源。

1.2　数据成为数据要素的必要属性

2023年1月，本文作者李凤华在学术报告中首次提出了数据成为数据要素必备的6个属性，即可用性、机密性、隐私性、可控性、交易性和仲裁性。

1) 可用性。在数据流通过程中需要确保数据随时随地满足授权用户合规使用的数据质量。数据如果只在特定场景下达到使用质量，只能是一种有限的共享，可用性则强调需要随时随地满足数据的使用质量指标。

2) 机密性。在数据流通过程中需要确保数据不被非授权留存，在传输过程中不被授权用户截取。

3) 隐私性。确保个人敏感信息（包括指纹、虹膜、身份证号、电话号、住址、过敏信息、疾病和药品使用状况、犯罪信息、出行、监控信息、财务状况、交通工具、信用记录、购买记录、品牌爱好、社交账号等）在数据流通中不被非授权留存、发布、泄露和使用。隐私性是在机密性基础上还强调个人信息流通与利用中的保护，更多强调的是流通与利用中的脱敏。

4) 可控性。确保数据在流通过程中全生命周期的使用可控，包括：访问、加工、删除、脱敏、流转管控、边界过滤、追踪溯源、违规判定、审计、取证等操作的可控。

5) 交易性。在数据流通过程中需要在确保数据合规、真实、准确的基础上，评估数据价值，定价数据资产，并进行权属转移。数据交易要求数据作为资产入表，就必须定价、交付，所以数据就必须出域。数据不出域不可能成为数据流通。

6) 仲裁性。数据在数据流通过程中发生争议时，采用证据交叉认证、审计等方式进行仲裁。在数据流通过程中发生争议不可避免，有争议就必须要仲裁，需要有可信第三方随时随地不间断支持仲裁的方案，并且保证不同政策体系在此基础上精准高效运行。

1.3　数据共享与数据要素流通的区别

数据共享是指在不同的组织、机构或个人之间，按照一定的规则和标准，相互开放和交换数据资源的过程，其目的是提高数据的利用效率，打破信息孤岛，促进跨部门、跨领域、跨地区的合作与协同。

数据要素流通是指将数据作为产品进行分类定价、流通和买卖，是数据要素价值和作用发挥的重

要途径，包括数据的确权、交易、使用及监测等环节。数据要素流通的市场环境涉及数据供需、数据安全、数据授权和数据交易规则等多方面因素。

数据受控共享重点关注机密性、完整性、访问控制，支持跨系统协同、移动办公等。数据要素流通重点关注权属确定、权益转移、使用验证、争议仲裁等。

1.4　数据确权

数据确权是通过对数据主体赋权使其对数据享有相应的法律控制手段，从而在一定程度或范围内对数据具有排除他人侵害的效力。鉴于数据可复制的特性，传统的物理阻隔方法难以有效地保护数据的获取和使用。因此，必须依靠对数据进行确权，否则很难保护相关主体的数据权益。

现有的数据确权主要是基于文件哈希等简单的技术，仅支持对确权文件的哈希值进行哈希存证，以达到确权的目的。由于哈希算法的“雪崩效应”，使得多维度海量数据确权变得不可实现，在实际应用中迫切需要一种支持数据权属属性容毁的确权机制。

1.5　数据交易

数据交易是数据需求方通过支付费用从数据提供方获取数据的过程。数据交易涵盖了数据要素交易方、监测方在要素价值、数据权属、交易内容、契约细节、履约验证等方面的交易关联内容。

数据交易的本质是数据的权益转移和使用授权，数据要素频繁流通会因保管不善、越权滥用、脱敏防护不够等因素导致隐私泄露。数据交易的安全则主要涉及交易攸关方的匿名性、可认证性、权限控制，交易内容的机密性、完整性、不可否认性，交易契约的不可篡改性、可验证性等内容。

2　数据要素流通模型及主要环节

数据要素流通是在网络通信基础设施、数据流通基础设施上确保数据符合预期的交易与使用，数据要素流通模型如图1所示，其环节包括采集与治理、数据确权、数据资产发布、数据交易与安全交付、数据使用、数据销毁、流通存证与仲裁。其中，采集与治理是数据持有者对数据进行清洗、分类分级、元数据管理等操作，便于后续的数据流通利用；数据确权由权威机构确定数据归属，并在数据所有权和权益权纠纷时，利害关系人可以通过仲裁的方式解决争议，其关键技术包括权属属性提取、权属可信登记、权属离线验证、抗毁的权属仲裁等；数据资产发布是将确权后的数据资产在交易平台发布，从而方便数据要素的交易，其关键技术包括数据目录可信构建、资产价值估值等；数据交易与安全交付是指在数据交易过程中攸关方依据交易方意愿磋商交易契约，并根据契约准确完整及时地把数据交付给数据使用方，其关键技术包括与攸关方信誉评估、交易智能撮合、履约自动核验、交易数据安全撤销、权属回退等；数据使用是指数据

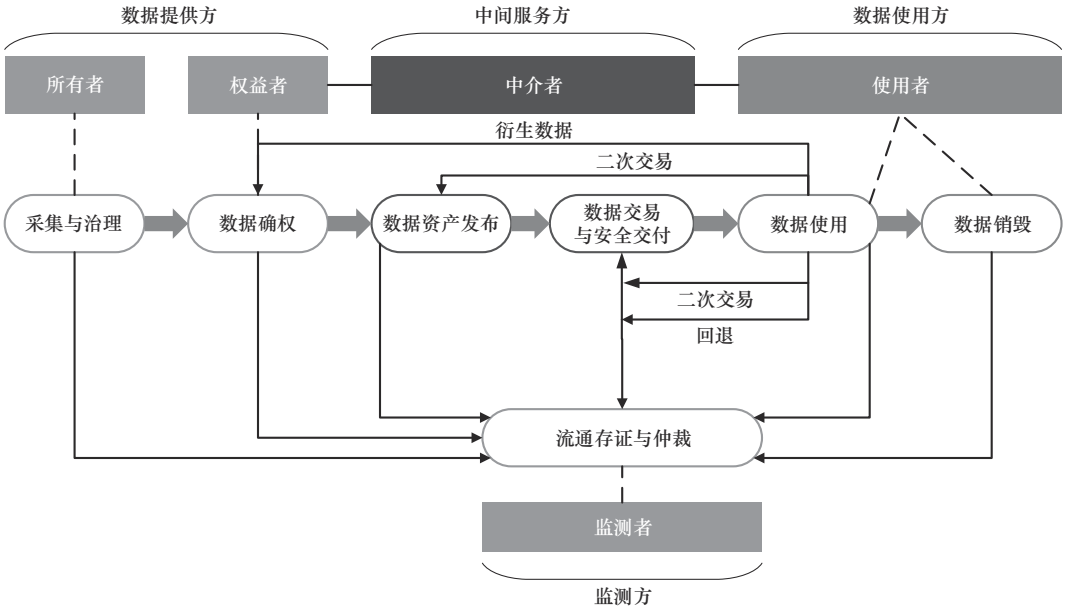

图1　数据要素流通模型

使用方根据契约约定对数据进行加工利用，实现数据增值，数据使用应确保在流转全流程中按照契约约束实现使用控制，并能对使用过程进行远程验证，数据可以根据契约再次发布并实现二次交易，加工后的衍生数据可以再次确权进入数据要素流通；数据销毁是在数据达到使用目的后根据契约或者数据所有者要求销毁数据；流通存证与仲裁是在数据要素流通的全流程中实现存证，并且在发生纠纷时利害关系人可以通过仲裁的方式解决争议。

3　数据要素流通的研究范畴及关键技术

3.1　数据要素流通的研究范畴

数据要素流通需要传统技术拓展应用和新技术创新并重，在价值评估、数据利用、安全保障、数据交易、资产处置等方面形成一系列创新机制。为此，梳理数据要素流通的研究范畴包括安全可信交易、数据可控利用、数据按需保护、隐私按需保护、流通利用监管 5 个部分。

3.1.1　安全可信交易

安全可信交易包括数据交易磋商、数据安全交付、数据交易撤销、资产权益转移、数据流转控制等技术范畴。数据交易磋商的安全机制为交易攸关方达成不可抵赖、不可篡改的买卖契约提供安全保障；数据安全交付为买方收到的数据与购买协议对应的标的物一致、买卖双方关于数据交付过程的不可否认提供安全机制；数据交易撤销为交易双方在交易进行中、交易完成后提供友好协商、回退到交易前状态的机制；资产权益转移为数据资产交易过程中的资产权益属性变更、变更的可追溯性提供安全支撑；数据流转控制为数据传播路径、传播范围提供控制机制。

数据交易磋商通过交易流程的定制与重构、多方动态磋商与交易契约签署，实现交易流程柔性重组，防止交易信息泄露、交易信息篡改、交易过程欺诈等问题，实现多方交易磋商过程的公平性、机密性、不可篡改和不可抵赖，支撑数据资产交易契约可信签署与安全流通共享。

数据安全交付采用契约约束的多方制衡交付、跨域交割执行、交付数据权属验证、交割执行一致性验证等技术措施，解决交易数据的防窃取与篡改、多方协同跨域可信交付等问题，实现交易数据交付流程可定制、可核验、可追溯，支撑交易数据的可信可控交付与权益转移。

数据交易撤销是数据交易的售后保障、交付数据按需退换时，采用契约废止后的多方动态磋商、退还数据可验证删除、数据权属同步返还等技术措施，解决数据资产交付后的交易撤销可信可控等问题，实现数据资产回退、交易状态回滚以及数据资产权益返还，支撑数据交易解约。

资产权益转移采用多方协同权属证书更新、交付数据资产权益跟踪、权益状态一致性核验等技术措施，解决交易后的数据权属可信变更、权属变更可追溯、权益变更重发布等问题，实现可信可靠的权属证书更新、发布，支撑数据资产可信转移、数据二次交易及交易回退。

数据流转控制采取权限按需调整、异常行为感知、多源融合分析等技术措施，对跨系统流入流出、网络传输路径上的交易数据流转行为进行检测、实时分析与监控，支撑交易数据的全流程管控以及违规溯源。

3.1.2　数据可控利用

数据可控利用包括数据分类分级、数据容毁确权、资产登记发布、数据使用控制、数据增值利用等技术范畴。数据分类分级用来区分数据类别，支撑数据资产目录建设和资产入表，还用于区分数据的敏感性及其管控措施；数据容毁确权用于界定海量数据的权属，是数据交易的前提，支撑数据资产入表；资产登记发布用于多源目录自动归集，实现数据资产快速检索；数据使用控制用于限定数据接收方按契约使用，控制数据使用范围，约束二次传播，解决使用控制的远程验证问题；数据增值利用用于数据使用方根据交易契约的约束，对数据要素进行统计、加工、模型训练等。

数据分类分级指把具有相同属性或特征的数据归成同一类别，并给出类别标签和分级标签，解决海量异构数据的目录构建、数据敏感性表征、基于分类分级的数据流转与使用控制等问题，支撑数据登记、发布，以及数据按需保护、隐私按需脱敏、数据流转控制和数据利用。

数据容毁确权采用数据特征刻画和容毁判定、数据权属登记证书、权益变更项登记等技术措施，解决海量数据的特征库构建、近似度高的数据权属区分、数据篡改仲裁、权属登记证书与数据关联关系不可分离、权属界定与可信变更等问题，实现部

分权属属性损毁情况下的可容毁可仲裁的数据确权，支撑数据流通与利用过程中的权益争议裁决。

资产登记发布对来自不同机构的数据目录自动合并，对权属属性在可控范围内的进行发布并动态调整，实现确权数据安全登记、数据目录可信发布、发布信息一致性检测，支撑数据使用方搜索和查询服务。

数据使用控制采用交易契约关联的数据使用策略、数据使用策略与数据的安全绑定、数据使用策略可信执行、策略执行远程验证等技术措施，解决数据域内非授权访问、数据域外使用不受控等问题，实现对数据在使用方的使用控制及远程验证，支撑交易后数据依照合约使用。

数据增值利用采用包括信用计算、联邦学习、大模型等技术对数据进行加工，实现数据要素的增值，支撑数据使用方根据交易契约的约束对数据要素实现有效利用。

3.1.3 数据按需保护

数据按需保护包括身份交叉认证、密码技术应用、机密计算、密文计算、安全多方计算等技术范畴。身份交叉认证贯穿于数据流通利用的全流程；密码技术应用贯穿于通信协议和应用协议、传输安全、存储安全等环节或业务系统中；机密计算和密文计算为数据利用的计算过程提供安全保障；安全多方计算为数据交换和计算结果交换过程提供安全保障。安全多方计算和密文计算通常联合使用，机密计算和密文计算需要依据算力的能效比来合理选择。

身份交叉认证面向现有数据流通方式中认证系统模式多样、认证机制差异、认证接口不统一等特点与问题，采用统一身份认证系统，实现分布式多技术体制的数据流通利用设施的身份交叉验证、分域接入认证、安全隔离、匿名访问、信息访问动态授权等，解决人机认证、设备/网络接入可控问题，支撑信息访问授权与权限调整。

密码技术应用面向数据流通的高通量安全传输、多源多模态数据联合计算的高并发与低时延等应用需求，采用高能效密码计算架构、优化加解密流量控制策略、密码服务高并发调度等技术措施，解决高效模型推理、密码服务带宽占用、透明加密存储等问题，提高密码处理的吞吐量和宽带利用率，支撑数据高速安全流通和后台基础设施应用的高性能密码服务需求。

机密计算在数据处理过程中将敏感数据隔离在被称为可信执行环境（TEE, trusted executive environment）的受保护区域中，并在可信执行环境中处理数据，支撑对计算环境的可信性、可控性和机密性有特殊要求的应用。

密文计算适用于数据利用的多方信息出域且计算环境不可控的情况下进行联合计算、统计分析、外包计算、密文数据检索等场景，采用密文出域明文不出域的方式，参与方在密文上计算，保护参与方参与运算的原始数据，并安全分享计算结果，支撑数据出域计算时原始数据的安全性防护和数据加密存储时的统计查询。

安全多方计算在数据利用的多方信息安全共享过程中，针对数据提供方中未分享的原始数据不出域需求，采用不经意传输、秘密分享、混淆电路、同态加密等密码技术，通过多轮信息秘密交换的数据安全计算，实现数据计算或交换过程中未分享的原始数据不出域，支撑计算结果安全共享。

3.1.4 隐私按需保护

隐私保护包括隐私计算、隐私信息识别、隐私信息求交、隐私挖掘评估、个人信息合规审计等技术范畴。在隐私计算中，根据接收方保护能力或者交易价格进行按需脱敏交付，延伸控制保障数据流通与利用过程的按需迭代脱敏控制和删除控制，主要解决数据出域的隐私保护问题；隐私信息识别用于对数据的隐私进行判断，以支撑数据分类和结合场景的分级；隐私信息求交实现买卖交易之前匿名匹配双方感兴趣的内容；隐私挖掘评估支撑脱敏效果评估；个人信息合规审计是用于评估数据流通利用设施对个人信息防护的合规程度，主要评估知情权、脱敏和删除的合规性。

隐私计算由本文作者李凤华、李晖等[5-6]提出，面向隐私信息泛在流通利用、跨域受控共享等应用需求，采用隐私动态度量、按需脱敏控制、迭代延伸控制、保护效果评估等技术措施，解决差异化脱敏、脱敏效果评估、数据交易的差异化定价与脱敏、多轮交易价格差异化联动、隐私传播控制、跨域隐私信息延伸控制等问题，实现隐私信息全生命周期保护，支撑隐私信息合规流通利用。

隐私信息识别抽取海量隐私信息特征，采用关键词匹配、自然语言理解等技术措施实现多模数据

的隐私识别与发现，支撑实现隐私信息的分类分级，并以此为基础，支撑数据流通利用过程中的延伸控制、信息流转控制、脱敏与删除、数据定价与多轮交易。

隐私信息求交是两方或者多方在不暴露自身原始数据的情况下实现共有隐私信息发现的技术，解决数据发现的隐私泄露问题，支撑数据发现、数据联合利用与建模。

隐私挖掘评估指对脱敏后的数据进行大数据关联分析，以发现其中是否包含被脱敏的隐私信息。隐私挖掘评估通过逆向方式判断隐私脱敏效果，以发现隐私泄露的潜在风险，支撑数据流通利用中的隐私信息脱敏技术迭代。

个人信息合规审计收集数据处理过程中的行为信息，并与其他相关日志审计信息进行多源融合分析，给出数据流通过程中存在的潜在安全风险，实现数据流通利用的合规检查与安全审计，支撑数据流通利用过程中的风险识别与预警。

3.1.5　流通利用监管

流通利用监管包括全流程可控存证、流通风险评估、流通争议取证、违规事件溯源、异常行为处置等技术范畴。全流程可控存证支撑异常行为发现和个人信息保护的合规审计；流通风险评估支撑数据流通利用设施的风险评估与预警、全域的风险评估与预警；流通争议取证支撑跨数据流通平台的证据交叉认证、证据链生成，并提升证据的公信力，有效支撑违规溯源；违规事件溯源支撑安全事件的溯源与追责；异常行为处置在线阻断异常行为、安全事件蔓延。

全流程可控存证是指在数据流通的确权、发布、交易交付、使用等环节记录和存储数据要素在某个时间点、某个信息系统的流转状态以及操作记录等，通过低开销、抗泄露、防篡改的存证方法解决存证信息防篡改、传输过程抗泄露、传输带宽高消耗的问题，实现数据流通全流程的全域存证，支撑数据融合分析、争议仲裁和违规溯源。

流通风险评估关注流通过程中数据资产面临的威胁、存在的弱点以及这些因素综合作用可能带来的风险，给出相关系统的风险态势与预警，为风险安全防护应对预案提供基于证据的信息和依据。

流通争议取证是利益攸关方在数据流通过程中发生权益冲突时，对存证信息通过关联分析形成证据链的过程，解决数据提供方、中间服务方、数据使用方等多方跨系统证据交叉验证、证据链构建等问题，实现权益冲突时争议仲裁信息的可信获取，支撑责任判定、争议仲裁和违规溯源。

违规事件溯源对存证信息进行挖掘获取线索，通过多源证据融合分析，构建数据违规流转血缘关系链，解决证据多源零碎、泄露点离散动态的违规事件精准定位问题，实现违规行为判定与溯源，支撑违规追责。

异常行为处置针对数据非法访问、篡改和破坏等异常行为，采取敏感信息识别与细粒度过滤、网络整体或信息部分阻断等技术措施，实现数据流通异常行为全域多点协同处置、处置效果研判等功能，支撑异常行为的威胁最小范围封闭与在线阻断，阻止异常流转信息在全域蔓延。

3.2　安全可信交易的关键技术

安全可信交易的关键技术包括交易磋商与安全交付、权益转移与交易回退等。

3.2.1　交易磋商与安全交付

交易磋商与安全交付需要解决数据交易动态多方可信磋商机制与契约可信签署、契约约束的交割协同执行与一致性核验等关键问题。

1) 数据交易动态多方可信磋商机制与契约可信签署

数据交易需要建立可信的磋商机制和契约签署方式，确保公开透明的协商过程，防止信息篡改和欺诈行为，满足法律法规的要求，减少纠纷发生，可采用密码学技术和不可篡改存证等机制，确保交易安全，有效防止数据泄露和篡改的风险。多方可信磋商机制的实现需要建立参与方的身份信任，在此基础上通过密码协议确保参与者之间的安全通信、数据交换以及协议执行的公平性。

2) 契约约束的交割协同执行与一致性核验

根据契约约束构建协同执行模型，在协同执行的过程中增加实时访问和验证协调执行的状态操作，减少一致性核验时间，确保交割按照契约规定进行，提升协调执行的透明度、自动化和高性能。契约创建后，各方应制定契合自己的协同执行计划，明确自己的责任和角色，以确保契约规定的条件和期限得以满足。通过智能合约技术，各方根据契约规定的条件编码智能合约，并根据契约履行自己的义务，最终通过一致性验证节点验证交割协同

执行的一致性，确保契约的规定得以遵守，交易没有被篡改。

3.2.2 权益转移与交易回退

依据交易磋商得到的契约，数据资产交付需要解决权益的可信转移以及交易回滚情况下的数据跨域可验证删除与权属同步返还等关键问题。

1) 权益的可信转移

权益转移主要涉及确权系统、原数据权属者、新数据权属者3个主体。确权系统负责颁发与更新权属证书，原数据权属者希望转移某数据的权属者身份，新数据权属者希望将某数据权属者身份转移给自己。

权益转移包含3种情况，第一种情况是在原数据权属者基础上追加新数据权属者，第二种情况是新数据权属者替换原数据权属者。这两种情况的权益转移都需要原数据权属者向确权系统发出同意转移证明，由确权系统签发新的权属证书并发布，同时作废旧的权属证书。第三种情况是数据副本的权益许可授权，原数据权属者向确权系统发出许可授权证明，由确权系统签发原权属证书的副本并发布。

2) 交易回滚情况下的数据跨域可验证删除与权属同步返还

交易回滚是指在某些情况下，交易需要撤销或回退到交易之前的状态。在这种情况下，需要确保涉及的数据可以跨域进行可验证删除，并确保权属的同步返还。

为实现上述需求，可以通过全流程存证对数据操作进行记录。当交易回滚时，数据使用方对数据的删除操作保存在存证系统中，并向数据权属者提供可验证的证据，同时通过确权系统在数据权属证书中增加回退的记录，确保交易所涉及的权属信息同步返还给原数据权属者。通过交易存证系统建立监督和审计机制，确保交易回滚过程的透明度和可追溯性。

3.3 数据可控利用的关键技术

数据可控利用的关键技术包括可仲裁容毁确权、权属登记与验证、数据使用控制、数据增值利用等。

3.3.1 可仲裁容毁确权

数据要素的可仲裁容毁确权需要解决数据权属属性提取、容毁确权仲裁等关键问题。

1) 数据权属属性提取

数据权属属性可根据类型分为特殊属性和一般属性，特殊属性是能够唯一标识数据所有者身份的单一属性，而单条一般属性则不具备唯一标识的能力，需要多条一般属性组合才能唯一标识数据所有者的身份。针对待确权的数据，需要提取能够确定数据权属的特殊属性或者一般属性的组合。

2) 容毁确权仲裁

容毁确权仲裁可以采用相似性度量方法，通过对比确权系统中已有的知识图谱与新申请确权数据所计算得到的知识图谱，判断是否一致。通过对确权系统中的知识图谱进行关键词检索，本质上可将确权仲裁转化为2个图谱集合的相似性度量。

3.3.2 权属登记与验证

权属登记与验证协议主要涉及确权系统、数据权属者、数据接收者3个主体的交互。确权系统负责颁发权属证书，用于验证数据权属者身份；数据权属者拥有某数据，并希望通过确权系统保障该数据的所有权；数据接收者接收来自数据权属者的数据，并希望验证数据权属者的身份。

权属登记与验证协议的完整流程可分为2个阶段，一是权属证书发行，数据权属者与确权系统交互获得权属证书；二是权属验证，数据接收者与确权系统交互验证权属。

权属证书发行阶段，数据权属者向确权系统提供希望声明所有权的数据信息；确权系统对该数据生成随机的协同抗毁数据以及相关的使用规则，数据权属者提取能唯一确定数据权属的权属属性，按照使用规则协同抗毁数据一起生成仲裁码，将仲裁码返回确权系统；确权系统检查该仲裁码是否被登记过，如果没有，则在数据库中记录数据权属者的身份与仲裁码的对应关系，并生成权属证书，将权属证书返回给数据权属者。

权属验证阶段，数据权属者在向数据接收者发送数据时，需附上由确权系统颁发的权属证书。数据接收者想要验证收到的数据是否确实属于数据权属者，需要向确权系统提交数据权属者的身份、数据权属属性和权属证书序列号；确权系统将该权属证书的真实性和有效性结果返回给数据接收者；数据接收者使用确权系统的公钥验证数据权属者的权属证书，并利用权属证书验证数据权属。

3.3.3 数据使用控制

数据要素泛在流通中具有交易多轮动态、交易数据出域多次流转、使用控制自适应变更等特点和应用需求，迫切需要解决数据要素跨域多轮受控使用的问题。

首先建立面向数据跨域流转的延伸控制模型。延伸控制模型定义了跨管理域交换后对数据使用的授权模式，是确保数据使用者按照数据权属者意愿使用的根本。因此，需在深入研究数据跨域流通场景下细粒度访问控制模型的基础上，设计数据权属者对数据预期共享意愿的归一化描述方法，包括数据使用者使用数据的条件（如可流转轮数、可用次数）、数据使用者的权限（如允许对何种数据要素执行何种操作、不允许对何种数据要素执行何种操作）和数据使用者的义务（如数据必须在某位置下才能继续流转、数据必须加密存储、数据存储14天后必须执行最高程度的删除）；基于多源数据使用日志，构建涵盖“谁、何时、何处、对数据要素作何处理”等要素的数据域内/域间传播链，设计基于共享意愿和数据要素传播链等要素的混合授权机制，确保数据在全生命周期的受控使用。

3.3.4 数据增值利用

数据要素流通的数据使用环节中，可以通过信用计算、联邦学习、大模型等关键技术实现数据要素的增值。

1) 信用计算

在数据流通领域，信用计算是指数据使用方利用数据提供方和数据使用方共同拥有的用户数据，根据信用评估模型对用户信用状况进行计算。

信用计算需要对数据提供方的原始数据进行计算以保证计算结果的准确性。因此，信用计算需要数据提供方的数据在数据使用方脱敏，且等效于数据提供方脱敏。同时，数据从数据提供方到数据使用方的安全传输过程以及在数据使用方的计算过程中，数据使用方无法获得数据明文。

2) 联邦学习

联邦学习是多方利用自身数据完成部分模型的训练、中心节点完成模型汇集的一种分布式的机器学习架构。在训练过程中，合作方之间交换训练中间结果和模型参数，而不交换数据本身，自然而然不存在数据出域而导致的原始数据泄露，但中间结果交换本身没有防泄露的机制，仍然存在部分数据泄露的问题。

根据参与训练数据维度的特征，联邦学习可分为纵向、横向以及迁移3种不同的模型迭代算法。联邦学习算力不需要集中，可充分利用分布式算力，减少最终模型需求方算力设备的资金投入。联邦学习的数据不需要出域，满足原始数据不出数据提供方本地的愿望，可促进跨组织协作。联邦学习属于人工智能研究范畴，可应用于数据增值服务、衍生数据产生，但不能应用于数据流通与利用场景。

3) 大模型

大型语言模型（LLM, large language model）简称大模型，是参数量非常大的深度学习模型，其训练过程通常包括预训练和微调2个阶段。其中，预训练阶段在大规模数据集上进行，使模型学习到丰富的语言和语义知识；微调阶段在特定任务的数据集上进行，使模型适应特定的任务需求。

大模型通常具有更高的计算能力和更强的泛化能力，可有效支撑数据要素流通。首先，大模型具备强大的特征提取和表示学习能力，可以从大量数据中提取有用的特征和知识，有助于提高数据的可读性和可理解性，为数据流通提供更好的支持；其次，大模型具备强大的数据清洗和预处理能力，能够从大量数据中提取出有用的信息和知识，从而提高数据的质量，有助于减少数据流通中的误差和歧义；最后，大模型在训练过程中可以对数据进行深入的分析和处理，从而提取出更加有用的信息和知识，提高数据流通的准确性和效率。

3.4 数据按需保护的关键技术

数据要素流通过程中的数据按需保护面临数据传输的机密性、完整性、不可否认性等安全问题，可用已有的身份认证、传输加密、完整性校验、数字签名等技术解决。此外，针对数据利用过程中的数据安全保护，也提出了具有代表性的安全多方计算、密文计算等密码方案。

安全多方计算旨在解决多个组织或个体在需要共同进行数据分析或决策时保护各自敏感信息的需求。通过特定的协议设计，允许各方在不直接交换或暴露未分享的原始数据的情况下，共同参与实现数据的计算和分析，达到未分享的原始数据不出域、只交换计算结果的数据使用效果。典型的安全多方计算协议可由包括秘密共享、不经意传输、混

淆电路等在内的技术实现。安全多方计算的计算复杂性和通信复杂性很高，严重制约了其在数据保护中的广泛应用。

密文计算能够在不解密的情况下对加密数据执行计算，以达到对明文的相同计算效果，并且计算结果也是加密的。密文计算技术是数据出域的安全计算技术，但是通用的全同态加密计算开销极大，部分同态加密技术适用场景有限且计算效率不高。因此密文计算难以适应大规模数据要素流通的安全计算。

3.5 隐私按需保护的关键技术

数据要素流通中，个人隐私信息伴随传播，所有权与使用权分离。隐私计算是数据出域流通隐私按需保护的基本支撑技术。

3.5.1 数据脱敏后的出域流通

利用隐私信息描述方法，结合给定发送者的隐私保护需求以及场景描述信息，生成其控制策略。利用隐私计算中的控制策略可信传递机制，定义数据流通过程中的传播链，发送者随后将策略与数据绑定传递给接收者，接收者验证数据绑定的策略，判断自己是否满足发送者的约束条件集合、传播控制操作集合等，若满足，则可对文件进行允许范围内的合法操作，上述方法可应用于数据要素多次流通场景下的脱敏迭代控制。在此基础上，基于场景适应的隐私信息动态度量、按需脱敏、隐私保护效果评估、侵权行为取证与融合分析、多副本完备删除等技术，实现数据要素流通全生命周期的隐私信息保护。数据出域前脱敏降低了可用性，不适用于信用计算等高计算精度需求的场景。

3.5.2 数据出域的安全计算

大模型、信用计算等场景需要原始数据出域，在数据使用方实现数据脱敏的安全计算，从而获得高精度的计算结果，并对不可信的数据使用方保护数据提供方的数据安全和个人信息安全。需要解决的关键问题包括：①数据从数据提供方到数据使用方的传输安全；②数据使用方不能获得数据提供方的明文数据，但仍可利用数据提供方的数据实现高效的大模型训练或者信用计算，因此数据使用方需在一个可信计算环境中实现数据解密，对明文实现模型训练或者信用计算并得到计算结果。

3.6 流通利用监管的关键技术

数据要素流通利用监管的关键技术包括低开销、抗泄露、防篡改存证，数据血缘等。

3.6.1 低开销、抗泄露、防篡改存证

针对数据要素流通过程多方利益冲突、存证海量与全流程存证、第三方可信仲裁等特点和需求，需要建立低开销、抗泄露、防篡改的存证机制，从而支撑数据要素流通的全流程监测和争议发生时的可信仲裁。

存证系统主要涉及信息系统、本地存证系统、中心存证系统3个主体。信息系统包括各类数据要素的处理系统，此类系统中对数据要素的操作日志、操作记录等证据信息会按照监测要求上传至存证系统；本地存证系统指部署在信息系统所在单位的存证系统，对信息系统上报的完整证据信息进行管理和存储；中心存证系统是数据要素流通的监测系统，此系统部署在中心监测机构，主要是对信息系统上报的摘要证据信息进行管理和存储。

首先，需要设计安全协议实现信息系统与存证系统的交互，确保存证信息的完整性和上传方身份的不可抵赖性。其次，本地存证系统中保证有全量的操作日志等信息，其中包含大量的机构敏感信息，用于信息系统运营方自身的监测。中心存证系统一般部署于第三方机构，如果全量保存各信息系统的信息，一是开销大，二是会产生信息系统运营方信息泄露的风险。因此，需要设计低开销、抗泄露、防篡改的存证机制实现信息系统的第三方存证。

3.6.2 数据血缘

针对数据要素流通中数据来源广泛、数据之间关系复杂、多种类型数据快速增长的特征，需要研究数据关联影响分析、数据溯源、数据价值评估技术，构建数据血缘，解决多层复杂逻辑处理后的数据难以理解、难以应用、难以评估价值、难以定位的问题。

数据血缘是指数据全生命周期中数据产生、处理、加工、融合、流转、销毁的关联关系，本质是关联关系的准确性与时序性，可反映数据流通与数据演化的全过程。数据的这种关联关系与人类的血缘关系比较相似，所以被称为数据血缘关系。由于数据血缘记录了数据产生、交易、交换、处理、销毁等各个环节的使用情况，因此通过构建数据血缘图，可支撑逆向追溯数据演化过程；发生争议时利用数据血缘图，还原数据在不同环节/不同时刻利益攸关方的状态，支撑数据争议仲裁。

4　数据要素流通的技术展望

4.1　可仲裁容毁确权

数据要素流通过程中，解决数据资产权益的登记、发布、搜索以及数据要素部分特征被有意或无意损毁情况下的确权仲裁，是保证数据提供方权益的前提。确定和提取不同模态数据的权属唯一性的权属属性集合，同时还能在部分权属属性损毁的情况下实现数据权属的仲裁判定是未来需要研究的关键技术。

4.2　数据交易与安全交付

数据可信交易和安全交付重点关注交易过程中信息交互的机密性、完整性、不可抵赖性和公平性，数据资产化要求数据要素需要出域交付。未来需要重点研究多方动态可信磋商机制、契约可信签署、契约约束的交割协同执行与一致性核验、权益转移、交易回滚状态下的数据跨域可验证删除与权属同步返还等关键技术。

4.3　使用控制与监测

在数据要素流通的过程中，如何保证多轮交易过程中的数据按照数据提供方与使用方达成的契约要求实现迭代延伸的使用控制是数据流通的真正挑战，需要建立数据要素流通全流程的数据使用控制规范、可信执行机制，同时，建立与之相应的第三方监测机制，达到低开销、抗泄露、防篡改的目标，为争议仲裁提供支撑。

4.4　争议仲裁

数据要素流通过程中不可避免地存在争议，需要研究通过存证数据的跨系统关联分析和一致性分析技术，及时精准地发现违规线索；在此基础上，对多方证据进行交叉验证，重点研究数据交易的契约与证据的自动化分析技术，生成违规证据链，实现违规行为的及时发现和溯源取证。

5　结束语

数据要素流通已经成为国家数字经济发展壮大的必然要求，然而数据要素流通过程中面临的数据要素可仲裁容毁确权、数据要素资产化要求的数据出域可信交易与安全交付、数据要素流通全流程使用控制、基于隐私计算保护出域数据要素中的个人信息、数据要素流通全流程的存证监测和违规溯源等都需要建立全面的安全保障支撑技术体系，需要学术界、产业界和主管部门大力协同，从理论、技术、产品、标准和管理法规方面一体推动，才能真正促进数据要素的有序合规流动，释放数据的应用价值和效能，促进我国数字经济健康发展和行稳致远。

参考文献：

[1] Federal data strategy, data ethics framework [EB/OL]. (2020–09–04)[2024–03–06].
[2] General data protection regulation[EB/OL]. (2018–05–25)[2024–03–06].
[3] European data governance act [EB/OL]. (2022–04–06)[2024–03–06].
[4] Data act[EB/OL]. (2023–11–27)[2024–03–06].
[5] 李凤华, 李晖, 贾焰, 等. 隐私计算研究范畴及发展趋势[J]. 通信学报, 2016, 37(4): 1-11.
LI F H, LI H, JIA Y, et al. Privacy computing: concept, connotation and its research trend[J]. Journal on Communications, 2016, 37(4): 1-11.
[6] LI F H, LI H, NIU B, et al. Privacy computing: concept, computing framework, and future development trends[J]. Engineering, 2019, 5(6): 1179-1192.

[作者简介]

李凤华（1966– ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、隐私计算、数据安全。

李晖（1968– ），男，河南灵宝人，博士，西安电子科技大学教授、博士生导师，主要研究方向为密码与信息安全、隐私计算、数据安全。

牛犇（1984– ），男，陕西西安人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为隐私计算、数据安全。

邱卫东（1973– ），男，江西修水人，上海交通大学教授、博士生导师，主要研究方向为大数据安全、隐私保护。