

## 数据价值链研究进展\*

李正辉 许燕婷 陆思婷

**摘要:**数据价值的创造过程可划分为基本价值和增值价值两大类。数据价值链在数据的采集、传输、存储、分析及应用等环节上实现数据的基本价值创造,在数据流转过程中实现价值增值。本文从数据流程、数据处理技术和数据增值三种视角系统阐述数据价值链的概念,并追踪其所经历的信息经济—数字技术嵌入—双链融合的演进过程。文章聚焦数据生产与应用流程链条以及数据与产业融合的实体链条,重点分析数据价值创造机制;同时,系统总结数据价值评估方法,从微观和宏观视角考察数据对利益主体的福利效应;进一步,全面审视数据价值链的生态模式和治理体系,探讨数据价值链生态的研究进展;最后,展望了数据价值链研究的拓展方向。

**关键词:**数据价值链 价值创造 价值评估 数据生态

数据兼具基础性战略资源和关键性生产要素属性,其创造的价值可分为基础价值和增值价值。数据价值链强调链条上数据采集、处理、存储、分析和应用环节中基础价值的创造以及流转过程中价值增值的实现。同时,数据价值链与其他产业链深度融合以产生更大价值或实现数据赋能。随着数据价值链领域的科学系统研究逐步形成,本文以数据价值链为研究对象,系统梳理其概念及演进阶段、价值创造机制、价值评估与福利效应、价值链生态四个方面的内容,并对未来的研究方向进行展望。

### 一、数据价值链的概念及演进阶段

尽管数据价值链概念是从传统价值链概念扩展而来,但其在Kasim et al.(2012)首次提出数据价值链的概念之后已经成为一个新兴领域。该领域因数据本身的交叉性和复杂性而呈现出独有的特性和演进路径。从不同学科视角来看,数据价值链的定义呈现出多样性,主要集中在数据流程、数据处理技术和数据增值三个典型视角。尽管这三种视角关注的焦点各有侧重,但共同揭示了数据价值链的三大重要特性:有序性、动态性和价值创造。数据价值链的演进可以追溯到信息技术应用给各类主体带来的福利效应(信息经济阶段),经历了数字技术的逐步嵌入(数字技术嵌入阶段),成型于双链融合(数据价值双链融合阶段)。

#### (一)数据价值链的概念界定

统计学家更多从数据搜集、整理、分析、应用的全流程定义数据价值链。基于数据流程的数据价值链是指从数据搜集到决策的全链条和流程,其基本原理是通过数据链条式的连贯应用,促进各利益相关者之间的协作,从而最大限度地挖掘和实现数据的潜在价值。数据价值链的框架涉及数据搜集、整理、分析、可视化和决策等系列活动。Hu et al.(2014)将大数据系统分解为数据生成、数据采集、数据存储和数据分析四个顺序模块,并将数据价值链定义为一个系统架构。Curry(2016)将数据

\*李正辉,广州大学经济与统计学院、广州大学金融研究院,邮政编码:510006,电子邮箱:lizh@gzhu.edu.cn;许燕婷,广州大学经济与统计学院,邮政编码:510006,电子邮箱:2112164136@e.gzhu.edu.cn;陆思婷,广州大学金融研究院,邮政编码:510405,电子邮箱:lusiting@gzhu.edu.cn。基金项目:国家社会科学基金重大项目“数据价值链效能识别与统计测度研究”(22&ZD163)。感谢匿名审稿专家的宝贵意见,文责自负。

价值链定义为能够让数据产生价值的获取、数据分析、数据管理、数据存储和数据使用等链式步骤。Visconti et al.(2017)认为数据价值链是大数据被捕获、存储、处理和共享的连续步骤。Li et al.(2019)认为数据价值链是由数据收集、存储、分析和驱动四个环节共同构成并不断产生价值的业务模型。Faroukhi et al.(2020b)将数据价值链视为链式模型,强调数据在整个链式流程内的各个环节中实现价值。链式模型由数据生成、预处理、存储、分析、可视化、共享与传输六个环节依次组成。在该模型中,数据被视为原材料和主要资源,可由最初的状态逐步转化为可以间接提取价值的形式。

数据科学家将数据处理技术区分为不同活动,并从数据处理技术视角定义数据价值链。他们认为数据由低到高经历四种表现形式,进而形成四种数据类型:未经任何处理的“原始数据”,通过关系连接被赋予意义的“信息”,通过预处理、分析、计算、存储过程演化而成的“知识”,通过处理和可视化过程实现价值的“智慧”(Singh, 2020)。数据价值链是在链条式数据类型的基础上将各种活动与相应的技术对应并寻找能够产生价值的活动,使数据在被利用过程中创造价值并实现货币化。不同类型的数据处理技术对应不同环节的现实目标,例如,基于立方体的能耗模型多用于数据预处理环节,基于长短记忆的模型则多用于生态状态预测(Ma et al., 2020)。一个环节可能涵盖多个技术,一个技术也可能贯穿多个环节,特别是批处理、在线分析、流处理等数据处理技术为数据提供了更大的增值空间。Kriksciuniene et al.(2015)认为,数据价值链是促进所有部门合作、使数据被连贯使用的技术链条,该链条能够通过数据信息技术提供丰富的数据分析活动,为企业或产业链条提供价值。Semlali et al.(2020)提出,数据价值链是通过各项数据持续系统地加工,实现数据价值创造和增值的链式流程。该流程由数据输入端集群、数据流和价值输出端三个核心平台与关键技术来支撑。数据输入端主要负责各种多源数据的抽取、按照既定规则进行有效数据的装载与分发;数据流促进数据资产在平台中持续循环流动;价值输出端将数据价值充分传递到各个利益相关者。

经济学家将数据视为生产要素,他们更多地从数据增值角度定义数据价值链。数据通过与经济活动各个环节嵌入和融合而产生价值。数据要素在经济活动链条中可以是中间产品或最终产品。数据作为中间产品,在驱动企业或产业的内部决策、管理与协调、技术与管理创新等环节体现数据增值作用,最终体现为竞争力提升。数据作为最终产品,通过流转与信息技术融合,为客户提供高质量的服务等方式提升数据价值,典型案例如定制化的解决方案、智能化的产品体验以及基于客户需求的创新性服务等。Petrova-Antonova et al.(2018)提出,数据价值链是决策辅助系统,该系统包含数据获取、结构构建、分析以及模型的构建与应用等系列活动,这些活动与数据产品、服务直接或间接相关。Attard et al.(2016)认为,数据价值链是数据生产和流通过程中各参与者(如数据生产者和数据消费者)进行数据发现、管理、解释、分发和利用五类相互关联又具有一定独立性的系列操作,强调数据作为产品在经济活动中的增值和价值创造作用。Wirén et al.(2019)和Faroukhi et al.(2020a)认为,数据价值链核心是数据价值增值过程,即数据产品在价值链的各个阶段积累价值的可重复过程。其中,Wirén et al.(2019)将数据价值链分为数据化、数字化、连接、储存、分类、模块化、交叉分析和个性化八个阶段。Faroukhi et al.(2020a)将数据价值链分为数据生成、获取与预处理、储存、分析、可视化和共享与传输六个阶段。

## (二)数据价值链的演进阶段

随着信息经济和数字技术的推进,数据价值链经历了三个显著的演进阶段:信息经济、数字技术嵌入以及数据价值链双链融合阶段。信息经济阶段驱动力源于数据等信息给各类市场主体带来的福利效应;数字技术嵌入阶段驱动力源于技术在数字经济领域的广泛应用与推进;双链融合阶段驱动力一方面源于数据从生产到消费链条流程产业化,另一方面源于数据流程链条与产业链条的充分融合。

1. 信息经济阶段(1990—2000年)。信息经济阶段的显著特征在于信息化给各类主体带来的福利效应。信息化主要通过数据体现,同时伴随着移动通信技术在商业、社交媒体、搜索引擎等典型应用场景出现。以数据为载体的信息化为消费者、生产者、政府和社会等主体创造了更多价值。消费

者可以利用、分享和创造各种信息来满足其个性化与多样化的需求,进而提升其生活质量和幸福感,数据价值通过消费者使用过程得到增值(Zhang et al.,2017)。在线市场中评论信息是一种关键信息源,其包括质量信息(反映商品质量水平)、匹配信息(反映商品是否符合消费者偏好)和准确信息(真实可信的评论信息)三大特征,会对消费者福利产生不同影响。不同类型消费者对不同的评论信息特征存在不同的敏感度和偏好,这导致评论信息对消费者产生的福利具有异质性。生产者则通过开发、获取、协调和整合信息资源扩展现有知识库,以获得低成本、高增值的竞争优势(Gaimon & Ramachandran,2020)。如Koutsoukis & Mitra(2003)认为,在信息经济链条中,企业信息处理分为信息输入、加工、输出和反馈四个关键阶段。在这个框架下,生产者能够系统性地整合内外部信息,通过有效的信息加工和反馈机制,实现对市场和生产环境的敏锐洞察,打造具备可持续竞争优势的业务模式。政府通过信息管理系统等技术更高效地收集、分析和利用各类信息,提升政府治理的便捷性和公平度。Weerakkody et al.(2018)研究表明,政府能通过信息系统的链式流程高效管理废弃车辆。该链式流程首先由信息系统自动接收市民上报的废弃车辆信息,其次是信息核对系统对市民提供的信息进行验证,最后是相关部门对已核对通过的废弃车辆进行及时处理。社会可以通过信息技术推动各方面的发展和进步,提升社会创造力(Gaimon & Ramachandran,2020)。例如,文化市场可以将文物、古籍、艺术、语言等事物转化为相关信息并上传云端,随后推送到不同终端和场景,以提升社会信息化程度,从而提升社会创造力。

2. 数字技术嵌入阶段(2000—2010年)。数字技术嵌入阶段的显著特征是将数字技术智能化地应用于数据处理链条和产业链条。一方面,数字技术在数据链条式流程中得到广泛应用。Hadoop分布式框架通过结合批处理和流处理模式,能同时处理有限的历史数据流和无限的实时数据流(Bhathal & Dhiman,2018)。其核心组件包括分布式文件系统和分布式计算框架。分布式文件系统将大规模数据分割存储在集群的多个节点,允许系统并行处理这些数据块以实现高效的读写操作。分布式计算框架将任务分为Map阶段和Reduce阶段,通过在分布式计算节点上并行执行这两个阶段,处理分布在分布式文件系统上的数据。机器学习算法无须依赖显式的编程规则,计算机能够通过训练数据集从数据中学习并做出分析和预测。Yeh et al.(2013)通过对传统的信用评价方式和基于随机森林、粗糙集理论等机器学习算法的信用评价方式的比较发现,机器学习算法在预测等方面更有效和精准。Alizadeh et al.(2023)提出了银行业客户流失模型,它融合主观数据和客观数据,采用随机森林算法分析银行业客户的流失原因和预测流失概率。另一方面,数字技术与产业链的融合得到了一定程度的应用。Kim et al.(2020)阐述了物联网技术在农业中的应用架构,该架构可分为感知层、网络层和应用层三个层次。感知层是用于识别和发送农业生产过程中各种对象及其对应参数数据的设备层,主要包括传感器节点和传输器;网络层是指用于数据传输、接收与处理的通信层,主要包括本地网关和云端服务器;应用层是指用于农业应用的功能层,主要包括管理系统、监测系统、控制系统和无人机械系统。Lv et al.(2022)结合工业物联网和数字孪生技术提出了危险气体泄漏跟踪的具体应用架构,该架构分为感知层、边缘层和控制层三个层次。感知层负责采集和监测危险气体的数据;边缘层利用数字孪生技术对气体的迁移和扩散进行模拟,从而实现气体泄漏的检测和定位;控制层则负责分析和解决具体问题。

3. 双链融合阶段(2010年至今)。双链融合阶段标志着一个全新数据价值链阶段的到来,该阶段的显著特征是,信息经济逐步演进为数据流程与价值创造链条,数字技术嵌入逐步演进为产业链条融入数据支持,最终形成双链条融合。链条之一是数据价值的创造链条,是指数据生产到应用的全流程链式过程。Kasim et al.(2012)提出了一个基于云平台实现数据处理、数据安全和数据分析的数据服务框架,包括数据收集、管理、整合、分析、模拟和可视化六个环节。其中,前三个环节获取和管理不同类型和结构的数据;数据分析环节提取数据中有价值的信息;数据模拟环节利用获得的信息对假设场景进行探索并对未来事件进行预测或优化产品服务;数据可视化环节直观地展示复杂的数据分析结果。Faroukhi et al.(2020b)认为,数据价值链包含数据生成、收集、分析和交换四个核心环

节。数据交换环节的核心是数据服务平台,是基于正常的商业交易规则构建的数据输出主体。Visconti et al.(2017)将数据价值链延长至数据消费和数据货币化环节。数据消费是指数据的使用者通过数据服务平台获取并利用数据的过程。数据货币化是指数据提供者通过数据服务平台将数据转换为经济效益的过程,如数据出售和数据订阅。链条之二是数据价值的实体链条,指数据沿着生产链条流动与创造价值的过程。Eling & Lehmann(2017)基于Porter价值链模型构建保险数据价值链,链条主要涉及市场营销、产品发展、销售、承销、理赔管理等环节。数据在链条的各个环节中发挥不同的作用,如在市场营销和产品发展阶段进行市场分析和趋势预测,在销售和承销阶段进行保险定价和销售渠道优化,在理赔管理阶段进行事故分析和流程优化。Müller et al.(2019)则提出了一个数字化的木材供应链框架,覆盖从森林到终端用户的一系列活动,包括采伐规划、采收组织和控制、采伐作业、木材运输和物流、木材销售等环节。这个框架利用地理信息系统模拟、实时数据传输、运输追踪技术和提供个性化服务,优化整个木材供应链。数据价值的创造链条为数据价值的实体链条提供数据支持和服务,而数据价值的实体链条对数据价值的创造链条提出需求和反馈,呈现出两链条相互交织、相互促进的复杂特征。

## 二、数据价值链的价值创造研究

数据价值链具有典型的双链特征:其一是以数据生产和使用流程为链条,涵盖数据搜集、存储、分析、流转与应用;其二是数据与产业融合的实体链条,实体链条是以数据沿着生产链条数据流动与价值创造相伴而动的过程。数据实体链条更多的是作为数据价值创造的渠道。研究数据价值链的价值创造,更多根据链条运行机制分析其价值创造机制。

### (一)数据运行环节中价值创造研究

数据流程链条从数据生成与获取开始,以数据处理与分析为核心,终于数据流转与使用。数据生成与获取环节的价值主要体现在高质量数据的获取;数据处理与分析的价值主要体现在数据潜在规律的洞察;数据流转与使用环节的价值主要体现在数据转化为实际业务行动。

1. 数据生成与获取环节的价值创造机制。高质量数据的获取是数据生成与获取环节的价值来源。高质量数据的特性主要包括可靠性、稀缺性和时效性。可靠性指数据反映客观事实的真实和准确程度,是该环节价值产生的重要基础(Li et al., 2019)。获取可靠数据需要可靠的数据源,以避免数据的篡改、伪造与损坏等。政府部门、权威机构、专业平台等通常可以提供高质量的数据,而一些不明身份的网站、社交媒体、个人账号等可能提供低质量或虚假的数据。可靠性数据的获取还需要验证数据内容,以避免数据的错误、偏差、噪声等,常用方法包括数据校验、数据清洗、数据标准化等。稀缺性是指数据的供给量相对于需求量的不足,或数据的获取难度和成本较高,或数据的独特性和不可替代性较强,是该环节的重要价值来源(Lynn, 1991)。稀缺数据具有较高的市场价格和竞争力,可为获得这些数据的企业建立起独特的竞争优势。例如,越详细的个人信息越能够准确地对用户进行画像从而提高广告的精准度。时效性是指数据能够及时更新和反映变化的程度,是该环节可持续创造价值的重要保障(Müller et al., 2019)。例如,传感器技术与物联网可以实时地监测和收集各种类型的数据,以反映环境、设备或其他实体的状态变化。

2. 数据处理与分析环节的价值创造机制。数据处理与分析环节的价值创造源于把握数据中蕴含的规律,进而获得相关信息;该环节与嵌入相关的流程包括数据融合、场景关联、完善流程。Villar et al.(2018)认为多源数据的融合,如来源于企业自身产生或收集的数据、企业从外部渠道或平台获取的数据以及源于不同流程或阶段上的数据的融合,能够提高信息准确性、降低信息不确定性、增强系统可靠性、提升建模过程中数据的匹配性,从而实现从数据到信息的价值创造。异构数据使数据类型和数据格式存在较大可变性。对异构数据进行处理能够将数据映射到一个共同的数据模型中,以实现信息融合,提升信息提取效率和建模效果,进而创造数据价值(Visconti et al., 2017)。多源异构数据通过ETL流程、数据仓库、API与数据集成工具等技术进行融合,进而在流转中创造数据价

值(Semlali et al., 2020)。数据关联到医疗和零售等业务场景后,数据价值能够得到显著提升。例如,Bellini et al.(2023)将患者的个体差异和生活方式等影响健康状况的动态数据、影像数据和临床数据的实时数据融合,可以更及时和准确地分析患者的健康状况。Zhao et al.(2023)认为,消费者在电商平台的购物行为数据、在社交媒体的互动数据、用户属性数据等数据的融合能对消费者进行细致的画像和特征分析。Falcao et al.(2021)结合实体店模型和实时库存监控优化自助收银方式,实现高达96.4%的数据准确性,进而提高了销售状况分析的准确性。

3. 数据流转与使用环节的价值创造机制。数据流转与使用环节的价值创造源于数据的交换,将数据价值全部转换为货币价值。该环节包括数据包装和数据流通过程。数据包装是指对数据处理与分析阶段的信息加工,展示数据的价值和使用价值。数据供给方主要通过自用、共享和销售等方式进行数据流转。Li et al.(2019)认为,对数据分析阶段输出结果的包装能够提高数据的可用性、可理解性和可交互性,使数据更适应业务场景或用户需求,进而实现数据价值的转换。数据流转是指供给方将数据分析结果传递给需求方,需求方利用这些信息进行具体的业务行动,以实现数据价值向经济效益的转换。Visconti et al.(2017)以10家不同行业的企业为案例,提出了数据货币化的9个策略:将数据用于内部操作以保持数据的专有性;进入新的业务领域,实现产品或服务的差异化;将专有数据以出租、租赁、临时授予等形式授权给客户使用并约定按客户实际销售或使用情况收取费用;与商业合作伙伴共享数据以最大化利益;向资产所有者、其他相关方等销售优质数据产品;向所有的数据需求方提供数据的付费访问;与商业合作伙伴进行数据的买卖或交换;将数据作为广告的载体,利用数据来吸引用户的注意力或提高用户的参与度,从而获得广告商的付费;通过数据融合生产新的数据产品和服务,如数字新闻和数字游戏。Müller et al.(2019)提出,利用森林生态数据(树种分布、树龄、森林结构、土壤质量等数据)可以制定更高效更具可持续性的伐木计划,如利用采收木材现场数据自动化监测和控制木材砍伐和加工过程,利用木材运输和交通状况、道路条件等实时数据选择最佳运输路径。

## (二) 实体链条价值创造研究

实体链条更强调数据与产业融合。数据实体链条涵盖多个参与主体,从市场角度看,参与主体分为生产者和购买者,因此数据实体链条价值可由生产者、购买者或多源驱动创造。生产者驱动数据实体链条的价值主要源自生产过程中的研发和制造阶段;购买者驱动数据实体链条的价值主要来自营销和服务流程;多源驱动数据实体链条的价值主要依赖市场参与者的协作和反馈机制。

1. 生产者驱动数据实体链条的价值创造机制。生产者驱动数据实体链条的价值源于数据要素化,即将数据作为生产要素植入生产的研发和制造过程中。生产者将数据融入生产活动,运用价值倍增、资源优化、投入替代等方式实现数据的价值创造。Ghasemaghaei & Calic(2020)在研究公司创新绩效时发现,数据会作用于公司创新绩效,而具有不同特征的数据会对公司创新绩效产生异质影响。多类型的数据能够通过丰富企业的客户数据促进企业新知识和新思想的产生,进而推动企业开发更符合消费者需求的新产品;高频率更新的数据使企业能够及时洞察市场变化趋势,帮助企业快速产生新想法并迅速将其转化为创新产品。Kim et al.(2020)认为,企业可以通过传感器、仪表和监测设备实时监测能源使用情况,将实时采集的数据进行深度分析并识别能源使用的模式和趋势。基于这一深入分析,企业能够建立智能控制系统,通过自动化和智能调控对设备和系统进行精细化的能源管理。Brandas et al.(2016)指出,数据能在合理分配劳动资源方面发挥关键作用:一方面,企业通过监控加班情况、生产效率以及员工流失率等关键数据,能够更准确地进行人力需求规划,确保在不同业务周期和季节性需求变化时有足够的员工资源,避免因人力短缺或过剩而导致生产效率下降;另一方面,企业通过分析员工工作技能和绩效等数据,提高招聘的准确性并更好地匹配人才与职位需求。Na et al.(2022)提出,企业可以通过自动化控制系统减少制造过程中对人力的需求和干预,提高生产的稳定性和精确性。

2. 购买者驱动数据实体链条的价值创造机制。购买者驱动数据实体链条的价值源于数据在服

务过程中提供信息。具体而言,在产品或服务的营销和服务流程中,购买者将其需求反馈给生产者,促使生产者通过个性化营销、客户关系管理等方式优化现有服务,进而实现数据的价值创造。Zhao et al.(2023)认为,通过分析消费者在购物平台上的浏览、搜索和购买行为等个性化数据,可以辨别出每个购买者的兴趣类型是属于受个体特征、长期购物历史等因素影响而形成的稳定购物兴趣,还是因季节、促销活动或社会事件等瞬间引发的暂时购物兴趣,进而识别购买者的行为模式并生成对应的个性化推荐内容。为确保推送的产品或服务更符合消费者当前的兴趣和需求,这种数据驱动的个性化推荐系统具有离线和实时两种推荐方式(Roy & Dutta,2022)。离线推荐系统通过深度分析历史数据,在用户未发出具体需求时预测其可能感兴趣的内容。在离线分析中,数据分析促进推荐决策的内容,进一步提升了数据的价值。实时推荐系统通过分析购买者在平台的浏览、搜索或交互行为,即时调整推荐内容。此外,Del Vecchio et al.(2021)强调,购买者与企业的多维互动数据在客户关系管理和服务优化方面发挥了关键作用:一方面,企业通过实时沟通渠道与购买者直接互动,在购买者提出问题或遇到困难时迅速响应,提升客户满意度;另一方面,企业能够通过分析购买者与企业互动的历史数据,更好地理解购买者的需求、行为模式和服务偏好,进而改进服务和优化流程。

3. 多源驱动数据实体链条的价值创造机制。多源驱动数据实体链条的价值源于数据协同,即数据实体链条的多个主体协同作用,使数据在实体链条中充分流通和应用,进而创造更大的数据价值。Soosay & Hyland(2015)认为,供应链中各个主体之间实时共享和分析数据能够提高供应链的透明度、灵活性和响应能力,从而提高客户满意度和利润率。在制造公司与供应商、物流公司和零售商的协作中,零售商通过销售数据和客户反馈分析市场需求,制造公司结合市场需求利用物联网和传感器等设备调整生产计划,物流公司通过调整运输路线和提高运输效率迅速地将产品从供应商送达零售商。这种高效的运作不仅有助于降低整体运营成本,还提高了整个供应链的灵活性,彰显了数据对不同主体所创造的价值(Nimmy et al.,2019)。随着成员之间协调水平的不断提高,供应链中的信息流也随之增加,进而提高供应链的绩效和竞争力,通过主体绩效提升而创造数据价值(Eltantawy et al.,2015)。在医疗领域,所有可用知识源的结合能够提高医疗质量和安全性,并为未来医疗领域的发展带来新的可能性(Shaik et al.,2023)。例如,通过整合患者的电子病历、实验室检查和医学影像等多维度数据,医疗机构能够更全面地了解患者状况,从而提高医疗决策的准确性。医疗系统通过监测患者病理数据,使得医生能够更及时地响应患者的需求和提供必要的医疗支持。制药公司在新药研发过程中结合患者的病历数据、用药反馈等信息,以更好地了解药物的效果和副作用,加速新药的研发和上市,为患者提供更优质的治疗选择。数据通过在医疗领域各利益相关者之间的流通、分析和应用,实现了更大的数据价值创造(Verboven et al.,2022;Bellini et al.,2023)。

### 三、数据链中的价值评估与福利效应研究

数据在生成、流转、融合等环节和链条中均会体现价值,同时各环节和链条又会涉及参与数据价值链的利益相关者。数据创造机制在链中的不同环节具有差异,数据价值评估方法也存在差异。在数据价值全链条中涉及许多利益相关者,各利益相关者关注的焦点是数据为参与数据价值链的各主体所带来的福利效应。参与数据价值链的主体主要分为宏观和微观两大类,故研究集中于数据为这两类主体所带来的福利效应。

#### (一)数据链条中的数据价值评估方法

数据全链条中的各主体专业背景差异使得数据价值评估方法出发点具有差异,形成了资产观和资源观两类典型的数据价值评估视角。资产观认为数据是一种资产,即具有基于产品使用价值的客观存在形态,而这种资产是由机构或法人所控制的。资源观则将数据视为一种能够促进各类主体增值的要素,其对个人或组织的经济效益更多表现在融合赋能上,有时甚至为间接贡献。尽管两种视角采用的数据价值评估方法有所差异,但两类方法都认同数据在经济活动中具有实际的价值。

1. 资产观的数据价值评估方法。资产观的数据价值评估方法将数据视为数据价值链中各参

与个体或组织所拥有且可进行交易的客观存在形态,认为数据价值是数据作为资产的价格。Coyle & Manley(2023)将数据视为无形资产,从企业估值视角介绍成本法、收益法和市场法三种数据价值评估方法。成本法认为数据价值取决于生产信息的成本,利用企业获取、收集、整理、分析与应用数据的成本对数据进行估值。收入法将数据价值与数据的未来收益联系起来,利用数据在未来经济生命周期内带来的净现金贴现值衡量数据价值。市场法认为数据价值取决于数据目前的市场情况,利用市场上数据可比产品的市场价来评估数据价值。Fang(2018)结合前景理论和实物期权定价模型构建数据价值评估模型。首先,该模型认为数据资产符合美式看涨期权的特征,即数据资产持有者在未来的某个时间、某个价格拥有购买或放弃购买的权力,并且数据资产的未来价格具有较高的不确定性。其次,引入前景理论的价值函数和权重函数计算数据资产持有者各状态下的数据资产价值。最后,以最小二乘蒙特卡洛模拟法为例进行数据价值评估的仿真分析。Sofia & Edward(2020)提出利用市场竞争信息反映数据的真实价值,其基本思路是,由数据供应者或需求者发起一个拍卖,让其他需求者或供应者对数据进行竞价,最终以最高出价或最低出价作为数据的成交价,即数据价值。Hao et al.(2023)提出了基于协议定价的数据估值法,其基本逻辑是,数据的使用者和提供者在数据交易过程中基于对自身数据需求、成本、收益、风险等因素的分析,通过中介机构或直接进行协商的方式,采用不同的博弈模型,寻求在交易实现双方最大效用的数据价格。在协议定价的框架下,最终达成的交易价被视为数据的实际价值。

2.资源观的数据价值评估方法。资源观的数据价值评估方法将数据视为可用于参与数据价值链的个人或组织内部经济活动的客观存在形态,认为数据价值是数据作为在各链条中投入要素所带来的效益或价值变动。由于数据价值很大程度上取决于数据用途,因此数据价值可分为货币和非货币两种形式。货币形式的数据价值体现在数据要素融入后的经济效益变动,例如利润增加和成本减少等。Coyle & Manley(2023)以优步公司的营业收入为例,基于反事实模拟法评估数据价值。他们通过改变机器学习算法中运用的数据量,计算优步公司运营过程中数据价值的上限。估计结果显示,数据为优步公司带来了高达47%的收入,该估计结果被用来进一步测算数据的价值。Farboodi et al.(2022a)提出利用充分统计量法对金融数据进行估值,以数据纳入投资者信息集前后的超额收益变动作为数据的价值。非货币形式的数据价值体现在数据要素融入后的效能变动,例如创新能力提升和生产效率提高等。Ghasemaghaei & Calic(2020)认为,数据价值体现在创新产品的数量方面,多类型、高频率更新的数据使企业能够及时洞察市场变化趋势,帮助企业快速产生新想法并迅速将其转化为创新产品,进而通过创新产品数量间接估算数据价值。Kim et al.(2020)认为,数据的价值体现为资源消耗的减少量,数据在农业生产活动中的整合可以实现农业生产的精准化,进而减少农业生产的资源消耗和环境污染,并以减少的消耗量来估算数据价值。Na et al.(2022)认为,数据的价值体现在企业运营效率的提升方面。企业拥有关于业务流程、结果的数据后,能够更好地监测和优化运营过程,提高运营效率,从而以效率提升值来测算数据价值。

## (二)数据链条中的微观主体福利效应

数据链条与市场产品从生产到消费的链条具有同步性,故参与数据价值链的微观主体主要包括产品的供需双方,即企业和消费者。数据能够给各类数据价值链主体带来福利效应,但这种福利效应可正可负。

1.微观主体福利正向效应。数据对企业和消费者福利的正向效应表现形式存在差异性。数据的有效利用能降低企业风险与成本、提高企业经营效率,为消费者提供个性化服务策略和提高消费者满意度。数据作为新型生产要素具有非竞争性、非排他性,边际成本递减和边际收益递增等特征,这些特征是数据提升企业生产经营效率、实现价值创造能力倍增、增加消费者福利的微观基础(Nimmy et al.,2019)。数据的本质是减少不确定性,降低风险溢价和资本成本,帮助企业增加投资,提高利润(Eeckhout & Veldkamp,2022)。企业通过历史数据的收集与分析,对市场风险进行识别与量化分析,并借助预测模型理解市场趋势、客户行为,以及资产类型、预期收益与风险等关键因素。

同时,企业通过实施及时的监测机制获取相关信息。这一综合性的数据分析使企业做出更精确的预测,有效降低未来业务环境中的不确定性(Kim et al.,2020),有助于企业更深层次地理解市场动态,为战略决策提供可靠的基础,增强企业的竞争力与长期稳健性,最终提升企业福利。此外,数据分析使企业能够实时监控整个供应链,迅速识别潜在的问题,即时调整和采取应对措施,提升供应链中各环节的性能和效率,降低运营成本,提高企业的整体竞争力(Soosay & Hyland,2015;Dhurandhar et al.,2015)。

大数据驱动的市场营销使得企业重新审视消费者与产品之间的关系。企业通过挖掘消费者数据,对消费者群体进行精细划分,提供个性化服务,以推动产品选择的多元性与市场灵活性的提升(Farboodi et al.,2022b)。消费者在这一框架下享有更为广泛的选择权,企业根据个体需求和偏好选择最适合自身的消费方案,从而提高整体的消费满意度,实现消费者福利的最大化。数据分析使得定价策略更灵活地随市场波动而调整,通过对销售数据和趋势的分析,企业能够更为准确地实施定价策略,实现对成本的有效控制(Chen & Schwartz,2015)。这种数据驱动的定价策略有助于创造更富竞争力的价格结构,使消费者受益于更具吸引力的价格,提升消费者剩余(Dubé & Misra,2023)。

2. 微观主体福利负向效应。数据对企业的负向福利效应体现在扩大企业差距方面。数据不平衡扩大了企业间的差距,尤其对弱势企业构成了更大的生存挑战。数据对消费者的负向福利效应主要体现在为“杀熟”、隐私侵犯等。部分企业利用自身对熟悉客户的优势,滥用数据权力,剥削熟悉客户消费者利益,侵犯消费者隐私,进而降低消费者的福利水平。企业之间的数据不平衡表现在数据资源和分析技术的差异。比如,中小型或新兴企业等弱势企业的数据库量较少、数据类型相对单一,而大数据技术依赖于大量且多样化的数据输入以减少不确定性,因此弱势企业大数据技术应用受限(Begenau et al.,2018;Farboodi et al.,2022b)。内部专业知识缺乏是弱势企业面临的另一个关键问题。Lam et al.(2017)指出,内部专业知识对于最大化发挥大数据潜力至关重要。企业需要拥有足够的科学、数据分析和数据管理能力才能确保有效地从数据中提取有价值的信息。许多中小企业或新兴企业缺乏这些专业知识,导致他们无法有效利用现有的数据资源,甚至导致数据分析的成本超过其带来的收益。由于这些弱势企业难以准确预测市场风险和业务趋势,资本成本较高,可能采取风险溢价较高的价格策略来弥补潜在的不确定性(Hennessy & Whited,2007),导致它们的产品或服务相对于成熟大型企业更加昂贵,从而增加客户流失的风险,影响市场份额,长期来看,这些弱势企业可能面临被市场上的优势企业取代或吞并的风险。

企业在对消费者数据进行共享时存在侵犯消费者隐私的现象,有时甚至会出现欺诈行为,降低消费者福利。消费者与企业之间往往存在信息不对称现象,企业在数据共享中处于信息优势地位,而消费者通常缺乏对数据使用和保护的深入了解,同时容易受到误导性数据营销的影响,这在很大程度上影响他们对产品和服务做出理性决策,从而削弱其利益(DeFusco et al.,2021)。企业在共享消费者数据过程中存在数据滥用现象,即企业可能会超出其获得的事先授权范围,以侵犯消费者的隐私权为代价来追求自身利益,或者向第三方披露消费者数据,导致消费者的隐私受到侵犯或泄露给未经授权的利益相关方(Kshetri,2014;Jones & Tonetti,2020)。

### (三)数据链条中的宏观主体福利效应

数据链条与国民经济各链条实现充分融合,体现为数据要素对经济整体的影响,即宏观主体福利效应,其影响可正可反。正向影响是围绕社会福利提升进行研究,如通过数据共享改善公共服务,优化资源配置,进而提升社会福利;负向影响是围绕数据在各个领域的应用导致数字鸿沟等社会福利下降进行研究。

1. 宏观主体福利正向效应。宏观主体福利效应主要源于数据共享。政府数据和私人数据均可以实现共享。政府数据共享与开放对于改善公共服务、创造公共价值具有重要作用;私人数据共享有助于优化资源配置,提升经济效率。政府多个领域充当主要数据的生产者和收集者,因此政府数据被广泛地视为一种资源。政府数据开放旨在主动、免费、无差别地向社会主体提供政府获得的数据,以激励其利用政府数据创造更大的经济价值。政府数据的开放对于改善公共服务和创



造公共价值具有重要作用。一旦政府数据能够以公开的形式获取,将为相关利益方提供巨大的机遇(Hitz-Gamper et al., 2019)。政府数据共享也能够提高政府运行的透明度(Park & Gil-Garcia, 2022)。开放政府数据能够在多个关键领域对社会产生显著积极影响,涵盖提高政府透明度和落实问责制、提升政府效率和有效性、促进公民的包容与赋权、推动经济增长等方面。通过开放获取公共资金支持的数据,可实现更广泛范围的公共投资回报,为决策者提供解决复杂问题所需的数据支持。此外,公众参与大规模数据集的分析,并通过协作网络共享和创造信息,有助于提高信息的共享和创新水平(Chun et al., 2010; Park & Gil-Garcia, 2022)。

私人数据共享有助于优化资源配置,提升经济效率,促进创新和发展。数据链条中的不同组织、企业或部门能够更充分地共享的私人数据进行分析,从而更精准地调整运营策略、优化产品或服务,提高生产效率和竞争力。开放银行旨在促进贷款机构竞争、使融资方受益,银行业采用开放式的技术和商业模式,允许第三方服务提供商通过开放的API与银行客户数据和服务进行交互(He et al., 2020),这种开放性模式旨在促进更多的创新、提供更多的选择以及改善金融服务的质量和效率。在供应链管理领域,数据供应商的数据共享可以帮助企业区分忠诚消费者和非忠诚消费者,从而实现数据供应商和企业之间的双赢局面(Chen et al., 2001)。不同行业的竞争对手可以共享客户数据,实施有针对性的定价,具有积极的社会福利效应(Jentzsch et al., 2013)。

2. 宏观主体福利负向效应。宏观主体福利负向效应主要源于数据链条中不同群体之间的数字鸿沟,进而引起的信息孤岛现象。数字鸿沟是指在不同地区、社会群体或经济阶层之间存在的数字技术和数据获取能力的差异,数字鸿沟的存在导致了信息孤岛的形成。信息孤岛是指某些地区或社会群体由于数字技术和信息获取能力滞后,与先进群体形成信息断层,进而使得这些地区或群体无法充分融入数据网络,形成了相对封闭的数据局部化现象(Steyaert & Gould, 2009; Liao et al., 2022)。一方面,信息孤岛的存在不仅妨碍了社会信息的全面流通,还限制了这些地区或群体在全球信息时代的参与和竞争力(Lv et al., 2018)。信息孤岛直接导致了数据碎片化,即不同地区或群体产生的数据以不同的标准和格式存在。这使得数据难以整合和共享,限制了数据的全面性和有效性(Steyaert & Gould, 2009)。这种碎片化的数据状态成为社会发展中的障碍,影响了信息的有效流通和利用,显著降低了社会福利。数字鸿沟限制了信息的流通,阻碍了合作机会的实现,从而影响了数据共享的协同效应。这将导致企业错失一些潜在的合作机会,降低了整个社会的生产力。合作机会的缺失不仅造成了经济损失,也影响了社会创新和发展的潜力。另一方面,信息的局部化和碎片化意味着一些地区或群体无法充分利用新兴技术和获取商业机会,这不仅限制了经济发展,还降低了整个社会的经济效益。因为信息难以共享,新兴技术的应用也受到限制,导致经济增长潜力的损失。信息孤岛会导致一些地区或社会群体相对于其他地区或群体处于信息劣势地位(Lv et al., 2018),加剧了社会的不平等现象,使得信息受限的地区或群体在社会和经济活动中面临更大的挑战,这种不平等不仅影响了个体和群体的发展,也削弱了整个社会的稳定性和包容性(Yuan et al., 2023)。

除了数据鸿沟引致信息孤岛给经济社会带来负向效应外,数据共享过多也会导致市场低效,进而出现负向福利效应。例如用户间数据存在关联性,市场上用户共享数据会引发相关联者数据泄露,降低数据价值。企业为了防止数据创造性遭受破坏,会选择囤积数据,从而导致数据的低效使用(Jones & Tonetti, 2020; Acemoglu et al., 2022)。

#### 四、数据价值链的生态研究

数据价值链的生态是与数据有关的组成部分相互作用、相互依赖和相互协作进而实现数据价值创造、传递和增值的动态系统。这一生态涵盖生态模式和生态治理两个关键方面:生态模式是描述数据如何在生命周期内流动、转化和被利用的规律;生态治理是指对数据价值链生态做出的指导和控制。数据价值链生态模式的持续运作依赖于数据价值链生态治理。

### (一)数据价值链的生态模式

数据价值链的应用广泛以及不同主体在不同领域的应用模式具有显著差异,相应地形成数字技术、在线平台、全面融合三种典型的数据价值链生态模式。数字技术模式的关键在于企业通过数字技术实现产品和服务、业务流程和经营等方面的创新;在线平台模式的关键在于以在线数据平台为媒介,实现数据的交换和共享;全面融合模式的关键在于全面整合数据,实现数据的多维度、多场景、多目的的持续利用。这三种生态模式体现了数据价值链生态中数据在不同主体、领域和层级之间的协调和协作。

1. 数字技术模式。数字技术模式主要体现在传统行业的数字化转型方面,注重数据的内部流动。不同行业的市场需求不同,采取的数字化转型策略也有所差异。航空业的数字化转型旨在提升服务效率和优化客户体验,自助航空服务是实现这一目标的关键手段。在线销售机票不仅减少了对旅行社佣金的支付,还有效降低了航空公司的运营成本(Yoon et al., 2006)。航空公司借助指纹和人脸识别等技术,减少乘客办理登机手续和安检的时间,改善客户体验(Heiets et al., 2022)。金融业的数字化转型旨在于在风险可控情况下获得最大化的投资收益,智能化投资是关键手段。金融机构利用先进的算法和人工智能技术分析投资者的风险偏好、财务目标和市场情况等信息,生成适合投资者需求的投资组合,并根据市场波动实时调整投资策略。这种投资模式能够避免人为错误和情绪影响,提高投资决策的准确性和效率(Niemand et al., 2021)。零售业的数字化转型旨在为消费者提供灵活的零售服务,个性化营销和沉浸式体验是关键手段。电商平台基于消费者的购物偏好和行为习惯,利用智能推荐系统为消费者提供个性化的商品推荐、定价策略和营销活动;通过虚拟现实、增强现实等技术为消费者创造更具沉浸感和参与度的购物体验(Buss et al., 2021)。此外,不同行业数字化转型的提升效应具有异质性。Na et al.(2022)在研究数字化转型与绩效时发现,数字化转型显著提升了绩效水平,其中,对市场依赖度较高的行业的提升效果尤其明显。

2. 在线平台模式。在线平台模式主要体现在在线数据平台的搭建和运营方面,强调数据的外部流通。在线平台的运作模式主要是数据交易和数据服务交易两种模式。在数据交易模式中,在线平台通过连接数据持有者和数据需求者,促成数据的买卖、租赁、授权和分成等交易。数据持有者在数据交易市场中展示其数据信息,数据需求者通过搜索、选择和交易获取所需的数据。这一市场的收费模式主要包括订阅收费、会员制收费、按需付费和增强型收费(Visconti et al., 2017)。随着数据规模和数据价值的增加,这种集中式数据交易市场出现了一些潜在问题,例如数据泄露和数据篡改。为实现安全、高效、透明和公平的数据交易,Sober et al.(2023)提出了基于区块链的数据市场系统。该系统在数据交易市场中引入区块链,直接连接数据持有者和数据需求者,不依赖任何第三方机构,实现了去中心化的数据交易。该系统首先通过数据需求者在数据筛选模块输入的内容,定位符合需求的数据;其次,在数据需求者和数据持有者的交互中,确定出售的数据及价格;最后,当买家完成支付后,系统运行买家任务并将结果返还给买家。在数据服务交易模式中,在线平台围绕数据需求方提供多样化的数据服务。Li et al.(2019)以亚马逊为例,介绍其通过数据服务部门(Amazon Web Services, AWS)提供的多种与数据相关的服务。AWS提供的仓库选址咨询服务基于消费者的地理位置和需求数据,不仅能够分析仓库的最佳位置、规模和布局,从而降低企业成本和提高运输效率,同时还可以根据市场的变化和消费者的反馈实时调整仓库的配置和策略,以适应不同的情况和需求。AWS提供的商品改进服务,能基于消费者的反馈和评价数据,评估商品的竞争力和满意度,从而为商品的改进、创新和营销提供依据和建议。AWS提供多种类型的云存储服务,允许数据持有方将其数据安全、可靠和高效地存储在云端,同时根据自身需求选择不同的存储类型、存储级别和存储位置等。

3. 全面融合模式。全面融合模式主要体现在数据的跨领域协作利用方面,关注数据的跨界流动。该模式包含数据整合和多领域的数应用两个关键步骤。数据的全面整合要求将各种来源、类型和格式的数据整合成统一的、可相互操作的数据资源池,这为不同领域数据的集中式管理和数据的协调提供了便利;多领域的数应用指的是不同领域的相关利益主体共用数据资源池并展开合

作,以推动创新和提高整体效能。数据整合是多领域数据应用的基础,多领域的数 据应用最大程度释放了数据整合后的价值,两个步骤相辅相成并形成良性循环(Villar et al.,2018)。医疗数据的应用涵盖患者治疗、医学研究和医疗管理等多个领域(Verboven et al.,2022;Bellini et al.,2023)。例如,医疗人员利用患者的基因组数据、生理参数和临床历史等数据制定更精准的治疗方案;研究人员整合各医疗机构和实验室的医疗数据进行更深入的疾病发病机理研究、新疗法的开发等工作;医疗机构通过分析患者就诊数据和医院运营数据,科学分配医疗资源,缓解医疗系统的压力;疾病防控机构通过分析患者的就诊历史、病毒监测数据追踪疫情迹象,及时采取措施阻止疾病的传播;保险公司通过分析患者的医疗数据,评估患者风险,制定更合理的保险策略。地理空间数据的应用同样贯穿城市规划、交通规划和环境保护等多个领域(Silva et al.,2018)。例如,城市规划者分析土地利用、人口分布和交通流量等数据,优化交通网络和路线,改善城市布局;环境机构监测自然资源的利用和环境变化,为可持续发展制定战略和政策提供支持。

## (二)数据价值链的生态治理

数据价值链的生态治理涵盖数据管理、数据安全监管和数据确权三个方面:数据安全监管确保数据在从生成到利用的整个过程中合规和安全地流动,是数据持续创造价值的前提条件;数据管理直接关系到数据的可信度和有效性,为数据创造更高层次的价值提供必要支持;数据确权确保数据的价值能够被公平地分配给相关利益主体,促进数据价值链生态的均衡发展。这三个方面的有机结合推动数据价值链在合规、安全和高效的生态中持续运转。

1. 数据管理。数据管理是指组织、存储、处理、分析和维护数据的系列活动,确保数据在整个生命周期内能够满足相关利益主体的需求,并成为可信、可用和可理解的资源。在数据从生成到利用的链式过程中,不同专业背景的人对数据管理的理解存在差异,这种差异产生于数据收集和数 据利用两种视角的数据管理方式。基于数据收集视角的数据管理关注数据在整个流程中能否被有效地结构化、理解和获取。Godoy & Saadatnejad(2017)提出利用元数据驱动法简化数据的管理和使用。该方法不仅详尽描述了数据集,包括数据的起源、内容、结构等方面的元数据信息,而且可以根据用户需求来可视化数据和导出元数据。Olsson & Hartley(2019)考虑到不同研究场景下的数据需求差异,开发用于管理数据的命令行工具。该工具首先将数据和元数据打包成一个统一的整体,即数据集。其次,它访问整个数据集和单个数据文件的元数据,并对其进行一致性检查。再次,它将数据集存储在不同的存储系统,以实现数据的可移植性和可访问性。最后,它提供应用程序编程接口,将数据集合并到现有的管道和工作流中,从而为数据的分析和利用提供更灵活的途径。基于数据利用视角的数据管理关注数据的可信度和可重用性。Fong(2001)强调数据管理的关键在于数据质量控制程序、数据的审计跟踪和最终数据库的质量量化。数据质量控制程序是指对数据的收集、录入、验证、修改和分析等关键过程的监督和检查,确保数据的准确性和完整性;数据的审计跟踪是指对数据的来源、变更、访问和使用等历史信息的记录,保证数据的可追溯性、可审查性和可复核性;数据库质量量化是指对数据质量的评估,保证数据的可靠性和可比较性。Even et al.(2007)指出,低质量数据会导致组织效率低下和资本损失,因此建议采取以下措施提升数据仓库的质量水平:一是自动化收集数据,以减少数据输入错误,提高数据的准确性;二是持续检查、纠正和清理数据以提高数据的完整性和消除数据的不一致性。

2. 数据安全监管。数据安全监管是为管理和监督数据的处理和流动而采取的措施,这包含对个人隐私的保护和对跨境数据流动的监管。个人隐私的保护是指防止个人隐私被非法收集、存储、使用、传输或公开而采取的措施。在全球范围内,个人隐私保护的监管立法模式可分为欧盟和美国两种模式,二者的区别主要体现在法律框架、基本原则和执行机构等多个方面(Makulilo,2020;Batlle & Van Waeyenberge,2023)。欧盟具有统一的个人隐私保护法,即《通用数据保护条例》(General Data Protection Regulations,GDPR)。GDPR赋予公民对个人数据的控制权,要求企业承担更多数据保护责任。欧盟设有专门的数据监管部门,负责监督和执行隐私法规,并对违规行为实施高额罚

款。美国的隐私法规涉及多个领域,如《隐私权保护法》《电子通讯隐私法》等,强调自由市场和行业自律,允许企业在一定范围内自主处理个人数据。美国的数据监管机构较为分散,如联邦贸易委员会(Federal Trade Commission,FTC)和联邦通信委员会(Federal Communications Commission,FCC),对违规行为的制裁取决于各个法规的规定。数据跨境流动监管是指对数据从一个国家或地区传输到另一个国家或地区时所进行的监督和管理。根据监管目的和方式,数据跨境流动监管可分为本地化和国际合作管制。数据本地化的主要目的是保障国家数据主权,要求数据在其产生或使用的国家或地区内进行存储、处理和管理,限制甚至禁止跨境传输。俄罗斯规定所有在俄罗斯境内收集的个人信息必须存储在俄罗斯境内的服务器上,未得到个人信息主体同意或法律许可的情况下,不得向境外传输(Shelepov,2022)。国际合作管制是促进数据跨境流动进行有序协作的协调机制,旨在维护数据流动和合作的利益。Zheng(2021)认为数据跨境流动监管可分为欧盟、美国和中国三种范式。欧盟寻求个人信息保护与数据自由流动的平衡,强调数据目的国应具备充分的数据保护水平;美国提倡数据能够在全全球市场上自由传输,以获得更多的商业利益;中国侧重保障数据的安全和合法使用,规定在数据跨境传输之前必须开展事先审核和批准,以确保敏感数据不被滥用或泄露。

3. 数据确权。数据确权是指为保障数据所有权和数据价值合理分配而采取的系列措施,关键在于数据所有权的界定。Thouvenin & Tamò-Larrieux(2021)提到个人数据不仅是一种信息资源,更是构成数据主体个人特征的重要组成部分,其独特性使个人数据不可被视为传统财产。在这一理念框架下,个人数据所有权赋予是不恰当的,因为个人无法自由放弃或完全让渡对其个人特征的权利。Saarikko et al.(2017)在处理数据所有权分配的冲突时指出,不存在一种适用于所有情况的通用方法,因此数据所有权的处理应根据特定国家、企业和合同来进行灵活调整。例如,在某些情况下,产生物联网数据的个体可能被认为拥有数据所有权,强调数据生成者对其数据的掌控权;在其他情况下,数据所有权被视为属于收集和分析数据的一方,如设备制造商,强调数据采集和处理者的权利。Asswad(2021)将数据所有权分为多个层次,涵盖从完全授权到存在一定限制的情况。第一层是数据所有权本身,授权人员被赋予对数据的完整所有权,包括对数据的控制权和决策权。这一层次强调数据所有者对其数据的绝对掌控,使其能够自主决定数据的用途、共享和修改。第二层是有限制的数据所有权,这些限制可能有利于第三方或公众,使他们能够主张一定的权利,反对数据所有者的专有权,以确保其他利益相关主体的权益。第三层是数据处置权和使用权,数据所有者在考虑一定限制的前提下,决定将数据使用权授予第三方。在这个层次上,数据所有者失去了对数据的主要控制权和决策权,只能在一定范围内对数据处理和利用进行监督或干预。Janecek(2018)强调,数据所有权应该根据数据的不同类型和特征进行区分。针对非个人数据,数据所有权可视为财产权。数据所有者能够对数据进行排他性、可转让性和可处分性的支配,即可以自由地获取、使用、转让、删除和保护数据,而不受任何限制或干预;针对个人数据,数据所有权可视为控制权。数据主体能够对数据进行自主性、可撤销性和可反馈性的控制,即可以根据自己的意愿和利益决定数据的使用或处理的方式、范围和目的,以及随时撤回或修改自己的同意或偏好。

## 五、研究展望

随着数字经济的蓬勃发展,数据价值得到充分发挥。本文在系统梳理数据价值链概念和演进阶段的基础上,以数据双重链条为主线阐述数据价值创造机制,对数据价值链中的价值评估方法和福利效应进行总结,进一步讨论数据价值链模式和生态治理研究。目前文献在数据价值链相关方面已取得了一定成果,但仍有以下几个问题值得进一步研究。

1. 数据价值链及其研究内容体系构建。通过对数据价值链概念的系统梳理可以发现,数据价值链是一个新兴领域,其研究内容尚未形成完整体系。从研究内容体系看,可以数据双重链条为主线,以参与数据价值链的利益相关者效应为核心,形成由数据价值链核心特征、数据估值与数据资产定价的基础理论问题、数据要素融合与链效能的形成机制、数据互联的复杂性、数据福利等构成的内容体系。

2. 数据价值链的位置测度。数据价值链位置对新质生产力的形成具有重要作用。数据价值链位置包括数据本身在生产运行环节的位置,也包括数字技术在全球数据价值链的位置。新质生产力是由生产要素创新性配置、产业深度转型升级、技术革命性突破而催生的当代先进生产力。针对面向新兴领域的新质生产力,既需要客观准确测度数据价值链各个运行链条中的位置,又需要客观认识数据技术价值链位置。目前,数据价值“链条”特征并没有得到足够关注,如“链融合”“链重构”等问题。另外,全球价值链研究融合数据技术在链条中的位置、融合程度、跨境数据流动等也应受到关注。

3. 数据价值链效能识别与测度问题。现有研究更多讨论数据价值,故文献更关注数据价值,但数据流转和融合是价值产生的必备条件;数据在链式中实现流转和融合,进而产生价值,这涉及数据价值链效能识别与测度问题。在数据价值测度方面,现有文献并没有把数据生产和流转纳入“链条”中,特别是数据流转在“链条”中的价值测度没有被充分挖掘,而原有数据价值测度技术过于“静止”。“链”效能测度技术更多应用在传统的价值链(如产业链)中,没有考虑数据与产业之间的充分融合问题。价值链位置测度侧重产业之间的关联性,其测度方法对新兴技术特别是数字技术的纳入成分较少。基于此,根据数据价值链“双链”特征,研究数据价值链的基本特征、数据效能形成机制、多链融合效能识别和测度等问题都是进一步的研究方向。

4. 多链联动调控机制结合问题。现有研究侧重某一链条的生态,对“多链”生态纳入较少。现有的许多链条均存在数据驱动特征,不同链条的主体具有差异性,数据价值链的效能研究要求实现多主体高效协作,研究物理空间、社会空间和信息空间中数据串联、信息融合和业务协同方法,研究数据驱动的跨链耦合、多链联动的价值关系发现。同时,数据价值链生态与数据驱动的跨链耦合具有强关联性,其联合调控机制有赖于对数据价值链的构造;在全球数据价值链攀升过程中更需要一国经济主体在价值链生态位上具有显著优势。

5. 数据价值链的风险管理与风险传播问题。数据的赋能和安全具有双重特征,与传统价值链相比,数据价值链的结构必然更加复杂和多变。故从信息与数据安全、产业安全和价值链风险等视角分析不同模态下潜藏的数据价值链风险,探索不同模态数据价值链中数据要素蕴含的风险隐患,以及数据要素与不同产业链融合所产生的市场风险、信息传递风险、道德信用风险、系统风险、战略风险等常见风险,这些都是进一步研究需要关注的方向。

#### 参考文献:

- Acemoglu, D. et al. (2022), “Too much data: Prices and inefficiencies in data markets”, *American Economic Journal: Microeconomics*, 14(4):218—256.
- Alizadeh, M. et al. (2023), “Development of a customer churn model for banking industry based on hard and soft data fusion”, *IEEE Access*, 11:29759—29768.
- Asswad, J. (2021), “Data ownership: A survey”, *Information*, 12(11), No.465.
- Attard, J. et al. (2016), “Data value networks: Enabling a new data ecosystem”, 2016 IEEE/WIC/ACM International Conference on Web Intelligence, IEEE.
- Battle, S. & A. Van-Waeyenberge (2023), “EU-US data privacy framework: A first legal assessment”, *European Journal of Risk Regulation*, DOI:10.1017/err.2023.67.
- Begenau, J. et al. (2018), “Big data in finance and the growth of large firms”, *Journal of Monetary Economics*, 97:71—87.
- Bellini, V. et al. (2023), “From Big Data’s 5Vs to clinical practice’s 5Ws: Enhancing data-driven decision making in healthcare”, *Journal of Clinical Monitoring and Computing*, 37(5):1423—1425.
- Bhathal, G.S. & A.S. Dhiman (2018), “Big data solution: Improvised distributions framework of hadoop”, Second International Conference on Intelligent Computing and Control Systems, IEEE.
- Brandas, C. et al. (2016), “Data driven decision support systems: An application case in labour market analysis”, *Romanian Journal of Information Science and Technology*, 19(1—2):65—77.
- Buss, K.P. et al. (2021), “Systemic rationalization 4.0. How competition and business models are shaping digitalization in retail, logistics and financial services”, *Berliner Journal für Soziologie*, 32(1):35—68.

- Chen, Y. & M. Schwartz (2015), "Differential pricing when costs differ: A welfare analysis", *RAND Journal of Economics*, 46 (2):442–460.
- Chen, Y. et al.(2001), "Individual marketing with imperfect targetability", *Marketing Science*, 20(1):23–41.
- Chun, S.A. et al.(2010), "Government 2.0: Making connections between citizens, data and government", *Information Polity*, 15(1/2):1–9.
- Coyle, D. & A.Manley(2023), "What is the value of data? A review of empirical methods", *Journal of Economic Surveys*, DOI:10.1111/joes.12585.
- Curry, E.(2016), "The big data value chain: Definitions, concepts, and theoretical approaches", in: J.Cavanillas et al. (eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Bata in Europe*, Springer.
- DeFusco, A.A. et al.(2021), "Measuring the welfare cost of asymmetric information in consumer credit markets", NBER Working Paper, No.29270.
- Del Vecchio, P. et al.(2022), "A structured literature review on Big Data for customer relationship management (CRM): Toward a future agenda in international marketing", *International Marketing Review*, 39(5):1069–1092.
- Dhurandhar, A. et al.(2015), "Big data system for analyzing risky procurement entities", 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery.
- Dubé, J.P. & S.Misra(2023), "Personalized pricing and consumer welfare", *Journal of Political Economy*, 131(1): 131–189.
- Eeckhout, J. & L.Veldkamp(2022), "Data and market power", NBER Working Paper, No.30022.
- Eling, M. & M. Lehmann (2017), "The impact of digitalization on the insurance value chain and the insurability of risks", *Geneva Papers on Risk and Insurance — Issues and Practice*, 43(3):359–396.
- Eltantawy, R. et al.(2015), "Towards supply chain coordination and productivity in a three echelon supply chain: Action research study", *International Journal of Operations & Production Management*, 35(6):895–924.
- Even, A. et al.(2007), "Economics-driven data management: An application to the design of tabular data sets", *IEEE Transactions on Knowledge and Data Engineering*, 19(6):818–831.
- Falcao, J.D. et al.(2021), "ISACS: In-store autonomous checkout system for retail", *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMVUT)*, 5(3), No.99.
- Fang, Y.(2018), "The real option approach for assessment of big data asset based on prospect theory", 2018 International Conference on E-Business, Information Management and Computer Science.
- Farboodi, M. et al.(2022a), "Valuing financial data", NBER Working Paper, No.29894.
- Farboodi, M. et al.(2022b), "Where has all the data gone", *Review of Financial Studies*, 35(7):3101–3138.
- Faroukhi, A.Z. et al.(2020a), "A multi-layer big data value chain approach for security issues", *Procedia Computer Science*, 175:737–744.
- Faroukhi, A. Z. et al.(2020b), "Big data monetization throughout big data value chain: A comprehensive review", *Journal of Big Data*, 7(1):1–22.
- Fong, D.Y.T.(2001), "Data management and quality assurance", *Drug Information Journal*, 35(3):839–844.
- Gaimon, C. & K.Ramachandran(2021), "The knowledge value chain: An operational perspective", *Production and Operations Management*, 30(3):715–724.
- Ghasemaghaei, M. & G.Calic(2020), "Assessing the impact of big data on firm innovation performance: Big data is not always better data", *Journal of Business Research*, 108:147–162.
- Godoy, O. & B.Saadatnejad(2017), "ACCESS climate data management", *Ambio*, 46:464–474.
- Hao, J. et al.(2023), "The evolution of data pricing: From economics to computational intelligence", *Heliyon*, 9(9), e20274.
- He, Z. et al.(2020), "Open banking: Credit market competition when borrowers own the data", NBER Working Paper, No.28118.
- Heiets, I. et al.(2022), "Digital transformation of airline industry", *Research in Transportation Economics*, 92, No. 101186.
- Hennessy, C.A. & T.M.White(2007), "How costly is external financing? Evidence from a structural estimation", *Journal of Finance*, 62(4):1705–1745.

- Hitz-Gamper, B.M. et al.(2019), “Balancing control, usability and visibility of linked open government data to create public value”, *International Journal of Public Sector Management*, 32(5):457–472.
- Hu, H. et al.(2014), “Toward scalable systems for big data analytics: A technology tutorial”, *IEEE Access*, 2:652–687.
- Janecek, V.(2018), “Ownership of personal data in the Internet of Things”, *Computer Law & Security Review*, 34(5): 1039–1052.
- Jentzsch, N. et al.(2013), “Targeted pricing and customer data sharing among rivals”, *International Journal of Industrial Organization*, 31(2):131–144.
- Jones, C.I. & C.Tonetti(2020), “Nonrivalry and the economics of data”, *American Economic Review*, 110(9):2819–2858.
- Kasim, H. et al.(2012), “Data value chain as a service framework: For enabling data handling, data security and data analysis in the cloud”, Eighteenth International Conference on Parallel and Distributed Systems, IEEE.
- Kim, W.S. et al.(2020), “A review of the applications of the Internet of Things (IoT) for agricultural automation”, *Journal of Biosystems Engineering*, 45(4):385–400.
- Koutsoukis, N. & G.Mitra(2003), *Decision Modelling and Information Systems: The Information Value Chain*, Springer Science & Business Media.
- Kriksciuniene, D. et al.(2015), “Process optimization and monitoring along big data value chain”, *Business Information Systems Workshops*, 228:75–86.
- Kshetri, N.(2014), “Big data’s impact on privacy, security and consumer welfare”, *Telecommunications Policy*, 38(11):1134–1145.
- Lam, S. et al.(2017), “Leveraging frontline employees’ small data and firm-level big data in frontline management: An absorptive capacity perspective”, *Journal of Service Research*, 20(1):12–28.
- Li, W.C.Y. et al.(2019), “Value of data: There’s no such thing as a free lunch in the digital economy”, Bureau of Economic Analysis Research Paper, <https://www.bea.gov/research/papers/2018/value-data-theres-no-such-thing-free-lunch-digital-economy>.
- Liao, S. et al.(2022), “Revisiting the development trajectory of the digital divide: A main path analysis approach”, *Technological Forecasting and Social Change*, 179, No.121607.
- Lv, J. et al.(2018), “Application on information island with information visualization and software engineering”, Fifth International Conference on Systems and Informatics, IEEE.
- Lv, Z. et al.(2022), “Cross-layer optimization for industrial internet of things in real scene digital twins”, *IEEE Internet of Things Journal*, 9(17):15618–15629.
- Lynn, M.(1991), “Scarcity effects on value: A quantitative review of the commodity theory literature”, *Psychology Marketing* 8(1):43–57.
- Ma, S. et al.(2020), “Big data driven predictive production planning for energy-intensive manufacturing industries”, *Energy*, 211, No.118320.
- Makulilo, A.B.(2020), “The long arm of GDPR in Africa: Reflection on data privacy law reform and practice in Mauritius”, *International Journal of Human Rights*, 25(1):117–146.
- Müller, F. et al.(2019), “Digitization in wood supply—A review on how Industry 4.0 will change the forest value chain”, *Computers and Electronics in Agriculture*, 162:206–218.
- Na, C. et al.(2022), “Digital transformation of value chains and CSR performance”, *Sustainability*, 14(16), No.10245.
- Niemand, T. et al.(2021), “Digitalization in the financial industry: A contingency approach of entrepreneurial orientation and strategic vision on digitalization”, *European Management Journal*, 39(3):317–326.
- Nimmy, J.S. et al.(2019), “Literature review on supply chain collaboration: Comparison of various collaborative techniques”, *Journal of Advances in Management Research*, 16(4):537–562.
- Olsson, T.S.G. & M.Hartley(2019), “Lightweight data management with dtool”, *PeerJ*, 7, e6562.
- Park, S. & J.R.Gil-Garcia(2022), “Open data innovation: Visualizations and process redesign as a way to bridge the transparency-accountability gap”, *Government Information Quarterly*, 39(1), No. 101456.
- Petrova-Antonova, D. et al.(2018), “Modelling of educational data following big data value chain”, Eighteenth International Conference on Computer Systems and Technologies, IEEE.
- Roy, D. & M.Dutta(2022), “A systematic review and research perspective on recommender systems”, *Journal of Big*

- Data*, 9(1), No.59.
- Saarikko, T. et al.(2017), “The Internet of Things: Are you ready for what’s coming?”, *Business Horizons* 60(5): 667–676.
- Semlali, B. E. B. et al. (2020), “SAT-ETL-Integrator: An extract-transform-load software for satellite big data ingestion”, *Journal of Applied Remote Sensing*, 14(1), No.018501.
- Shaik, T. et al.(2023), “A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom”, *Information Fusion*, 102, No.102040.
- Shelepov, A.(2022), “Approaches of BRICS countries to data regulation”, *Vestnik Mezhdunarodnykh Organizatsii — International Organisations Research Journal*, 17(3):212–234.
- Silva, B.M. et al.(2018), “Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities”, *Sustainable Cities and Society*, 38:697–713.
- Singh, P.J.(2020), “Breaking up Big Tech: Separation of its data, cloud and intelligence layer”, *Data Governance Network Working Paper*, No.9.
- Sober, M. et al.(2023), “A blockchain-based IoT data marketplace”, *Cluster Computing — The Journal of Networks Software Tools and Applications*, 26(6):3523–3545.
- Sofia, D. & A. S. Edward (2020), “Auction based game theory in cognitive radio networks for dynamic spectrum allocation”, *Computers & Electrical Engineering*, 86, No.106734.
- Soosay, C.A. & P.Hyland (2015), “A decade of supply chain collaboration and directions for future research”, *Supply Chain Management — An International Journal*, 20(6):613–630.
- Steyaert, J. & N.Gould (2009), “Social work and the changing face of the digital divide”, *Journal of Social Work*, 39(4):740–753.
- Thouvenin, F. & A.Tamò-Larrieux(2021), “Data ownership and data access rights”, in: M.Burri(ed.), *Big Data and Global Trade Law*, Cambridge University Press.
- Verboven, L. et al. (2022), “A treatment recommender clinical decision support system for personalized medicine: Method development and proof-of-concept for drug resistant tuberculosis”, *BMC Medical Informatics and Decision Making*, 22(1), No.56.
- Villar, A. et al.(2018), “Integrating and analyzing medical and environmental data using ETL and Business Intelligence tools”, *International Journal of Biometeorology*, 62(6):1085–1095.
- Visconti, R.M. et al.(2017), “Big data-driven value chains and digital platforms: From value co-creation to monetization”, in: A.K.Somani & G.C.Deka(eds.), *Big Data Analytics: Tools and Technology for Effective Planning*, Chapman and Hall/CRC.
- Weerakkody, V.J.P. et al.(2019), “A case analysis of E-government service delivery through a service chain dimension”, *International Journal of Information Management*, 47:233–238.
- Wirén, M. et al.(2019), “Big data value chain: Making sense of the challenges”, *Lecture Notes in Artificial Intelligence*, 11701:125–137.
- Yeh, C.C. et al.(2013), “A hybrid KMV model, random forests and rough set theory approach for credit rating”, *Knowledge-Based Systems*, 33:166–172.
- Yoon, M.G. et al.(2006), “Impact of e-business on air travel markets: Distribution of airline tickets in Korea”, *Journal of Air Transport Management*, 12(5):253–260.
- Yuan, K. et al.(2023), “The impact of digital divide on household participation in risky financial investments: Evidence from China”, *Changing Societies & Personalities*, 7(1):113–129.
- Zhang, T. et al.(2017), “Welfare economics of review information: Implications for the online selling platform owner”, *International Journal of Production Economics*, 184:69–79.
- Zhao, J. et al.(2023), “A cross-platform personalized recommender system for connecting E-commerce and social network”, *Future Internet*, 15(1), No.13.
- Zheng, G.(2021), “Trilemma and tripartition: The regulatory paradigms of cross-border personal data transfer in the EU, the US and China”, *Computer Law & Security Review*, 43, No.105610.



## Research Progress in the Data Value Chain

LI Zhenghui, XU Yanting and LU Siting  
(Guangzhou University, Guangzhou, China)

**Summary:** Data value creation activities are divided into fundamental and incremental value activities. Data value chain facilitates the creation of fundamental data value and the value enhancement during the circulation process in the various chains of data collection, transmission, storage, analysis, and application. In light of the absence of a scientific and systematic research paradigm in the current academic research in this field, this paper focuses on the data value chain, systematically elucidating its conceptual framework, evolutionary stages, operational mechanisms, value assessment, and the value chain ecosystem. Additionally, it provides insights into future research directions.

Initially, the paper conducts a systematic examination of the conceptual framework and evolutionary stages inherent in the data value chain. Diverse professional backgrounds contribute to the formulation of diverse definitions for the data value chain, resulting in three typical conceptual perspectives based on data flow, data processing technology, and value augmentation. The evolutionary stages of the data value chain originate from various welfare effects brought by information (the information economy stage), undergo gradual embedding of digital technologies (the digital technology embedding stage), and ultimately take shape through the convergence of dual chains (the data value dual-chain fusion stage).

Subsequently, the paper analyzes the mechanism of data value creation, focusing on the data production and usage process chains and the physical chains integrating data with industries. The data value chain demonstrates typical dual-chain characteristics, where one chain encompasses the process from data production to usage, while the other involves the integration of data with industries. Research on value creation during data operations primarily analyzes its mechanism based on the functioning of these chains, while value creation in physical chains emphasizes the value generated through the integration of data with industries.

Furthermore, the paper systematically summarizes methods for assessing data value, elaborating on the welfare effects of data on stakeholders from both micro and macro perspectives. According to different assessment perspectives, data value assessment methods are categorized into asset-oriented and resource-oriented approaches. Participants in the data value chain are mainly divided into two major categories: macro and micro-level ones. Originating from the theory of firms in a market economy, the micro-level participants in the market include both suppliers and demanders of products, namely, businesses and consumers. The micro-level welfare effects of data primarily focus on the benefits that data brings to various entities in the market. The macro-level welfare effects of data concentrate more on the impact of data elements on the overall economy.

Moreover, the paper provides an overview of research progress in the ecological aspects of the data value chain, covering its models and governance. The data value chain's ecological models describe how data flows, transforms, and is utilized throughout its lifecycle, featuring three typical patterns of the digital technology model, online platform model, and comprehensive integration model. Ecological governance refers to the guidance and control on the data value chain ecosystem, encompassing data management, data security supervision, and data rights confirmation.

In conclusion, the paper outlines future research directions for the data value chain. Although existing literature has made certain achievements in various aspects of the data value chain, several issues still warrant further investigation, such as the characteristics of the data value chain and its related content system construction, the analysis of the value chain pattern and its effects along with data integration, the identification and measurement of data value chain empowerment, the integration of multi-chain coordinated control mechanisms, and the risk management and propagation in the data value chain.

**Keywords:** Data Value Chain; Value Creation; Value Assessment; Data Ecosystem

**JEL Classification:** C80, D60, I30

(责任编辑:木子)

(校对:何伟)