

# 数据价值实现视角的数据产业 定位构成与中国规模测算\*

蔡跃洲 何甜甜

**摘要:**本文着眼于数据价值实现过程中所涉环节主体,明确数据产业的功能定位和构成分类,剖析各部门间的价值转移链条,据此构建数据产业增加值测算框架,估算中国数据产业增加值规模。研究表明:(1)数据携带的有效信息是其潜在价值源泉,数据作为生产投入参与生产经营和价值创造后,潜在价值才转化为现实价值,数据才成为生产要素。数据的技术—经济特征使其能通过效率提升、数据融合、多场景复用实现价值创造。(2)数据产业是为数据价值实现提供基础性、专业化、支撑性服务或产品的经济活动部门,与提炼数据有效信息有关的产业都可纳入其中,具体可分为“直接核心层”“间接支撑层”和“前端基础层”。(3)前端基础层数字产业通过软硬件、基础设施等形成直接核心层和间接支撑层的固定资产,并在提供相关服务过程中将部分价值转移至数据资产或数据产品,最终投入生产运营、完成数据价值实现,从而形成一条价值转移链。(4)2017年以来中国数据产业增加值快速增长,年均增速为17.84%,2023年达到16637亿元,2025年预计达22418亿元,是数据作为关键要素作用日益强化的直接反映;“十五五”时期,中国数据产业增加值规模(名义值)有望达到5万亿元,成为重要的新兴支柱产业,这将为深化数据资源开发利用提供有力支撑。

**关键词:**数据价值实现 数据产业 规模测算 产业生态 价值转移链

## 一、引言

2010年前后,互联网、4G/5G通信、云计算、大数据(分析)、人工智能等新一代信息技术的不断突破和大规模商业化应用,全方位提升了数据生产(收集)、存储、传输、处理等各环节的效率,同时大幅降低了相关成本,“大数据”逐渐成为社会各

\* 蔡跃洲,中国社会科学院大学应用经济学院、中国社会科学院数量经济与技术经济研究所,邮政编码:100732,电子邮箱:caiyuezhou@cass.org.cn;何甜甜(通讯作者),中国社会科学院大学应用经济学院,邮政编码:102445,电子邮箱:1344228764@qq.com。基金项目:国家社会科学基金重大项目“我国数字经济发展状况测评研究”(23VRC088);国家自然科学基金重大项目“宏观大数据建模和预测研究”(71991475)。感谢审稿专家的宝贵意见,文责自负。

界关注的焦点 (Manyika et al., 2011; 蔡跃洲和马文君, 2021)。此后, 数据在经济社会活动中发挥着越来越重要的作用, 成为各种数字经济新模式、新业态不断涌现的要素基础。

数据中包含的有效信息是其成为生产要素、参与价值创造的源泉。有效信息的提炼和传递, 可以提高生产经营各环节的协同性, 进而提升效率, 创造出额外的价值。与此同时, 数据的非物质性、非竞争性、易复制性、非消耗性、可增值性等技术-经济特征, 进一步放大了生产经营各环节协同效应 (Levitin & Redman, 1998; 蔡跃洲和马文君, 2021)。一方面, 多源数据融合可以释放更多有效信息, 进一步提高协同性; 另一方面, 数据可以在不同场景中复用, 从而在更广范围内实现协同性的提升。这些也是释放“数据要素乘数效应”的重要内在机制。深化数据资源开发利用, 就是要发挥好上述机制的作用, 使得数据能够在更大范围内融合、更多场景中复用, 以充分实现数据价值。为此, 需要有更多的数据资源在更广范围内实现流通交易。国家数据局提出的“供得出、流得动、用得好、保安全”, 正是从数据供给、数据流通交易以及数据使用等环节共同发力, 在保证数据安全前提下确保上述机制有效运行, 充分释放数据价值。

开发利用数据资源、实现数据价值是一项复杂的系统工程。数据持有方、数据使用方、数据商、第三方中介机构及数据行政管理部门等利益相关主体 (stakeholder), 围绕数据价值实现在不同环节上承担特定功能并协同发展。在此过程中, 不同主体逐步形成专业化分工, 进而衍生出各类细分的数据产业, 如数据清洗、数据加工、数据评估、数据分析、数据合规等, 加上各种标准、制度、规则, 共同构成数据产业生态体系。因而, 数据产业生态体系是支撑数据价值实现和数据资源开发利用的重要基石。近年来, 国家有关部门发布的《“十四五”大数据产业发展规划》《关于促进数据产业高质量发展的指导意见》等政策文件, 为数据产业发展锚定了重点方向。然而, 数据价值的实现面临数据供给意愿不足、流通交易不畅、应用场景有待挖掘拓展等各种现实障碍。数据相关技术—经济特征所带来的权属划分不清晰、数据质量参差不齐、个人隐私及公共安全风险识别难等现实挑战, 是前述障碍产生的根源。而数据产业发展不充分、产业生态体系不完善则是上述障碍未能有效克服进而制约数据价值实现的重要原因。

数据在微观生产运营、宏观经济发展中发挥的作用已经得到广泛认可 (Jones & Tonetti, 2020; 谢康等, 2020; Farboodi & Veldkamp, 2021)。现有文献已基于不同角度展开对数字经济影响的研究: 微观层面通过“利益相关者法”<sup>①</sup>和特定业务场景下数据应用溢出效应的价值估算 (McKinsey & Company, 2021), 揭示了数据价值创造和生产率提升的内在机制; 宏观层面则通过数据资本的测度 (Calderón & Rassier, 2022; 刘涛雄等, 2023; 许宪春等, 2025) 与大数据产业、数字经济规模估算 (杨政等, 2022; IDC & the Lisbon Council, 2022) 衡量数据的经济增长贡献等。但数据价值实

<sup>①</sup>Anmut (2020), “Different Data Valuation Methodologies”, Available from: <https://www.anmut.co.uk/different-data-valuation-methodologies/>.

现的微观机理与宏观测算体系的有效衔接不足,数据和数据产业相关概念、范围边界划分等仍处于探讨阶段,且缺乏系统性理论分析框架和各界公认的标准(蔡跃洲,2024),从而在客观上制约了对数据产业发展状况的把握。

综上所述,为推动数据产业健康发展,以更好支撑数据价值实现和数据资源开发利用,需要对数据产业发展态势形成客观准确的认识。本文聚焦于数据通过影响资源配置进而直接或间接创造经济增加值机制与测算的研究目标,从(使用)价值创造的本质出发,在界定数据价值实现内涵的基础上,梳理数据价值实现的理论机制并对关联概念予以辨析;结合数据流通交易相关实践,对数据价值实现过程进行分解,厘清所涉及的相关环节、主体及其功能定位;据此确定数据产业的范围边界,并对其进行分层分类,从价值转移链条角度剖析各构成部分的内在关联。在此基础上,给出数据产业规模统计测算的基本思路框架,并尝试对中国数据产业的增加值规模进行测算,以期把握中国数据产业发展状况提供参考。

## 二、数据价值实现内涵机制及关联概念辨析

### (一)数据价值实现的内涵

广义的数据通常是指基于测度或统计产生的可用于计算、讨论和决策的事实信息;20世纪50年代后,随着电子计算机及关联信息技术的推广和应用,数据逐步指向以“0”“1”二进制编码形式存在的、需要借助计算机及关联设备进行存储和处理的记录(Farboodi & Veldkamp, 2021; Veldkamp & Chung, 2024; 蔡跃洲和马文君, 2021)。数字经济时代,狭义的数据被限定为以“0”“1”编码存在的统计调查记录;随着数据处理、分析技术的不断提升,“记录”的范围也由传统量化的统计、调查结果拓展为几乎所有能够从中获取有效信息或知识、且以“0”“1”二进制编码形式存在的记录,包括音视频、文本语料、程序代码等(Statistics Canada, 2019)。

据此,包含(携带)有效信息或知识是数字经济时代“数据”最为根本的特质,而有效信息(或知识)也是数据作为一种资源所具备潜在价值的源泉。根据DIKW模型,数据资源从数据(data)到信息(information)再到知识(knowledge)进而上升到智慧(wisdom)的形态跃迁,既代表着数据中所包含有效信息的不断凝练升华,更是其潜在价值不断显性化的过程(Ackoff, 1989)。当然,要将有效信息的潜在价值转化为真实的经济价值,则需要将数据以特定形态投入到生产经营中,同其他要素一道参与价值创造。参与创造增加值可以看作是数据作为生产要素实现价值创造的终极标志,而该过程可以称为“数据要素化和价值化”。

### (二)数据价值实现的理论机制

同其他有形、具象化的资源要素相比,数据具有非物质性、即时性、非竞争性、非排他性、易复制性、非消耗性、非置换性、可增值性等技术-经济特征(Levitin & Redman, 1998; 蔡跃洲和马文君, 2021),使其能够通过三种机制参与创造增加值并实现价值,详见图1。

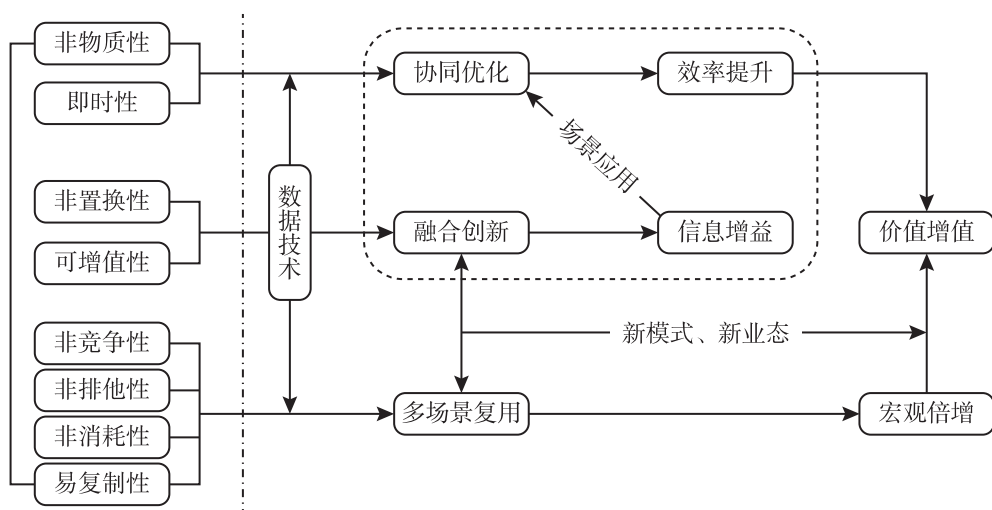


图1 数据要素乘数效应与数据价值实现机制

一是特定场景下的效率提升机制(蔡跃洲和牛新星,2021;戚聿东和王言,2026)。在每个具体的生产经营场景中,有效信息或行业知识的提取和传递,能够降低生产经营各环节、各主体间的信息不对称,增强各环节、各主体间的协同性,进而提高生产经营效率,实现额外的价值创造。

二是数据融合带来的信息和价值增益机制(蔡跃洲等,2025)。不同来源、规模的数据,通过整合通常会带来更多额外的有效信息或行业知识,并有望在特定场景下发挥出更大的效率提升作用,从而创造出更多的增加值,更大限度地实现数据价值。

三是多场景复用带来的宏观价值倍增机制(蔡跃洲和马文君,2021;李海舰和赵丽,2021)。数据的非竞争性、非排他性、非消耗性和易复制性决定了同一数据集可以同时多个场景中发挥出前两项价值创造机制的作用,从而在宏观层面产生价值创造的倍增效应。

上述三种机制中,效率提升机制是最根本的价值实现机制,是数据参与增加值创造的微观基础;数据融合带来的价值增益本质上还是效率提升机制在发挥作用,只是由于数据本身具有“非置换性”和“可增值性”,融合后的数据产生了信息增益,进而带来更大幅度的效率提升、资源配置优化以及催生新产品、新业态、新模式等(李娅和李卿芸,2025),从而创造更多增加值;关于宏观倍增机制,更多来自于数据所具备的“非竞争性”“可复制性”两项技术-经济特征。

另外,上述三种机制恰恰是数据要素乘数效应实现的三种途径。在数字经济实践中,要充分发挥乘数效应、最大限度实现数据要素价值化,很重要的前提条件就是在保证安全的同时实现数据充分流通。国家数据局于2024年提出的“供得出、流得动、用得好、保安全”,可以看作是上述前提条件的具象化表述。

### (三)数据价值关联概念辨析

需要特别说明的是,当前社会各界对“数据价值化”“数据资产化”“数据资本化”“数据资源化”等相关概念的理解存在一定程度的混淆,为夯实后文理论基础,本文对上述概念予以辨析和区分。

在会计核算和统计核算(SNA,国民账户体系)中,数据资产化同以往的研发费用资产化类似,都是将原本应该记录到成本或费用项下、直接冲抵当期收入(利润)的支出项目,转为能够长期发挥作用的资产。其中,在会计处理中,通常将其作为无形资产,而在经济统计(SNA)中则将之作为广义的固定资产。每年入账的数据资产,都会作为一种资本予以累积(Rassier et al., 2019; 罗玫等, 2023)。在马克思主义政治经济学框架中,购买的机器设备安装后形成了企业的固定资产,而这些机器设备同厂房等其他固定资产一道,“只有在它的买者手里,即在生产上使用它的资本家手里,才成为固定资本”。<sup>①</sup>换句话说,所有机器设备、厂房和其他建筑物,当它们被企业作为生产资料在生产中发挥作用时,都可以抽象为无差别的固定资本。数据资产化过程与有形机器设备的资产化过程是一致的,只是将有形的机器设备替换为无形的数据资源。因此,数据资产化后,也就完成了数据固定资本形成,或者说数据资本形成。从这个意义上来看,数据资产化就是数据资本化。

从资本形成视角来看,数据资产化就是数据资本化的过程。然而,在数字经济实践中,基于数据价值评估实现数据资产入表后,开展数据质押贷款、直接融资等活动时,不可避免地将数据与货币价值联系在一起。由于这些活动涉及各种金融资本,因此很多时候也被称之为“数据资本化”。不过,上述活动更应该被称之为“数据金融化”。前述马克思政治经济学意义上的数据资本化与金融意义上的数据资本化容易产生混淆。前者是抽象的、作为生产投入要素的固定资产或无形资产,而后者则指的是资金/现金意义上的资本。为避免概念上的混淆,可将这类活动界定为“数据金融化”或“数据资金化”。确切地讲,这类行为可以称之为“数据金融化”,是微观主体在数据价值化尚未完成之前,对预期收益的提前变现。而从宏观层面和全社会价值创造来讲,这类行为的完成,并不意味着数据参与价值创造、完成了价值化。

此外,数据只要被收集和存储,就具备潜在价值,就是数据资源。所以不需要经过专门清洗、处理转化,就已经是资源了,无须专门定义所谓“数据资源化”。而数据价值实现则可以看作是数据资源开发利用的结果和标志。

## 三、数据产业功能定位、构成分类及价值转移

数据产业作为数字经济时代的产物,其范围边界尚无公认的清晰界定。本部分将尝试梳理数据产业与数据价值实现之间的关系,以确定数据产业在数据价值实现过程中的功能定位,并以此为基础,确定数据产业所涵盖的范围边界,对其构

<sup>①</sup>《资本论》(第二卷),人民出版社2018年版,第179页。

成要素进行分类,进而厘清数据产业各构成部分在推动数据价值实现中的价值转移情况,以及由此形成的价值转移链和数据产业生态系统。

### (一)价值实现中的数据产业功能定位

数据价值实现的核心是数据中的有效信息在生产经营中得到利用。从宏观社会化大生产角度来看,数据价值的实现需要数据作为投入要素参与到生产经营各环节;从微观层面来看,数据在价值创造中的贡献,或者说数据实现的价值,最终体现在生产经营活动及其成效中。在农业经济以及工业经济时代,数据在生产经营和价值创造中已经发挥了重要作用,甚至衍生出以电信业为代表的独立产业部门,但数据及数据信息处理更多是生产经营活动的伴生物,收集处理过程较为简单,通常能被直接作为投入参与价值创造(蔡跃洲,2022)。然而,进入数字经济时代后,经济社会运行实践中,数据所携带的有效信息得到利用并非自然而然就能实现。

日益提升的算力、密集的互联网、无处不在的计算、索引识别和大规模分布式存储构成了大数据现象蓬勃发展的基础。高德纳公司(Gartner)、国际数据公司(IDC)及IBM等业界知名机构从数据相关特征来定义大数据,最早将其归纳为3V,即大容量(volume)、高速率(velocity)、多样性(variety);此后又增加了价值性(value)、真实性(veracity)、易变性(variability)等特征(Zikopoulos et al., 2012; Kitchin, 2014)。大数据具有“大容量”“真实性”“价值性”等特点,使得数据对经济社会运行的支撑作用日益凸显。2010年以来,以电子商务、网约车、外卖等为代表的各种消费互联网新业态,还有传统产业数字化转型过程中涌现的智慧农业、智能制造等各种模式创新,其运行背后的关键支撑都源于海量数据中有效信息的提取和运用。

与此同时,大数据所具备的“多样性”“易变性”“高速率”等特征,也意味着海量数据的有效信息提取,越来越成为高度专业化的技能;有效信息提取的相关活动,已不再是生产经营活动的伴生物。事实上,在绝大多数场景中,数据(资源)都需要进行加工、提炼、分析,才能作为生产要素投入使用,前期通常还会涉及数据的收集整理、流通交易及融合等环节。这些环节虽不能直接实现数据所含有效信息的价值转化,却是数据价值最终实现前的必要准备。随着数据资源规模的快速增长,上述环节变得愈发专业化、精细化,并逐步形成数据相关的新兴产业部门,由此便衍生出各个细分的“数据产业”。从这个意义上讲,数据产业作为大数据时代的产物,其功能定位是为数据价值实现提供基础性、专业化、支撑性服务或产品的经济活动部门,是数据资源开发利用的重要支撑。

### (二)数据产业范围边界与构成分类

根据前述功能定位,数据产业作为支撑服务数据价值实现的衍生物,所有与提炼数据有效信息相关的产业都可以纳入数据产业的范畴。当然,在数据产业具体定义和分类上,不同学者以及数据相关行政管理部门虽然也形成一些共识,但在划分视角和标准方面都存在差异,鲜有文献给出数据产业的明确范围边界及构成成分

类,具体情况大致如下。

第一,数据产业是大数据时代产物和新兴产业已成为共识。从既有的相关研究来看,部分国内外文献提出了“大数据产业”的概念,并从产出可操作信息(produce actionable information)或“大数据”产生、收集、存储、传输、组织管理等相关活动角度,对“大数据产业”进行界定(周曙东,2017;翟丽丽等,2023;Martin,2020;杨玫等,2022)。在数据相关宏观管理实践中,工业和信息化部出台的《“十四五”大数据产业发展规划》,明确指向“大数据产业”,将其定义为新兴产业;<sup>①</sup>但从该规划中大数据产业所涉及的具体内容来看,与国家发展改革委等部门出台《关于促进数据产业高质量发展的指导意见》(后文简称《数据产业高质量发展》)中的“数据产业”存在很大重叠。<sup>②</sup>而国家数据局发布的第二批名词解释,同样将数据产业界定为新兴产业,是“利用现代信息技术对数据资源进行产品或服务开发,并推动其流通应用所形成的新兴产业”。<sup>③</sup>

第二,数据产业作为服务于大数据信息知识提取(或价值实现)的新兴产业,涵盖了数据投入生产经营活动中与有效信息提取直接相关的各环节,包括从数据的收集到存储、从预处理到加工分析等。不同研究和机构对于这些直接相关环节的表述略有差别。基于数据价值链视角,Faroukhi et al.(2020)将数据投入使用前的各环节分为数据生成、数据采集、数据预处理、数据存储、数据分析、数据可视化。2021年11月,工业和信息化部发布的《“十四五”大数据产业发展规划》政策文件中,列举了“数据生成、采集、存储、加工、分析、服务”,<sup>④</sup>《数据产业高质量发展》提到了数据采集、存储、治理、分析、流通、应用等关键环节,国家数据局则表述为“数据采集汇聚、计算存储、流通交易、开发利用……”。<sup>⑤</sup>

第三,为数据有效信息提取提供间接支撑的活动也被纳入广义的数据产业范畴,包括数据安全、数据技术、数据基础设施服务等。例如,《“十四五”大数据产业发展规划》中指出“围绕‘数据资源、基础硬件、通用软件、行业应用、安全保障’的大数据产品和服务体系初步形成”,<sup>⑥</sup>相当于将数据技术相关硬件软件、数据安全保障以及行业应用等都纳入到广义的数据产业范畴;《数据产业高质量发展》和国家数据局的名词解释,将数据资源的开发利用、安全治理和数据基础设施建设纳入数据产业范畴;而周曙东(2017)则将大数据传输与组织管理、大数据价值服务等经济活动纳入到数据产业,从而将其边界拓展到提供间接支撑的数据技术产品服务乃至更为广泛的大数据与传统产业融合应用。

基于以上归纳及前述定位可以看出,虽然数据产业是数据资源衍生出来的新兴产业,但有着涵盖范围广、涉及链条长的特点。这一特点源于数据技术-经济特

①④⑥高婴劼,2022:《〈“十四五”大数据产业发展规划〉解读》,《数字经济》第3期。

②《促进数据产业高质量发展》,《人民日报》2024年12月31日。

③⑤国家数据局2025年3月29日发布的《数据领域名词解释(第二批)》中,定义数据产业是“利用现代信息技术对数据资源进行产品或服务开发,并推动其流通应用所形成的新兴产业,包括数据采集汇聚、计算存储、流通交易、开发利用、安全治理和数据基础设施建设等”。

征带来的价值实现的复杂性,也是造成数据产业范围边界不甚明晰的重要原因。为了对数据产业的范围边界进行更为清晰的界定和刻画,本文围绕“支撑数据价值实现”这个终极目标和定位,将其构成划分为三个层级,即直接核心层、间接支撑层和前端基础层。

直接核心层是指在数据价值实现之前,为提取数据有效信息直接提供服务或产品的经济活动部门,具体包括数据生成、收集、存储、清洗、标注、流通、加工、分析等环节。每一个环节的活动都会让数据中所包含的有效信息相比上一个环节更为凝练,更接近于被用作生产投入。间接支撑层是指那些虽不直接参与有效信息提取,但却为提取数据信息并最终实现数据价值提供保障支撑的经济活动部门,包括数据资产价值评估、数据安全合规审查等第三方服务,以及通信、算力等基础设施支撑服务。前端基础层是指与数据价值实现有关联但又明显不能划归为数据产业的经济活动部门,主要涵盖前端数据生成处理能力形成、数据基础设施建设等涉及的数字产业,是实现数据生成、收集、传输、处理等活动的物质技术基础。

### (三)数据产业价值转移链及产业生态系统

上述数据产业广义范围边界所涵盖的三个层级,围绕数据价值实现,在推动数据要素化、价值化过程中,不断进行价值的沉淀和后向转移,形成了一条价值转移链。

第一,价值从前端基础层所属各数字产业向直接核心层或间接支撑层转移。前端基础层各数字产业提供的数字产品服务,如射频识别设备、数字监控设备、通信设备制造、计算机设备制造等,将以软硬件设备、数字基础设施等方式形成直接核心层和间接支撑层的固定资产。这些固定资产是数据产业直接核心层和间接支撑层生成、收集比特数据,对外提供数据产品服务的物质技术基础。在上述过程中,前端数字产业所提供产品服务的价值,通过固定资本形成,沉淀于以固定资产形式呈现的数字设备、基础设施中。而这些固定资产在使用过程中,部分价值又会转移到数据收集、处理、分析及数据基础设施运行服务等环节中。于是,便完成了价值由前端基础层向直接核心层或间接支撑层的转移。

第二,直接核心层和间接支撑层所创造的价值,转化沉淀为数据资产或数据产品服务。数据产业直接核心层的数据收集、存储、流通、加工、分析等活动和间接支撑层的通信算力服务、第三方服务等,将原始数据所包含的有效信息不断凝练,形成各类数据资产或数据产品服务。在此过程中,直接核心层和间接支撑层各项服务所创造的增加值也逐步转移到数据资产或数据产品服务中,从而完成了价值的第二次转移。

第三,数据资产或数据产品服务参与价值创造,将其自身所包含的价值转移并体现在最终所实现的数据价值之中。数据价值实现过程,就是企业将积累的各类数据资产和购买的数据产品服务,作为要素投入生产经营活动,与其他要素一起创造增加值。这是数据有效信息参与价值创造、完成数据价值实现的最终环节,也是

数据要素化价值化的终点和目标。数据要素所贡献的价值,有一部分是用于补偿数字资产形成或数据产品服务获取方面的支出。这样,就相当于直接核心层、间接支撑层以及前端基础层各细分行业的一部分价值,被间接转移到数据要素最终所实现的价值中。

从前端基础层的相关数字产业,到中间各类直接、间接的数据服务,再到末端数据价值的最终实现,整个数据产业链背后对应着价值的逐次转移。前序环节对应细分行业所创造的增加值会通过固定资本形成等方式予以积累沉淀,其部分价值会逐步转移到下一个环节(见图2)。从数据生成、收集开始,经过若干环节,逐步形成可直接用作生产投入的数据资产或数据产品服务,并在投入生产经营活动后转变为数据要素,实现了数据要素化价值化。而数据要素所创造的增加值中,又有一部分是前序数据资产或数据产品服务转移而来的。

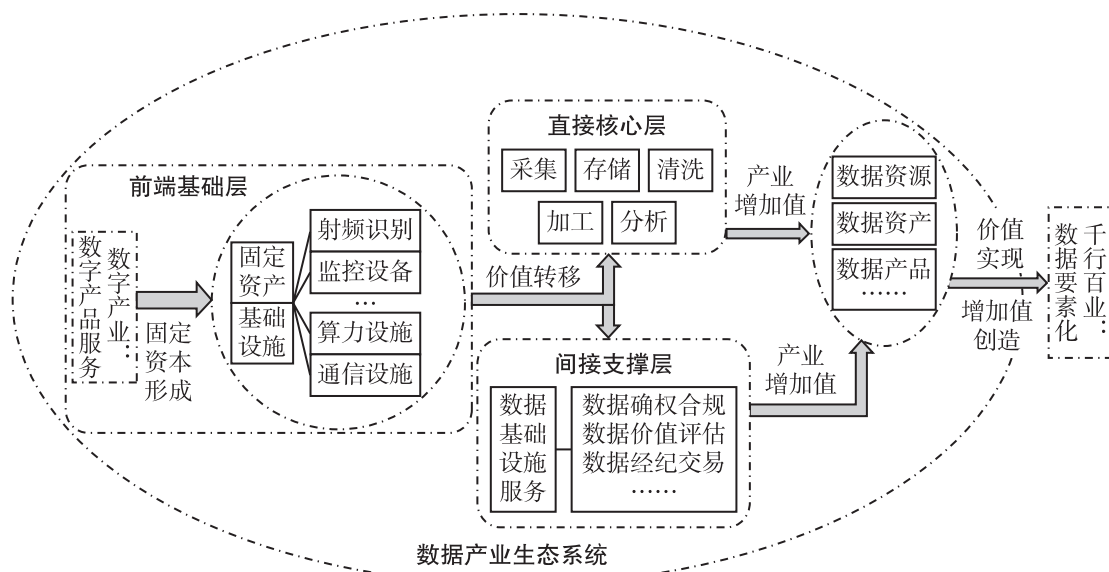


图2 数据价值转移链与数据产业生态系统

伴随着数据价值的流转,数据产业构建了一条完整的价值转移链。数据产业每个环节所对应的增加值,理论上都将由数据要素最终所实现的价值予以弥补,但这种补偿往往是滞后、错期且非一一对应的。毕竟每一次参与价值创造的数据要素,都只是前序环节相关活动成果的一部分。从更长的时间跨度和数据产业价值链整体来看,数据产业各细分部门创造的增加值都源于末端的数据价值实现,可以看作是对最终实现价值前的一种分配。

需要指出的是,上述价值转移和分配能否顺利实现,取决于数据产业价值链上各环节间的相互协调和有效衔接。而在数据价值化实践中,数据产业价值链的构成主体多元化、链路交织复杂等特征更为突出。首先,每个环节都由众多从事相关业务的微观主体组成。其次,数据所具备的非竞争性、非排他性、易复制性等技术一经

济特征使得数据在权属确定、风险控制等方面的复杂性更为突出,并增加了不同主体、不同环节协调衔接的复杂性。再次,数据作为数据价值化源泉的有效信息本身,在数据流通实践中具有信息不对称特征,使得各环节中出现的数据资产、数据产品服务面临定价难的现实困境,客观上制约了数据在不同环节间的顺畅流转。

事实上,数据产业价值链上的各类主体围绕数据价值实现,形成了一个庞大复杂的数据产业生态系统。在这个生态系统中,除了各类微观企业外,以监管部门和行业主管部门为代表的政府部门、第三方中介机构等也是重要的主体。数据产业价值链各环节的有效衔接和数据价值的顺利实现,需要在不断完善各种标准、规则、制度的基础上,促进利益相关主体的协调配合,从而保障数据产业生态系统的有机运转。

#### 四、数据产业规模统计测算思路框架

数据产业作为支撑数据价值实现的重要载体是数字经济时代的产物,其规模则可以成为衡量数字经济发展状况的重要指标。科学准确测算数据产业总量规模并开展系统化结构分析,是把握数字经济发展态势和推进数据产业有序发展的基础性工作。然而,现行官方统计体系是工业经济时代的产物,无法直接提供表征数据产业规模的统计指标及基础信息。因此,本部分将以前述第三部分确定的数据产业生态范围边界和构成分类为依据,在分析数据产业规模与数据规模测算关联的基础上,提出数据产业规模统计测算的基本思路,基于生产法搭建数据产业统计测算框架,并按照划分的三个层次,分别确定数据产业各层级所涵盖或涉及的细分行业。

##### (一)数据产业规模统计测算思路

2023年2月,中共中央、国务院印发的《数字中国建设整体布局规划》中指出,“到2025年,基本形成横向打通、纵向贯通、协调有力的一体化推进格局……数据资源规模和质量加快提升,数据要素价值有效释放……”。<sup>①</sup>由此,社会各界对“数据规模”的关注度再次攀升。由于数据具有非物质性、非消耗性、非竞争性、非置换性及可增值性等技术-经济特征,其规模表征相比其他传统要素显得更为复杂,特别是以货币计量的数据规模,可以是数据资产规模,也可能是数据交易规模,还可能是数据价值实现规模(蔡跃洲,2024)。由于现行以SNA为核心的官方统计体系缺乏直接表征数据或数据产业规模的统计指标,各类渠道给出的数据规模/数据产业规模基本上都是研究者在关联信息收集的基础上,运用特定测算方法得到的估算结果。不同场景或语境下,测算的标的对象及其涵盖范围往往存在巨大差异,同时由于不同机构、学者采用的测算方法以及依据的基础信息也不尽相同,使得测算结果之间不具备可比性。

从数据规模测算实践来看,早在本世纪初便有部分机构和学者尝试测算数据在价值创造方面的贡献。英国公平贸易办公室(Office of Fair Trading, 2006)通过走

<sup>①</sup>《中共中央国务院印发〈数字中国建设整体布局规划〉》,《人民日报》2023年2月28日。

访调查估算了2005年英国公共部门数据带来的生产者剩余和消费者剩余。<sup>①</sup>ACIL Tasman(2008、2009)针对了针对澳大利亚和新西兰空间地理数据所带来的广义经济收益的测算,其思路是考察地理空间数据对各行业生产率的影响,据此推算出其对GDP增长的影响。与之类似,Vickery(2011)借鉴ACIL Tasman的做法测算出2009年欧盟27个国家使用地理信息数据所产生的价值。与此同时,随着学术界和实务界对数据(无形)资产属性的认同,越来越多的学者在既有的SNA框架下开展数据资产规模的测算(Rassier et al., 2019;刘涛雄等, 2023;许宪春等, 2025);还有研究聚焦数据市场,对数据交易规模进行测算(IDC & the Lisbon Council, 2022;蔡跃洲和刘悦欣, 2022)。与数据价值规模类似,无论是数据资产规模还是数据交易规模,不同研究的测算结果也都存在较大差异。

数据产业规模测算与数据规模测算密切关联,却有着显著的差异。一方面,从测算的标的对象来看,前者是支撑数据价值实现的相关产业,后者则是数据本身的规模,包括数据价值实现(增加值创造)的规模。另一方面,测算数据产业规模,更多是对流量规模进行测算,即数据产业在特定时间段内的产值、增加值规模;而对于数据规模,既可以测算存量数据资产规模,也可以测算特定时间段内数据交易规模或数据价值实现规模。从数据相关产业规模测算实践来看,现有研究相对较少。Vickery(2011)估算了2009年欧盟空间地理信息产业规模,对照本文前述第三部分对数据产业的构成分类,其涵盖范围应涉及直接核心层中采集、存储、清洗、加工、分析各环节。杨玫等(2022)在对大数据企业进行界定的基础上,基于骨干企业大数据营收占比等信息,通过折算和加总估算出2016—2021年期间大数据产业累计产值达4.49万亿元,预计2025年全年产值能突破3万亿元。由于缺乏各界普遍接受的统计测算框架,上述测算结果尚未形成统一口径。为此,本部分拟将前述数据产业构成分类与以现行SNA为核心的官方统计体系相结合,从生产法角度对数据产业生态各组成部分所涵盖细分行业进行归集,构建数据产业规模统计测算框架,并尝试从行业数据产业关联度入手,提出相关参数设定依据。

采用生产法测算数据产业规模,是充分利用现有统计体系下可以提供的细分行业增加值进行统计核算,大致的思路及步骤如下:(1)分层确定所涉及细分行业:按照数据产业构成分类中“直接核心层”“间接支撑层”和“前端基础层”的划分,对照《数字经济及其核心产业统计分类(2021)》《国民经济行业分类(2017)》等文件,逐一确定数据产业生态体系三大模块(层级)各自所涉及的细分行业(可视具体情况细分到四位码或六位码)。(2)收集基础信息、设定相关参数:收集各关联细分行业的增加值、行业产值/产出或所属企业营业收入等基础信息,并设定表征每个细分行业与数字经济相关程度的行业数据业务调整系数、行业增加值率等参数。(3)分层加总,形成多口径数据产业规模:利用细分行业增加值测算结果及设定的关联系数,计算其归属于数据产业的增加值,即细分行业数据产业增加值,据此对数据

<sup>①</sup>Office of Fair Trading (2006), “The Commercial Use of Public Information (CUPI)”, OFT861, December.

产业生态系统三大模块各自所涵盖的细分行业数据产业增加值进行加总,从而获得直接核心层、间接支撑层和前端基础层各自所对应的增加值规模,以及三个不同口径的数据产业增加值规模。

(二)数据产业涵盖涉及细分行业梳理

梳理数据产业生态各层级所涉及的细分行业,是从生产法角度对其产业增加值规模开展统计测算的基础。

数据产业直接核心层即数据采集、存储、清洗、加工、分析等活动,其对应的细分行业,涉及《数字经济及其核心产业统计分类(2021)》中的“03 数据技术应用业”,而在《国民经济行业分类(2017)》中则涉及“门类 I:信息传输、软件和信息技术服务业”及“门类 L:租赁和商务服务业”“门类 J:金融业”,具体细分行业在上述两项分类中的名称及代码如下表 1 所示。在所涉及的 6 个(六位码)细分行业中,“030306 互联网数据服务”“030405 信息处理和存储支持服务”“040303 金融信息服务”相关活动大部分都归属于数据产业直接核心层,其增加值可以全部计入直接核心层规模。至于其他 3 个细分行业,则只有部分活动可以归属于数据产业直接核心层,在统计直接核心层规模时,需要结合行业经济活动的数据产业关联度剥离出行业的数据产业增加值贡献。

表 1 数据产业直接核心层涉及细分行业名称代码

| 代 码    |        |        | 名称                          | 说明                                                                                                     | 国民经济<br>行业代码及名称<br>(2017)                               |
|--------|--------|--------|-----------------------------|--------------------------------------------------------------------------------------------------------|---------------------------------------------------------|
| 大<br>类 | 中<br>类 | 小<br>类 |                             |                                                                                                        |                                                         |
| 03     | 0303   | 030306 | 互联网数<br>据服务                 | 指以互联网技术为基础的大数据处理、云存储、云计算、云加工、区块链等服务活动                                                                  | 6450 互联网数据<br>服务                                        |
|        |        | 030405 | 信息处理<br>和存储支<br>持服务         | 指供方向需方提供的信息和数据的分析、整理、计算、编辑、存储等加工处理服务,以及应用软件、信息系统基础设施等租用服务,包括在线企业资源规划(ERP)、在线杀毒、服务器托管、虚拟主机等             | 6550 信息处理和存<br>储支持服务                                    |
|        |        | 030407 | 地理遥感<br>信息及测<br>绘地理信<br>息服务 | 指各类地理遥感信息服务活动和遥感测绘服务活动,包括互联网地图服务软件、地理信息系统软件、测绘软件、遥感软件、导航与位置服务软件、地图制图软件等地理遥感信息服务,以及卫星定位测量、导航定位服务等遥感测绘服务 | 6571 地理遥感信息<br>服务<br>7441 遥感测绘服务<br>7449 其他测绘地理<br>信息服务 |
|        |        | 030408 | 动漫、游<br>戏及其他<br>数字内容<br>服务  | 指将动漫和游戏中的图片、文字、视频、音频等信息内容运用数字化技术进行加工、处理、制作并整合应用的服务活动,以及数字文化、数字体育等其他数字内容服务                              | 6572 动漫、游戏数<br>字内容服务<br>6579 其他数字内容<br>服务               |

续表 1

| 代 码    |        |        | 名称         | 说明                                                                         | 国民经济<br>行业代码及名称<br>(2017) |
|--------|--------|--------|------------|----------------------------------------------------------------------------|---------------------------|
| 大<br>类 | 中<br>类 | 小<br>类 |            |                                                                            |                           |
| 04     | 0403   | 040303 | 金融信息<br>服务 | 指向从事金融分析、金融交易、金融决策或者<br>其他金融活动的用户提供可能影响金融<br>市场的信息(或者金融数据)的服务,包括<br>征信机构服务 | 6940 金融信息服务               |
|        | 0404   | 040409 | 数字广告       | 指在互联网平台投放,以广告横幅、文本链<br>接、多媒体等形式,为外部客户提供宣传推<br>广服务的活动                       | 7251 互联网广告<br>服务          |

数据产业间接支撑层,包括通信、算力等生产性服务以及数据标准制定、数据确权与合规、数据经纪与交易等数据第三方服务。这些活动对应的细分行业涉及《数字经济及其核心产业统计分类(2021)》的“03 数据技术应用业”“04 数字要素驱动业”,而在《国民经济行业分类(2017)》中则涉及“门类 I:信息传输、软件和信息技术服务业”和“门类 M:科学研究和技术服务业”,具体细分行业在上述两项分类中的名称及代码如下表 2 所示。

表 2 数据产业间接支撑层涉及细分行业名称代码

| 代 码 |      |        | 名称                   | 说明                                                                                                                                | 国民经济                                      |
|-----|------|--------|----------------------|-----------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| 大 类 | 中 类  | 小 类    |                      |                                                                                                                                   | 行业代码及名称<br>(2017)                         |
| 03  | 0302 | 030201 | 电信                   | 指利用有线、无线的电磁系统或者光电系统,传<br>送、发射或者接收语音、文字、数据、图像、视频<br>以及其他任何形式信息的活动                                                                  | 6311 固定电信服务<br>6312 移动电信服务<br>6319 其他电信服务 |
|     | 0304 | 030402 | 信息<br>系统<br>集成<br>服务 | 指基于需方业务需求进行的信息系统需求分析<br>和系统设计,并通过结构化的综合布线系统、计<br>算机网络技术和软件技术,将各个分离的设备、<br>功能和信息等集成到相互关联的、统一和协调<br>的系统之中,以及为信息系统的正常运行提供<br>支持的服务活动 | 6531 信息系统集成<br>服务                         |
|     | 0304 | 030404 | 运行<br>维护<br>服务       | 指各种运行维护服务活动,包括基础环境运行<br>维护、网络运行维护、软件运行维护、硬件运行<br>维护、局域网安装调试服务、局域网维护服务以<br>及其他运行维护服务、网络技术支持服务等                                     | 6540 运行维护服务                               |
|     | --   | --     | --                   | 指专利、商标、版权、软件、集成电路布图<br>设计、技术秘密、地理标志等各类知识产权的代理、<br>转让、登记、鉴定、检索、分析、咨询、评估、<br>运营、认证等服务                                               | 7520 知识产权服务                               |

表 2 列示了 8 个六位码/四位码细分行业,其中,“030201 电信”“030402 系统集成服务”“030404 运行维护服务”都是《数字经济及其核心产业统计分类(2021)》中列示的六位码细分行业;而“7520 知识产权服务”则属于未在《数字经济及其核心产业统计分类(2021)》中出现的《国民经济行业分类(2017)》四位码细分行业。前 3 个细分行业依托基础设施或其他硬件设备为直接核心层提供技术支撑,虽然属于直接核心层相关细分行业运行不可或缺的前提条件,但真正支撑数据产业直接核心层运行的活动只占一部分。知识产权服务业的相关业务,如知识产权公证、知识产权法律服务与数据产业间接支撑层的“合规评估、公证及仲裁服务”在法律逻辑应用上具有相近性,以及软件著作权、知识产权许可/交易和商业秘密等,也随数据产业发展延伸至数据资产的核心载体(如算法、模型、数据产品),但这也只是其业务活动的一部分。因此,在核算各细分行业归属于数据产业间接支撑层的增加值规模时,也需要结合其业务活动的数据产业关联系数予以分拆。

数据产业前端基础层所涵盖的数字产业部门生产的硬件、软件等数字产品,是直接核心层和间接支撑层开展活动、提供服务的物质技术基础。前端基础层对应的细分行业涉及《数字经济及其核心产业统计分类(2021)》的“01 数字产品制造业”“03 数据技术应用业”以及《国民经济行业分类(2017)》四位码细分行业,具体的细分行业代码及名称详见表 3。前端基础层相关行业的产品服务会被直接核心层和间接支撑层用于资本积累并形成固定资产。某种意义上可以将其视为数据产业直接核心层及间接支撑层相关行业的中间投入。因此,这些行业用作电信、互联网及相关服务业中间投入的规模可作为测算其增加值规模的重要依据。

表 3 数据产业前端基础层涉及细分行业名称代码

| 代 码    |        |        | 名称            | 说明                                                                        | 国民经济<br>行业代码及<br>名称<br>(2017) |
|--------|--------|--------|---------------|---------------------------------------------------------------------------|-------------------------------|
| 大<br>类 | 中<br>类 | 小<br>类 |               |                                                                           |                               |
| 01     | 0101   | 010101 | 计算机整机<br>制造   | 指将可进行算术或逻辑运算的中央处理器和<br>外围设备集成计算整机的制造,包括硬件与软<br>件集成计算机系统的制造、来件组装计算机的<br>加工 | 3911 计算机整<br>机制造              |
|        | 010102 |        | 计算机零部<br>件制造  | 指组成电子计算机的内存、板卡、硬盘、电源、<br>机箱、显示器等部件的制造                                     | 3912 计算机零<br>部件制造             |
|        | 010103 |        | 计算机外围<br>设备制造 | 指计算机外围设备及附属设备的制造,包括输<br>入设备、输出设备和外存储设备等制造                                 | 3913 计算机外<br>围设备<br>制造        |

续表 3

| 代 码    |        |        | 名称             | 说明                                                                                                                                                          | 国民经济<br>行业代码及<br>名称<br>(2017)                             |
|--------|--------|--------|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|
| 大<br>类 | 中<br>类 | 小<br>类 |                |                                                                                                                                                             |                                                           |
|        |        | 010104 | 工业控制计算机及系统制造   | 指一种采用总线结构,对生产过程及机电设备、工艺装备进行检测与控制的工具总称;工控机具有重要的计算机属性和特征,如具有计算机 CPU、硬盘、内存、外设及接口,并有操作系统、控制网络和协议、计算能力、友好的人机界面;工控行业的产品和技术非常特殊,属于中间产品,是为其他各行业提供可靠、嵌入式、智能化的工业计算机制造 | 3914* 工业控制计算机及系统制造                                        |
|        |        | 010105 | 信息安全设备制造       | 指用于保护网络和计算机中信息和数据安全的专用设备的制造,包括边界安全、通信安全、身份鉴别与访问控制、数据安全、基础平台、内容安全、评估审计与监控、安全应用设备等制造                                                                          | 3915 信息安全设备制造                                             |
|        | 0102   | 010201 | 通信系统设备制造       | 指固定或移动通信接入、传输、交换设备等通信系统建设所需设备的制造                                                                                                                            | 3921 通信系统设备制造                                             |
| 03     | 0301   | 030101 | 基础软件开发         | 指能够对硬件资源进行调度和管理、为应用软件提供运行支撑的软件开发活动,包括操作系统、数据库、中间件、各类固件等                                                                                                     | 6511 基础软件开发                                               |
|        |        | 030102 | 支撑软件开发         | 指软件开发过程中使用到的支撑软件开发的工具和集成环境、测试工具软件等的开发活动                                                                                                                     | 6512 支撑软件开发                                               |
|        |        | 030104 | 其他软件开发         | 指其他未列明软件的开发活动,如平台软件、信息安全软件等                                                                                                                                 | 6519 其他软件开发                                               |
| --     | --     | --     | 体育场馆建筑和其他房屋建筑业 | 指体育馆工程服务、体育及休闲健身用房屋建设活动和其他房屋建筑业                                                                                                                             | 4720 体育场馆建筑<br>4790 其他房屋建筑业                               |
| --     | --     | --     | 建筑安装           | 指建筑物主体工程竣工后,建筑物内各种设备的安装活动,以及施工中的线路敷设和管道安装活动;不包括工程收尾的装饰,如对墙面、地板、天花板、门窗等处理活动                                                                                  | 4910 电气安装<br>4920 管道和设备安装<br>4991 体育场地设施安装<br>4999 其他建筑安装 |

续表 3

| 代 码 |    |   | 名称    | 说明                   | 国民经济<br>行业代码及<br>名称<br>(2017) |
|-----|----|---|-------|----------------------|-------------------------------|
| 大   | 中  | 小 |       |                      |                               |
| 类   | 类  | 类 |       |                      |                               |
|     |    |   | 建筑装饰、 |                      | 5011 公共建筑                     |
|     |    |   | 装修和其他 | 指对建筑工程后期的装饰、装修、维护和清理 | 装饰和装修                         |
| --  | -- |   | 建筑业   | 活动,以及对居室的装修活动        | 5012 住宅装饰                     |
|     |    |   |       |                      | 和装修                           |
|     |    |   |       |                      | 5013 建筑幕墙                     |
|     |    |   |       |                      | 装饰和装修                         |

(三)基础信息处理与参数设定

基于前文对数据产业分层所涉及细分行业的梳理,理论上可以在现有的国民经济行业分类和生产法统计核算框架下,收集到各细分行业的增加值,作为测算数据产业增加值规模的基础信息。现行的统计公开资料并不能完全直接提供细化至四位码、六位码的行业增加值数据,但很多细分行业可以收集到其行业总产值(总产出或者行业内主要企业营业收入)信息。如果能够得到合理的细分行业增加值率,便可大致估算出其增加值规模。据以假定,在国民经济行业分类中,细分行业(四位码)的增加值率与其所属大类行业(二位码)的增加值率基本一致,具体可表达为式(1)。这样利用大类行业(二位码)增加值率即可对细分行业增加值规模进行估算,见式(2)。将式(1)简单变换可进一步得式(3)。

行业增加值率= $\frac{\text{行业增加值}}{\text{行业产出}} \approx \frac{\text{细分行业增加值}}{\text{细分行业产出}}$  (1)

细分行业增加值=细分行业产出×行业增加值率 (2)

$\frac{\text{细分行业产出}}{\text{行业产出}} \approx \frac{\text{细分行业增加值}}{\text{行业增加值}}$  (3)

投入产出表包含了分行业产出、中间投入、增加值等信息,是本文设定行业增加值率的重要依据。需要特别说明的是,部分行业产出或行业营业收入数据可以通过历年《中国工业统计年鉴》和《第三产业统计年鉴》获取。由于国家统计局每五年才开展一次经济普查,投入产出表也并非年年编制,最新数据仅截至2023年,因此需要对中间间断缺失年份的产出、增加值等数据采用年复合增长率平滑的方法予以补齐。考虑到该来源数据统计口径的一致性,将利用细分行业产出数据和式(2)估算该细分行业增加值。

对于电信业、软件和信息技术服务业所属的部分细分行业,只能从工信部发布

的《中国信息产业年鉴》中获取历年的营业收入数据。《中国信息产业年鉴》同时也公布了电信业、软件和信息技术服务业等(大类)行业的营业收入(产出)及增加值数据,但与国家统计局相关年鉴公布的数据有差异。考虑到行业数据统计口径的一致性,我们将基于信息产业年鉴提供的行业大类和细分行业相关数据,利用式(3)估算各细分行业增加值。

在数据产业各层级所涉及的各细分行业中,有的行业所对应经济活动可全部归为数据产业,如互联网数据服务等。但有的细分行业所对应的经济活动只有部分可以归为数据产业,例如,门类M“科学研究和技术服务业”的中类“知识产权服务”,只有其中的“知识产权信息服务和法律服务”会有部分业务涉及数据相关服务。因此,在完成对各细分行业增加值的估算后,还需对后者的增加值进行拆分,将属于数据产业的那部分增加值从行业增加值中予以剥离。为此,需要综合该行业经济活动与数据的关联程度,通过数据相关业务结构占比等方式设定细分行业的“数据业务调整系数”,据以估算细分行业中真正归属于数据产业的增加值,见式(4)。

$$\text{细分行业数据产业增加值} = \text{细分行业增加值} \times \text{数据业务调整系数} \quad (4)$$

## 五、中国数据产业增加值规模及结构分析

本部分将在多方收集所涉细分行业产值、增加值以及参数设定依据等关键信息的基础上,依据前文提出的测算思路与测算框架,对中国数据产业生态各层级增加值规模进行测算。

### (一)直接核心层增加值规模估算

根据前述梳理,直接核心层涉及细分行业中互联网数据服务、信息处理和存储支持业以及金融信息服务,其经济活动几乎都是围绕数据开展的,其增加值可全部计入数据产业增加值。从《中国经济普查年鉴》中可以获取互联网数据服务、信息处理和存储支持业这两个细分行业2018年和2023年的营业收入;采用年复合增长率平滑法补齐缺失年份的行业营收数据。国民经济行业分类中,“互联网数据服务”属于“互联网和相关服务”,而根据投入产出表又可以获取“互联网和相关服务”历年的产出和增加值。基于式(1),假定二者具有相同的行业增加值率,进而可测算出历年“互联网数据服务”的增加值规模。类似地,“信息处理和存储支持业”属于“软件和信息技术服务业”,基于投入产出表也能估算出软件和信息技术服务业的增加值率,据以推算信息处理和存储支持业的增加值。至于金融信息服务,华经产业研究院给出了2018年至2023年的市场规模数据。考虑到金融信息服务具有“即时生产即时消费”的特性,金融信息服务的供需应基本平衡,因而将金融信息服务业市场规模近似地作为其产值规模,假定金融信息服务业的增加值率等于整个金融业的增加值率,据此即可测算出金融信息服务业的增加值规模。

“地理遥感信息服务”“动漫、游戏数字内容服务及其他数字内容服务”“数字广告”这三个细分行业,其经济活动并非全部围绕数据展开,其增加值只有部分可以计入数据产业增加值,因此,在估算其行业增加值后还需要设定相应的数据业务调整系数。《中国信息产业年鉴》提供了软件和信息服务业的产出数据及2018年之前“地理遥感信息服务”“动漫、游戏数字内容服务及其他数字内容服务”的产出数据,2018年之后则将二者合并为“数字内容服务业”并提供了相应的行业产出数据。根据式(3),借助软件和信息信息技术服务业整体增加值,可算出这两个细分行业的增加值规模。通过对两个细分行业涵盖范围所涉及的业务活动与直接核心层的数据价值链环节进行分析,“地理遥感信息服务”的具体业务活动本质上都是由软件与数据共同构成的,初步将其数据业务调整系数设定为70%;相比之下,“动漫、游戏数字内容服务及其他数字内容服务”的业务活动涉及的文字、音视频固然是数据,但文化创意等成分也占据很大比重,结合各细分领域数据直接核心层环节渗透度,将其数据业务调整系数设定为50%。结合《中国数字内容产业市场格局与投资观察》报告中获取的历年数字内容产业各细分市场占比,<sup>①</sup>通过加权平均可以大致得到两个细分行业整体的数据业务调整系数。<sup>②</sup>据此,利用式(4)便可大致测算出“地理遥感信息服务”和“动漫、游戏数字内容服务及其他数字内容服务”两个细分行业合计的数字内容服务业数据产业增加值规模。

数字广告业的基础信息来源更为分散。通过国家市场监督管理总局发布的《中国广告业发展指数报告》可获取各年份的广告业营收数据。基于MAGNA机构历年发布的全球广告预测,<sup>③</sup>可以获取中国数字广告占据广告销售额的比例,从而可以推算出各年数字广告的营业收入。利用投入产出表获取租赁和商务服务业产出及增加值数据,从而得到租赁和商务服务业整体的增加值率,根据式(1)和式(2)得到数字广告的增加值规模。数字广告细分业务结构包括搜索广告、视频广告、传统展示广告、零售媒体广告、社交与信息流广告、移动广告等,基于对其业务结构的分析可知其业务内容基本上是由数据和展示方式共同支撑的,故而将其数据业务调整系数设定为70%。代入式(4)即可估算出数字广告数据产业增加值规模。

基于以上六个细分行业对应数据产业增加值的估算,将其加总便可以得到数据产业直接核心层的增加值规模。从表4可以看出,2023年,直接核心层数据产业增加值规模为7108亿元。其中,互联网数据服务、信息处理和存储支持服务、数字广告三个细分行业对应的数据产业增加值分别为1335亿元、1921亿元和2818亿元,占直接核心层数据产业增加值的比重分别为18.8%、27%和39.6%。

①张立、吴素平,2020:《中国数字内容产业市场格局与投资观察(2019~2020)》,社会科学文献出版社。

②数字内容服务业数据业务调整系数,详见《经济学动态》官方网站本文链接附录。

③Mediabrand.MAGNA Archives - Mediabrand. Retrieved from MAGNA Archives - Mediabrand.

表4 直接核心层数据产业增加值规模估算结果 单位:亿元(名义值)

|             | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|-------------|------|------|------|------|------|------|------|------|------|
| 互联网数据服务     | 320  | 431  | 599  | 831  | 973  | 1140 | 1335 | 1515 | 1696 |
| 信息处理和存储支持服务 | 156  | 239  | 364  | 557  | 841  | 1271 | 1921 | 2365 | 2845 |
| 数字内容服务      | 146  | 183  | 231  | 270  | 431  | 502  | 584  | 676  | 765  |
| 金融信息服务      | 174  | 192  | 237  | 292  | 337  | 389  | 449  | 494  | 543  |
| 数字广告        | 822  | 1003 | 1237 | 1575 | 2102 | 2285 | 2818 | 3189 | 3592 |
| 合计          | 1618 | 2049 | 2668 | 3526 | 4685 | 5588 | 7108 | 8239 | 9440 |

注:本表细分行业数据产业增加值实际估算年份为2017年到2023年,2024年以及2025年的结果数据则是在历史年份估算基础上采用霍尔特线性趋势法得到的预测值(表5、表6同)。

(二)间接支撑层增加值规模估算

间接支撑层所涉及的4个细分行业,即电信、信息系统集成服务、运行维护服务及知识产权服务,其经济活动并非全部围绕数据开展,其增加值只有部分可以计入数据产业增加值。各细分行业的增加值同样基于收集的行业产出信息以及其所属二位码行业的增加值或增加值率,利用式(2)或式(3)进行估算。并在式(4)的基础上,结合各细分行业数据业务调整系数的设定,估算各细分行业数据产业增加值,进而汇总得到数据产业间接支撑层的增加值规模。

电信行业一方面主要依托固定互联网宽带接入服务,并以基础传输网络为载体,为存储数据、数据处理等相关活动,提供互联网接入配套应用设施服务;另一方面,以数据中心、云计算、大数据、物联网等为代表的电信新兴业务构成了支撑数据产业运转的重要部分。由于固定互联网宽带接入服务可以划分为家庭宽带接入服务、企业专线接入服务和固定无线接入服务,而家庭宽带接入服务的用户规模依然占主体,固定互联网宽带接入服务占比大约1/4,因此,将电信行业固定互联网宽带接入服务的数据业务调整系数设为25%。而新兴业务依据其数据产业关联度,数据业务调整系数设定为1。采用《通信业统计公报》披露数据测算部分固定互联网宽带接入服务和新兴业务收入占电信业业务收入的比重,根据式(3)假定该比重等于其增加值占电信业增加值的比重,则可以通过电信业的增加值数据和相关数据业务调整系数,进一步得到电信业数据产业增加值数据。

信息系统集成服务业和运行维护服务业,同属于软件和信息技术服务业,采用《中国信息产业年鉴》数据,基于式(3)和采用行业营业收入和增加值数据估算二者行业增加值。在数据业务调整系数设定方面,信息系统集成服务可细分为基础设施集成服务、应用集成服务和数据集成服务三大块,与数据密切相关的是基础设施集成服务和数据基础服务。据Allied Market Research信息技术IT服务系统集成服务业研究,<sup>①</sup>2021年至2025年,中国数据中心基础设施占系统集成市场份额稳定在

<sup>①</sup> Yogendra, B. et al. (2022), “System Integration Market: Global Opportunity Analysis and Industry Forecast, 2022–2031”, Allied Market Research.

63%左右,业务结构具有稳定性,因此我们将信息系统集成服务业的数据业务调整系数设为 60%—65%。运行维护服务业涵盖范围所涉及业务活动中大约 1/3 与数据服务相关,据此将其数据业务调整系数设为 35%。

知识产权服务基于前述估算思路,采用国家知识产权局发布的历年《全国知识产权服务业统计调查报告》可以获取知识产权服务业历年营业收入,以及基于投入产出表获取科学研究和技术服务业产出及增加值等数据可以估算细分行业增加值。对于数据业务调整系数的设定,知识产权服务主要包括知识产权代理服务、知识产权法律服务、知识产权运营服务、知识产权信息服务和知识产权咨询服务五大类,其中,知识产权法律服务和知识产权信息服务都与数据业务有密切关联。结合二者业务涵盖范围的分析,知识产权法律服务业务中超过一半都与数据直接相关,将其数据业务调整系数设为 60%;而知识产权信息服务业务结构中开展信息检索服务、数据加工服务等数据相关业务的机构占比近 90%,因此我们将其知识产权信息服务的数据业务调整系数设为 90%。将知识产权法律服务、知识产权信息服务在知识产权服务中的营收比重作为权重,与二者的数据业务调整系数进行相乘求和,可以得到知识产权服务整体的数据业务调整系数。<sup>①</sup>

基于以上设定的各细分行业数据业务调整系数,结合营收数据及增加值率,利用式(4)便可得到各细分行业对应的数据产业增加值规模,以及间接支撑层数据产业增加值规模。从表 5 可以看出,2023 年间接支撑层数据产业增加值规模为 5143 亿元,其中,电信、信息系统集成服务两个细分行业对应的数据产业增加值分别为 2633 亿元和 1705 亿元,占间接支撑层数据产业增加值的比重分别为 51.2% 和 33.2%。

| 表 5 间接支撑层数据产业增加值规模估算结果 |      |      |      |      |      |      | 单位:亿元(名义值) |      |      |
|------------------------|------|------|------|------|------|------|------------|------|------|
|                        | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023       | 2024 | 2025 |
| 电信                     | 529  | 1204 | 1507 | 1860 | 2184 | 2463 | 2633       | 2923 | 3191 |
| 信息系统集成服务               | 1368 | 928  | 970  | 1035 | 1179 | 1262 | 1705       | 1964 | 2261 |
| 运行维护服务                 | 204  | 344  | 345  | 458  | 405  | 478  | 455        | 526  | 575  |
| 知识产权服务                 | 138  | 188  | 202  | 210  | 398  | 319  | 350        | 416  | 461  |
| 合计                     | 2240 | 2664 | 3024 | 3563 | 4166 | 4522 | 5143       | 5829 | 6487 |

(三)前端基础层增加值规模估算

前端基础层中,通信系统设备制造业所生产的设备最终都会通过资本形成转化为数字基础设施,即该细分行业几乎全部与数据有关,因此将其数据业务调整系数设定为 1。通信系统设备制造业属于通信设备制造业,借助《中国工业统计年鉴》可获取通信系统设备制造业营收作为通信系统设备制造业产出,其中,2017

<sup>①</sup>关于知识产权服务业数据业务调整系数,详见《经济学动态》官方网站本文链接附录。

年和2018年数据缺失,分别根据《中国信息产业年鉴》和《中国经济普查年鉴》予以补齐。基于细分行业与其所属中类行业增加值率基本相等的假设,结合通信设备制造业增加值和产出数据,利用式(1)和式(2)便可得到历年通信系统设备制造业增加值规模,其中,2021年至2023年的增加值规模则是在2020年估算基础上,基于2023年投入产出表通信系统设备增加值数据直接进行年复合增长率平滑而得。

计算机整机制造、计算机零部件制造、计算机外围设备制造、工业控制计算机及系统制造、信息安全设备制造五个(六位码)细分行业都属于计算机制造;借助《中国工业统计年鉴》可得到细分行业和大类行业的营收数据作为产出(2017年和2018年缺失数据同上处理),基于细分行业与大类行业增加值率基本相等的假设,结合计算机制造业增加值数据,利用式(1)和式(2)便可得到计算机制造业五个细分行业的增加值规模。基于投入产出表可以得到与数据基础设施相关的电信业、互联网和相关服务业中,计算机制造业作为中间投入的规模,该投入规模与计算机制造业当年中间使用比值,可大致作为计算机制造业及其细分行业的数据业务调整系数,代入式(4)便可得到这五个细分行业对应的数据产业增加值规模。基础软件开发、支撑软件开发、其他软件开发三个细分行业属于软件和信息技术服务业,可以通过《中国信息产业年鉴》获取细分行业相应产品业务收入数据作为产出,继而采用式(3)和上述步骤估算出各自对应的数据产业增加值。

“体育场馆建筑和其他房屋建筑业”“建筑安装”“建筑装饰、装修和其他建筑业”三个细分行业涉及数据业务的活动主要与算力中心等数据基础设施建设有关。因此,估算这三个细分行业对应的数据产业增加值,可以围绕数据基础设施建设业务展开。从历年的《数据中心产业白皮书》中,可以获取2017年以来历年的在用数据中心机架规模,以5KW单机架平均建设成本15万元为参考,则可推算出每年的数据中心建设成本。据Dgtl Infra(2023)研究相关资料披露的数据中心建设成本结构,其中,土地和建筑外壳成本约占15%—20%,基本对应于“体育场馆建筑和其他房屋建筑业”涉及数据业务的产出;建筑装修的成本约为20%—25%,基本对应于“建筑安装”和“建筑装饰、装修和其他建筑业”两个细分行业涉及数据业务的产出。<sup>①</sup>据此,可以测算出数据基础设施建设中“体育场馆建筑和其他房屋建筑业”“建筑安装”“建筑装饰、装修和其他建筑业”三个细分行业合计的涉及数据业务产出,利用对应行业中类的行业增加值率,便可得到其对应的数据产业增加值。从表6可以看出,2023年,前端基础层数据产业增加值规模为4386亿元,其中,通信系统设备制造和软件开发两个细分行业对应的数据产业增加值分别为1438亿元和1264亿元,占前端基础层整体比重分别为32.8%和28.8%。

<sup>①</sup>Dgtl Infra(2023),“How Much Does it Cost to Build a Data Center?”, Available from: <https://dgtlinfra.com/how-much-does-it-cost-to-build-a-data-center/>.

表 6 前端基础层数据产业增加值规模估算结果 单位:亿元(名义值)

|          | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|----------|------|------|------|------|------|------|------|------|------|
| 计算机制造    | 27   | 40   | 40   | 45   | 95   | 162  | 253  | 321  | 394  |
| 通信系统设备制造 | 1843 | 2029 | 2166 | 1015 | 1140 | 1281 | 1438 | 1616 | 1815 |
| 软件开发     | 142  | 250  | 252  | 276  | 398  | 661  | 1264 | 1753 | 2286 |
| 房屋建筑业    | 91   | 122  | 177  | 234  | 337  | 468  | 647  | 780  | 920  |
| 建筑装饰     | 135  | 182  | 262  | 344  | 464  | 604  | 784  | 926  | 1075 |
| 合计       | 2238 | 2622 | 2897 | 1913 | 2433 | 3175 | 4386 | 5395 | 6490 |

(四)中国数据产业规模及结构趋势

基于前述表 4、表 5、表 6,可以加总得到 2017—2025 年各年中国数据产业增加值总规模,并计算出整体名义增速、各层级增加值增速及其在数据产业增加值总规模中所占比重。从下表 7 可以看出,2017 年以来数据产业增加值总规模呈现快速增长态势,由 2017 年的 6096 亿元提高到 2023 年的 16637 亿元,预计 2025 年可达到 22418 亿元,年均名义增速高达 17.84%,远高于 2017—2023 年间 GDP 年均 7.37% 的名义增速。

从结构来看,直接核心层所贡献的数据产业增加值在产业生态整体中的平均占比约为 37%,间接支撑层和前端基础层占比分别约为 34% 和 28%。其中,直接核心层比重提升平稳,间接支撑层则有较为明显的下降,到 2021 年直接核心层比重超过间接支撑层,并稳步提升且保持数据产业增加值贡献结构中的主导优势地位,而前端基础层比重则先降后稳步提高。从发展动态来看,直接核心层平均名义增速为 24.85%,间接支撑层为 17.35%,前端基础层为 16.6%,数据产业发展形成以直接核心层为创新极、间接支撑层为支柱、前端基础层为底座的增长贡献结构。

以上结果表明,作为一种新兴产业,数据产业涉及的经济活动已经具备相当规模,仅直接核心层创造的增加值就已超过 9 千亿元,而整体规模已经达到 2 万亿元,且仍保持着超高速增长势头。据此估算,“十五五”时期,中国数据产业生态所创造的数据产业增加值规模有望翻一番,接近或达到 5 万亿元,成为新兴支柱产业。数据产业增加值规模的快速增长,不仅反映出经济运行中数据要素地位的日益提升,也将为“十五五”时期深化数据资源开发利用提供有力支撑。

表 7 2017—2025 年数据产业增加值增速及构成 单位:%、亿元(名义值)

|      | 数据产业整体增加值 |        | 直接核心层增加值 |        | 间接支撑层增加值 |        | 前端基础层增加值 |         |
|------|-----------|--------|----------|--------|----------|--------|----------|---------|
|      | 总量        | 增速     | 占比       | 增速     | 占比       | 增速     | 占比       | 增速      |
| 2017 | 6096      |        | 26.54%   |        | 36.75%   |        | 36.71%   |         |
| 2018 | 7335      | 20.31% | 27.93%   | 26.61% | 36.33%   | 17.87% | 35.74%   | 17.16%  |
| 2019 | 8589      | 17.10% | 31.06%   | 30.23% | 35.21%   | 13.60% | 33.73%   | 10.49%  |
| 2020 | 9002      | 4.80%  | 39.17%   | 32.14% | 39.58%   | 17.76% | 21.25%   | -33.96% |
| 2021 | 11285     | 25.36% | 41.52%   | 32.89% | 36.92%   | 32.08% | 21.56%   | 27.19%  |
| 2022 | 13285     | 17.72% | 42.06%   | 19.27% | 34.04%   | 14.89% | 23.90%   | 30.47%  |
| 2023 | 16637     | 25.23% | 42.72%   | 27.19% | 30.92%   | 17.30% | 26.36%   | 38.13%  |
| 2024 | 19463     | 16.99% | 42.33%   | 15.92% | 29.95%   | 14.06% | 27.72%   | 23.01%  |

续表 7

|      | 数据产业整体增加值 |        | 直接核心层增加值 |        | 间接支撑层增加值 |        | 前端基础层增加值 |        |
|------|-----------|--------|----------|--------|----------|--------|----------|--------|
|      | 总量        | 增速     | 占比       | 增速     | 占比       | 增速     | 占比       | 增速     |
| 2025 | 22418     | 15.18% | 42.11%   | 14.58% | 28.94%   | 12.33% | 28.95%   | 20.31% |
| 平均   | --        | 17.84% | 37.27%   | 24.85% | 34.29%   | 17.35% | 28.44%   | 16.60% |

六、总结性评论

本文前述各部分立足于数据资源开发利用,从数据价值实现视角入手,在数据价值实现内涵辨析和机制梳理基础上,厘清数据价值实现过程中涉及的环节、主体,确定数据产业的功能定位、范围边界、构成分类,剖析数据价值实现过程中数据产业各构成部分间形成的价值转移链条。进一步地,本文着眼于把握数据产业发展状况,提出数据产业增加值规模统计测算思路和方法框架,据以估算中国数据产业增加值规模并进行结构分析。基于前述分析,围绕数据价值实现、数据产业生态和中国数据产业发展状况形成以下主要判断。

第一,数据所携带的有效信息是其潜在价值的源泉,而只有作为生产投入参与生产经营和价值创造,数据潜在价值才能真正转化为现实价值,数据也由经济资源切实转变为生产要素。数据参与增加值创造是数据价值实现的终极标志,该过程也被称之为“数据要素化和价值化”。除此之外,诸如数据资产形成、数据资产质押贷款或直接融资,虽然也涉及货币价值,但并非数据价值实现。

第二,数据所具备的非物质性、即时性、非竞争性、非排他性、易复制性、非消耗性、非置换性、可增值性等技术-经济特征,使其能够通过三种机制实现价值创造:特定场景下的效率提升机制、数据融合产生的信息和价值增益机制、多场景复用带来的宏观价值倍增机制。这些机制也是发挥数据要素乘数效应的三种实现途径。

第三,数据产业是为数据价值实现提供基础性、专业化、支撑性服务或产品的经济活动部门,为数据资源开发利用提供支撑。所有与提炼数据有效信息有关的产业都可以纳入数据产业的范畴,具体可划分为:(1)为提取数据有效信息直接提供服务或产品的“直接核心层”,包括数据生成、收集、存储、清洗、标注、流通、加工、分析等。(2)不直接参与有效信息提取,但却为提取数据信息并最终实现数据价值提供保障支撑的“间接支撑层”,包括数据资产价值评估、数据安全合规审查以及通信、算力等服务。(3)与数据价值实现有关联,但又明显不能划归为数据产业的“前端基础层”,主要是前端数据生成处理能力形成、数据基础设施建设等涉及的数字产业。

第四,数据产业所涵盖的三个层级,在支撑数据价值实现过程中,不断进行价值的沉淀和后向转移,形成了一条价值转移链。其中,前端基础层数字产业以软硬件设备、数字基础设施等方式形成直接核心层和间接支撑层的固定资产;直接核心层和间接支撑层依托固定资产提供服务,并将部分价值转移至数据资产或数据产品服务中;数据资产/产品服务作为生产投入参与到生产经营中,将其自身所包含

的价值转移并体现在最终所实现的数据价值中。

第五,在现行国民账户体系(SNA)下,对照《数字经济及其核心产业统计分类(2021)》《国民经济行业分类(2017)》,确定数据产业生态三个层级涉及的细分行业,构建数据产业生态体系对应的数据产业增加值规模测算框架。据此对中国数据产业增加值规模进行测算,结果表明:(1)2017年以来中国数据产业增加值整体规模呈现快速增长势头,到2023年达到16637亿元,2025年预计达22418亿元,年均增速17.84%,远高于同期GDP增速。(2)“十五五”时期,中国数据产业增加值规模有望接近或达到5万亿元,成为重要的新兴支柱产业。(3)数据产业增加值规模的快速增长,既反映出数据要素地位的提升,也表明数据产业发展将为“十五五”时期深化数据资源开发利用提供有力支撑。

第六,数据产业发展已形成以直接核心层为主导、间接支撑层为支柱、前端基础层为底座的增长贡献结构。基于实际测算结果,前端基础层增加值增速稳步大幅提升,远高于直接核心层和间接支撑层的增加值增速,数据流通与价值释放仍有较大空间。因此,在推进数字基础设施建设扩容提速的同时,要加快构建以数据商、数据流通服务平台和数据交易所为主体的多元化数据流通服务机构体系,协同推进基础设施服务效能的转化,畅通数据价值流、充分释放数据要素乘数效应,促进数据产业高质量发展。

#### 参考文献:

- 蔡跃洲,2022:《经济循环中的循环数字化与数字循环化——信息、物质及资金等流转视角的分析》,《学术研究》第2期。
- 蔡跃洲,2024:《价值视角的数据要素资源分类与统计测算思路》,《数字经济发展与治理》第1期。
- 蔡跃洲、林靖玲、陈楠,2025:《数据并购视角的数据资源潜在价值与价值实现》,《中国工业经济》第12期。
- 蔡跃洲、刘悦欣,2022:《数据流动交易模式分类与规模估算初探》,《China Economist》第6期。
- 蔡跃洲、马文君,2021:《数据要素对高质量发展影响与数据流动制约》,《数量经济技术经济研究》第3期。
- 蔡跃洲、牛新星,2021:《中国数字经济增加值规模测算及结构分析》,《中国社会科学》第11期。
- 李海舰、赵丽,2021:《数据成为生产要素:特征、机制与价值形态演进》,《上海经济研究》第8期。
- 李娅、李卿芸,2025:《数据要素价值研究进展》,《经济学动态》第12期。
- 刘涛雄、戎珂、张亚迪,2023:《数据资本估算及对中国经济增长的贡献——基于数据价值链的视角》,《中国社会科学》第10期。
- 罗玫、李金璞、汤珂,2023:《企业数据资产化:会计确认与价值评估》,《清华大学学报(哲学社会科学版)》第5期。
- 戚聿东、王言,2026:《数据要素参与价值创造和分配的理论溯源》,《经济学动态》第2期。
- 谢康、夏正豪、肖静华,2020:《大数据成为现实生产要素的企业实现机制:产品创新视角》,《中国工业经济》第5期。
- 许宪春、雷泽坤、胡亚茹,2025:《中国数据资本形成总额与数据资本存量测算》,《经济研究》

第3期。

杨玫、李玮、乔思渊、刘巍,2022:《中国大数据产业产值测算方法研究》,《大数据》第3期。

翟丽丽、刘晓珊、杨彩霞,2023:《基于 Logistic-熵的大数据产业生态系统演化路径研究》,《科技进步与对策》第1期。

周曙东,2017:《我国大数据产业统计范畴的分析与探讨》,《调研世界》第12期。

ACIL Tasman (2008), “The Value of Spatial Information: The Impact of Modern Spatial Information Technologies on the Australian Economy”, A Report Prepared for the CRC for Spatial Information and ANZLIC, March.

ACIL Tasman (2009), “Spatial Information in the New Zealand Economy: Realising Productivity Gains”, Prepared for Land Information New Zealand, August.

Ackoff, R. L. (1989), “From Data to Wisdom”, *Journal of Applied Systems Analysis*, 16(1): 3-9.

Calderón, J. B. S. & D. G. Rassier (2022), “Valuing the US Data Economy Using Machine Learning and Online Job Postings”, US Department of Commerce, Bureau of Economic Analysis.

Farboodi, M. & L. Veldkamp (2021), “A Model of the Data Economy”, NBER Working Paper, No. w28427.

Faroukhi, A. Z. et al. (2020), “Big Data Monetization Throughout Big Data Value Chain: A Comprehensive Review”, *Journal of Big Data*, 7(1): 3.

IDC & the Lisbon Council (2022), “European Data Market Study 2021-2023 (D2.1 First Report on Facts and Figures)”, Luxembourg: Publications Office of the European Union.

Jones, C.I. & C. Tonetti (2020), “Nonrivalry and the Economics of Data”, *American Economic Review*, 110(9):2819-2858.

Kitchin, R. (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage.

Levitin, A. V. & T. C. Redman (1998), “Data as a Resource: Properties, Implications and Prescriptions”, *MIT Sloan Management Review*, 40(1): 89-101.

Manyika, J. et al. (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey & Company.

Martin, K. E. (2020), “Ethical Issues in the Big Data Industry”, in: R.D. Galliers et al.(eds.), *Strategic Information Management*, Routledge.

McKinsey & Company (2021), *Achieving Business Impact with Data: A Comprehensive Perspective on the Insights Value Chain*, McKinsey Analytics.

Rassier, D. G. et al. (2019), “Treatment of Data in National Accounts”, BEA Advisory Committee, Bureau of Economic Analysis.

Statistics Canada (2019), *The Value of Data in Canada: Experimental Estimates*, Ottawa: Statistics Canada.

Veldkamp, L. & C. Chung (2024), “Data and the Aggregate Economy”, *Journal of Economic Literature*, 62(2):458-484.

Vickery, G. (2011), *Review of Recent Studies on PSI Re-Use and Related Market Developments*, Information Economics, Paris.

Zikopoulos, P. et al. (2012), *Harness the Power of Big Data: The IBM Big Data Platform*, IBM Press.

## Positioning and Composition of the Data Industry from the Perspective of Data Value Realization and Its Scale Measurement in China

CAI Yuezhou<sup>a,b</sup> and HE Tiantian<sup>a</sup>

(a: University of Chinese Academy of Social Sciences, Beijing, China;

b: Chinese Academy of Social Sciences, Beijing, China)

**Summary:** Data is a critical factor underpinning new models and new business forms in the digital economy. Consequently, activities focusing on effective information extraction from data have become increasingly specialized and refined, gradually fostering emerging data-related industrial sectors and giving rise to segmented branches of the data industry. As a fundamental support for data value realization, the data industry emerges alongside the development of the digital economy. Scientifical and accurate measurement and systematic structural analysis of the data industry constitute the foundational support for understanding the trend of the digital economy and promoting the orderly development of the data industry.

Accordingly, this paper sorts out the theoretical mechanisms of data value realization and clarifies relevant conceptual definitions. Then, this paper identifies key stages, participants, and their functional positioning, decomposes the process of data value realization, defines the compositional classification of the data industry, and analyzes the value transfer chains among different sectors. Furthermore, this paper establishes a fundamental framework for statistical estimation of data industry scale, and adopts the production approach to estimate the added value scale of China's data industry.

Based on the linkage between micro mechanisms of data value realization and the macro statistical measurement system, this paper draws the following conclusions: (1) Valid information embedded in data serves as the source of its potential value. Potential value can be transformed into real economic value only when data serves as a production factor input involved in operational activities and value creation. This value creation can be achieved specifically through efficiency improvement, data fusion, and multi-scenario reuse. (2) The data industry is the economic sector that provides foundational, specialized, and supportive services or products for data value realization. Based on the functional positioning across the value chain, it can be categorized into three layers: the direct core layer, the indirect support layer, and the front-end foundational layer. (3) Digital industries related to the front-end foundation layer form fixed assets for the direct core layer and indirect support layer through software/hardware and infrastructure, transferring part of its value to data assets or data products during service provision, ultimately contributing to production and operations to complete data value realization, forming a value transfer chain. (4) China's data industry added value has maintained rapid growth since 2017, with an average annual growth rate of 17.84%, reaching CNY 1663.7 billion in 2023 and projected to reach CNY 2241.8 billion by 2025, which directly reflects the increasingly prominent role of data as a key factor. During the 15th Five-Year Plan period, the nominal added value of China's data industry is expected to exceed CNY 5 trillion, becoming a vital emerging pillar industry that provides strong support for advancing data resource development and utilization.

**Keywords:** Data Value Realization; Data Industry; Scale Measurement; Industrial Ecosystem; Value Transfer Chain

**JEL Classification:** C82, L86

(责任编辑:金 禾)

(校对:木 丰)