

第一部分 数据要素与数据权利

第1章 数据要素

互联网与移动智能终端等信息技术的发展与普及促使全球数据量以几何级数速度增长。根据国际数据公司（IDC，International Data Corporation）预测，2025 年全球数据量将接近 175 ZB，中国届时有望成为全球最大数据圈。数据是驱动数字经济软硬件发展的原料与动力，被公认为数字经济时代的“新石油”“新货币”（Keller et al., 2020）。在此背景下，党的十九届四中全会首次将“数据”增列为生产要素，《“十四五”数字经济发展规划》强调数据要素是数字经济深化发展的核心引擎，并在首份要素市场化配置文件中强调加快培育数据要素市场。

1.1 数据的概念与分类

1.1.1 数据的概念

对数据概念的剖析可从词源、内容和形式三个角度进行。从词源上看，“data”一词在拉丁语中可追溯至“to give”（给予）的拉丁过去分词“dare”（敢于），意为“被给予之物”。从形式上看，数据的记录与存储随着信息技术的发展而不断演进，呈现出虚拟化、电子化的趋势。早期的数据主要是通过口头传承或书写形式进行记录和传递，如天文历法、贸易记录等。18 世纪，统计学家开始利用数据进行社会和经济研究。彼时，数据的记录方式仍以人力为主，多以纸笔为媒介，主要发挥描述这一基本功能。

随着科学技术的不断进步,生产生活的信息化纵深发展,数据的定义也在生成与记录方式、功能与应用价值等方面发生了相应变化。如今,数据通常以电信号的形式被存储和传输,并记录在磁、光或机械介质上,计算机更是成为记录、存储、处理、分析数据的重要工具,因而加拿大、中国等国家均先后将数据定义为“任何以电子或者其他方式对信息的记录”。

从内容上看,数据被认为是信息的载体,是对事实的记录与反映。Ackoff(1989)指出,数据是未经加工的对事物、事件和活动等的记录,美国质量学会(ASQ, American Society for Quality)、《新牛津美语辞典》(NOAD, New Oxford American Dictionary)也认为数据是“一组事实”。国际数据管理协会(DAMA, Data Management International)则进一步明确了数据对于事实的呈现格式,认为数据是以文本、图片、声音、视频等格式对事实的表现。国际标准化组织(ISO, International Standard Organization)将数据定义为“以适合于通信、解释或处理的正规方式来表示的可重新解释的信息”。近年来,数字经济在全球经济的地位不断提升,社会生产生活的数字化程度持续提高,大数据、人工智能等数据分析技术逐渐成熟,数据中蕴含的信息、知识逐渐走进了学界和业界视野。许宪春等(2022)便指出数据是已记录为数字化形式的事实,且能够从中获取信息和知识。

1.1.2 与数据相关的概念

在此有必要说明数据、信息和知识之间的关系。^①根据“数据—信息—知识—智慧”(data-information-knowledge-wisdom, DIKW)模型(Ackoff,

① 叶继元等(2022)《数据与信息之间逻辑关系的探讨》指出,关于数据与信息关系的代表性观点主要有“数据大于信息说”“信息大于数据说”“数据与信息等同说”和“数据与信息相对说”。本文在此采纳的是在信息科学学科中获得多数研究认可的“数据大于信息说”。当然,不可否认的是,数据、信息和知识之间未必是简单的递进式逻辑关系,这些概念之间可能存在着包含与交叉,值得进一步深入研究与讨论。梅夏英(2020)《信息和数据概念区分的法律意义》在法学语境下分析了数据与信息概念区分的标准和意义,感兴趣的读者可进一步阅读。

1989)，从数据到信息，再从信息到知识是逐层抽象、逐步深入的过程，前者均为后者的基础，如表 1.1 所示。数据是描述某种事物或现象的数字、符号或文字等信息的集合，是信息的载体或表现形式。数据本身没有意义，只有在经过处理和分析后才能转化为有用的信息。信息是从数据中提取出来的有意义的描述或解释，是对数据进行解释和解读的结果。信息具有描述、传递和解释的作用，能够提供有用的指导和决策支持。知识则是对信息的理解和应用，是融合经验、技能甚至价值观等背景信息的具有一定框架的复杂综合体（Davenport，1998）。知识是在不断学习和实践中积累得到的，是对信息的深度理解和应用。

表 1.1 数据、信息和知识概念辨析（郭琰、王磊，2021）

辨析维度	数 据	信 息	知 识
记录事物	○	○	○
建立联系	×	○	○
提供解释	×	○	○
积累经验	×	×	○
反馈实践	×	×	○

除了信息与知识外，与数据相关的概念还有数字。时至今日，“数字”这一概念至少包括三层含义。最为熟知的“数字”是用来表示数的书写符号，如阿拉伯数字、罗马数字等。通过前文对数据概念的剖析，不难发现，此种“数字”是数据的一种表示方式。其次，“数字”还可指代“数字信号”，数字化便是指将模拟信号转换成数字信号的过程。在现代信息技术中，数据的数字化已经成为一种基本趋势。数字信号是以二进制形式表示的数值序列，可以在计算机、通信设备等数字系统中进行处理、传输和存储。在数字通信中，数据经过编码、调制等处理后，被转换成数字信号，通过光纤、电缆、无线电波等介质进行传输。在接收端，数字信号经过解调、解码等处理后，还原成原始的数据信息。因此，数据和数字信号是相

互关联的，数字信号是数据在数字系统中的表现形式。此外，“数字经济”一词也扩展了“数字”这一概念的外延。数字经济是指利用数字技术和信息化手段，推动经济和社会发展的一种经济形态，包括“数字产业化”和“产业数字化”两个主要方面。其中，“数字”指的是数字技术和信息化手段在经济中的应用和作用。在数字经济时代，各行各业的一切活动和行为都将数据化。数据既是数字经济运行过程的结果，也是驱动数字经济进一步高质量发展的原料。

1.2 数据的特征

目前，大多数研究认为虚拟性是数据与传统生产要素的最主要差异（Jones and Tonetti, 2020），即数据以非实体的形式存在（李海舰、赵丽，2021），而这也意味着数据需要与其他载体结合才能发挥其作为生产要素的作用，如与信息通信（ICT）产品结合（Jorgenson and Wu, 2016）。此外，与土地等传统生产要素不同，数据并不是一种自然禀赋，而是需要在信息提供者愿意的基础上，结合数据采集者的资本、劳动及特定技术，才能够获得（刘涛雄等，2023）。因而，数据是一种生成品。

从一般经济特性来看，数据具有可复制性、非竞争性和部分排他性等性质。业界证据表明，高价值数据的获得通常依赖于高效的采集方法和高性能的数据基础设施，这意味着需要高昂的前期投入（Demchenko et al., 2013; Redman, 2020; 梅宏等，2023）。然而，一旦数据形成，复制数据、额外使用数据的边际成本均几乎为零（Shapiro and Varian, 1998）。可复制性的存在使得数据在理论上可以被无限复制和分享，这使得数据具有非竞争性（Jones and Tonetti, 2020），即数据的使用不会对其他人造成损失或排斥，不同的用户可以共享同一份数据，而且可以同时使用和分析数据。Cong 等（2021）、Cong 等（2022）等研究认为数据的非竞争性可进一步细



化为水平非竞争性、垂直非竞争性和动态非竞争性。数据的水平非竞争性，是指在同一经济中，一家企业对数据的使用并不会影响其他企业对同一数据的同时使用。数据的垂直非竞争性，是指在经济不同部门中，数据在一个部门（如生产部门）的使用并不会限制其在另一个部门（如创新部门）中的使用。数据的动态非竞争性，是指随时间推移，数据可以被多个主体以极小的复制成本使用。尽管数据具有理论上的非竞争性，但是在现实中，多数私营组织更加倾向于自留数据，而非共享数据。一方面，以数据服务公司为代表的组织可以利用数据的排他性获得收入（Varian, 2018）。另一方面，以互联网平台企业为代表的不少组织也将其积累的数据视为核心竞争力，甚至是商业秘密。因而，数据在现实中也经常表现出可排他性的特点。数据的非竞争性和可排他性也使得数据成为一种准公共品。

从价值特性上看，数据具有结合性、异质性和外部性等性质。一般地，数据中的有效信息越多，数据带来的价值也将越大。Veldkamp 和 Chung (2019) 的研究表明，只有当数据与其他生产要素结合之时，数据才能表现出规模报酬递增的特点。除与其他生产要素结合之外，数据这一生产要素的内部结合、组合也有望获得价值提升。这使得数据价值具有自生性（Glazer, 1993）和协调性（Babaioff, 2012），且在一定数量范围内，能够实现规模经济。其次，数据价值通常是异质性的。数据价值的异质性既体现在同一数据对不同使用者的价值不同，也体现在同一使用者在不同场景使用同一数据的价值不同，即数据价值存在用户异质性（Bergemann et al., 2018）和场景异质性（Short and Todd, 2017）。这是由于数据的价值是由数据的使用价值决定的，数据的使用价值是否能够发挥以及能够发挥多大比例则取决于使用者的知识、技术和能力以及应用数据的具体场景。此外，数据在价值实现的过程中还可能存在外部性。这种外部性既可能是正面的，也可能是负面的，前者多见于宏观经济研究，而关于后者的讨论则集中于微观个体层面。Agrawal 等（2018）、Cong 等（2021）指出，数据有望通

过促进新知识生产进而提高全社会生产效率和经济增长速度。Cr mer 等 (2019) 指出由于数据可被复用, 在平台经济模型中, 对数据的访问容易转化为竞争优势, 进而形成“数据—网络—活动”(Data-Network-Activities) 循环的自我强化过程, 实现递增的规模收益。尽管 Jones 和 Tonetti (2020)、李勇坚 (2022) 指出数据特性和“干中学”的网络效应能够促进数据正反馈循环的实现, 但是数据正反馈循环对竞争等带来的负面影响可能会对数据规模收益递增的持续产生负面影响, 使其不可持续 (Yun, 2020)。Farboodi 和 Veldkamp (2021) 则指出由于数据反馈循环等因素的存在, 可能会形成数据积累的马太效应, 长期经济增长率未必会提高。Acquisti 等 (2016) 指出消费者通常会面对个人数据共享和隐私之间的权衡, 电子商务平台在价格歧视等方面对个人数据的使用除可能会降低个人效用外, 也有可能减少整体社会福利。平台企业在从事经济活动中往往凭借其优势地位过度收集个人数据或对隐私保护并不重视, 这些也都会产生负外部性 (Acemogle et al., 2022; Carriere-Swallow and Haksar, 2019)。

从交易特性上看, 由于权属复杂、信息悖论、市场环境等问题及现状的存在, 数据交易天然具有安全脆弱性, 市场主体缺乏进入数据要素市场的激励。数据作为一种非自然存在的生成品, 在收集、存储、处理、分析、管理及应用的过程中不可避免地会涉及多个主体, 产权关系复杂, 且一旦数据遭到破坏、篡改和泄露, 难以进行追责。其次, 数据交易还易受到买卖双方信息不对称的影响。Arrow (1972) 指出数据交易可能存在着信息悖论, 即买方只有在了解信息之时才能判断其价值, 而当他知道信息的具体内容之时, 他已经免费获得了该信息, 不再需要支付对价。信息悖论的存在意味着在数据交易过程中, 买卖双方均存在较强的机会主义倾向, 进而扰乱市场秩序, 使得数据要素市场停摆。此外, 当前以数据技术标准、数据权属法律、数据交易法规等为代表的制度体系尚不健全, 数据交易相关监管部门的缺失等因素也进一步加剧了数据交易的安全脆弱性。



1.3 数据作为新型生产要素

1.3.1 数据何以成为生产要素？

生产要素是指生产过程中所必需的各种资源，是组织生产、经济增长的重要基础。生产要素的形态、种类和重要程度会随着社会经济和科学技术的发展而不断变化，反映着人类社会不同发展阶段的生产力水平。早期，土地、劳动和资本是最为重要的生产要素。在农业经济时代，土地和劳动是最为重要的两类生产要素。然而，在人类从农业经济逐渐转向工业经济后，资本要素的重要性在生产过程中得到不断加强。进入 20 世纪，第三次工业革命带来了计算机等新技术的应用，知识、技术和管理等生产要素在经济增长中扮演着越来越重要的角色。如今，随着互联网、大数据、云计算和人工智能等技术的快速发展，数字化和信息化成为社会发展的大趋势，数据在社会生产中的地位与作用日益凸显，成为生产要素符合历史逻辑。

数据成为数字经济时代的核心生产要素不仅符合生产要素发展的历史逻辑，还具备理论科学性。佩蕾丝（2007）认为历次技术革命都会出现至少一种关键生产要素，数量丰富且获得成本低是成为关键生产要素的两个前提。显然，在技术发展、终端普及的今天，获得大规模数据十分便利，且数据的可复制性等一般经济特性也降低了数据的获得成本。同时，在摩尔定律的作用下，数据收集、存储、处理、分析和应用的成本存在进一步下降的空间（刘玉奇、王强，2019）。这些因素都解释了数据能够成为数字经济时代核心生产要素的原因。宋冬林等（2021）指出数据符合马克思主义生产要素理论的两个基本特征：（1）新的生产要素是从原有生产要素中派生的；（2）新的生产要素应当能够显著提高生产力。数据主要体现为数字化的信息和知识。信息主要包括人类信息和自然信息两大范畴，前者存在于人化自然，后者来源于自在自然，本质上都属于自然要素。知识是

人类依靠智力在实践过程中对信息加工后的产物，本质上源于自然和劳动。可见，数据作为信息和知识的数字化形式，从根本上派生于自然要素和劳动要素。此外，数据对于生产力和企业生产效率的提高作用也得到了实证分析的支持（张叶青等，2021；王宏伟、董康，2022）。

1.3.2 数据如何成为生产要素？

经济学中讲的生产要素是指在生产过程中起决定性作用的资源。数据要素是指在数字经济运行及其市场主体生产经营过程中能够创造经济或社会价值的数据资源^①。对于用数部门而言，数据要素如劳动、资本、技术等要素一般构成价值创造的必要投入。而原始数据往往无法直接使用，需要加工后才可用。并非所有的数据都可被视为“数据要素”。“要素”这一概念强调数据在生产中价值创造的能力。对于数据要素型企业而言，数据要素产品是企业的最终产物。因而，数据要素既是产成品又是投入品。从数据的属性分析，部分数据产品可直接作为最终消费产品，部分可作为生产要素继续投入生产，部分两种属性兼而有之，因此在讨论数据要素这一概念时，有必要关注其具备的“产成品”和“投入品”双重属性。

作为上游的产成品和下游的投入品，数据需要满足一定的可用性后才能够成为生产要素。原始数据如果不经收集 and 清洗，便无法投入生产创造价值。因此，作为生产要素的数据是对原始数据收集、存储、分析等处理后形成的数据资源，这其中多环节的加工处理过程，是数据价值链延展的过程，也是数据的要素化过程。

数据要素化涉及盘活数据存量和激活数据增量，同时还要求将原始数据以更加有序、更为高效的方式组织起来形成高质量数据，便于后续加工、处理和分析。从技术层面来看，数据要素化过程需要进行数据采集、标注、

^① 数据资源指企业可接触到的一切数据，往往具有实时性、碎片化特征。



汇聚和标准化等过程。数据资产化^①则是在数据要素化基础之上，将数据纳入企业的资产项以体现其业务贡献与真实价值，并实现科学管理的过程。事实上，数据已成为互联网公司、数字技术公司的“资产”。随着数字化转型持续推进，数据资产的存有比例将持续增加，数据的资产化已成为数字经济时代不可逆转的趋势。作为最活跃的市场参与者，企业是否意识到数据应作为资产存在、是否将数据纳入财务报表、是否科学高效地管理与评估数据，在微观层面上影响企业决策与管理能力，宏观层面上则关乎到我国数据要素市场的活跃度。数据资产在国家标准中的定义为“合法拥有或控制的，能进行计量的，为组织带来经济和社会价值的数据资源”。数据资产往往是在数据资源的基础上进行加工而形成的，其侧重于数据的可控制、可计量、可使用的特征（罗玫等，2023）。数据形成资产以后，在数据价值链的各环节均有增值的可能性，这对于数字经济时代以用户数据为主的互联网企业及以物联网大数据为主的数据要素型企业充分释放价值而言尤为重要。可以说，数据资产化是数据价值化过程中的“第一次飞跃”。从应用层面看，数据资产化过程要求企业依靠自有或外购数据，通过数字技术形成以数据要素为基础的产品和服务。数据资本化是指把数据作为一种金融资产进而进行的一系列金融操作的过程，如数据入股、增信、质押等。数据资本化是数据价值化的“第二次飞跃”。在数据实现资本化后，资本市场将高度认可数据要素价值，数据要素发挥其价值的途径也将更多元。

本章参考文献

- Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets[J]. *American Economic Journal: Microeconomics*, 14(4), 218-256.
- Ackoff, R. L. (1989). From data to wisdom[J]. *Journal of Applied Systems Analysis*, 16(1), 3-9.

① 本书所讲的数据资产是指企业层面、会计意义上的数据资产。

- Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy[J]. *Journal of Economic Literature*, 54(2), 442-492.
- Agrawal A, McHale J, Oettl A, (2018). Finding needles in haystacks: Artificial intelligence and recombinant growth[M]//*The economics of artificial intelligence: An agenda*. University of Chicago Press, 149-174.
- Arrow K. J., (1972). *Economic Welfare and the Allocation of Resources for Invention*[M]. Macmillan Education UK.
- Babaioff, M., Kleinberg, R., & Paes Leme, R. (2012, June). Optimal mechanisms for selling information. In Proceedings of the 13th ACM Conference on Electronic Commerce (pp. 92-109).
- Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019, May). The impact of big data on firm performance: An empirical investigation. In AEA papers and proceedings (Vol. 109, pp. 33-37). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.
- Berczi, A. (1981). Information as a Factor of Production[J]. *Business Economics*, 14-20.
- Bergemann, D., Bonatti, A., & Smolin, A.(2018). The Design and Price of Information[J]. *American Economic Review*, 108(1): 1-48.
- Carriere-Swallow M. Y., Haksar M. V., (2019). *The Economics and Implications of Data: an Integrated Perspective*[M]. International Monetary Fund.
- Cong, L. W., Wei, W., Xie, D., & Zhang, L. (2022). Endogenous growth under multiple uses of data[J]. *Journal of Economic Dynamics and Control*, 141, 104395.
- Cong, L. W., Xie, D., & Zhang, L. (2021). Knowledge accumulation, privacy, and growth in a data economy[J]. *Management Science*, 67(10), 6480-6492.
- Crémer, J., De Montjoye, Y. A., & Schweitzer, H. (2019). Competition policy for the digital era. Report for the European Commission.
- Davenport T. H., & Prusak L. (1998). *Working Knowledge: How Organizations Manage What They Know*[M]. Harvard Business Press.
- Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013, May). Addressing big data issues in scientific data infrastructure. In 2013 International conference on collaboration technologies and systems (CTS) (pp. 48-55). IEEE.
- Drucker, P. (1992). The New Society of Organizations[J]. *Harvard Business Review*, 20, 281-293.
- Farboodi, M., & Veldkamp, L. (2021). A model of the data economy (No. w28427). National Bureau of Economic Research.
- Glazer, R. (1993). Measuring the value of information: The information-intensive organization[J]. *IBM Systems Journal*, 32(1), 99-110.



- Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data[J]. *American Economic Review*, 110(9), 2819-2858.
- Jorgenson, D. W., & Vu, K. M. (2016). The ICT revolution, world economic growth, and policy issues[J]. *Telecommunications Policy*, 40(5), 383-397.
- Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020). Doing data science: A framework and case study[J]. *Harvard Data Science Review*, 2(1).
- Redman, T. C. (2001). Data Quality: The Field Guide[M]. Digital Press.
- Shapiro, C., & Varian, H. R. (1998). Versioning: the smart way to[J]. *Harvard business review*, 107(6), 107.
- Short, J., & Todd, S. (2017). What's your data worth?[J]. *MIT Sloan Management Review*, 58(3), 17.
- Varian H., (2018). Artificial intelligence, economics, and industrial organization[M]//*The Economics of Artificial Intelligence: an Agenda*. University of Chicago Press: 399-419.
- Veldkamp, L., & Chung, C. (2019). Data and the aggregate economy[J]. *Journal of Economic Literature*, forth coming. <https://www.aea-web.org/article?id=10.1257/jel.20221580&&from=f>.
- Yun, J. M. (2020). The Role of Big Data in Antitrust. The Global Antitrust Institute Report on the Digital Economy, 7.
- 郭琰、王磊. 2021. 科学认识数据要素的技术经济特征及市场属性 [J]. 中国物价, 第 5 期.
- 何伟. 2020. 激发数据要素价值的机制、问题和对策 [J]. 信息通信技术与政策, 第 6 期.
- 胡锴、熊焰、梁玲玲、邵志清、汤奇峰. 2023. 数据知识产权交易市场的理论源起、概念内涵与设计借鉴 [J]. 电子政务, 第 7 期.
- 黄丽华、郭梦珂、邵志清、秦璇、汤奇峰. 2022. 关于构建全国统一的数据资产登记体系的思考 [J]. 中国科学院院刊, 第 10 期.
- 卡罗塔·佩蕾丝. 2007. 技术革命与金融资本: 泡沫与黄金时代的动力学 [M]. 北京: 中国人民大学出版社.
- 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 [J]. 上海经济研究, 第 8 期.
- 李勇坚. 2022. 数据要素的经济学含义及相关政策建议 [J]. 江西社会科学, 第 3 期.
- 刘玉奇、王强. 2019. 数字化视角下的数据生产要素与资源配置重构研究——新零售与数字化转型 [J]. 商业经济研究, 第 16 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 [J]. 清华大学学报 (哲学社会科学版), 第 5 期.
- 梅宏、杜小勇、金海、程学旗、柴云鹏、石宣化、靳小龙、王亚沙、刘驰. 2023. 大数据技术前瞻 [J]. 大数据, 第 1 期.



- 梅宏 . 2022. 数据如何要素化：资源化、资产化、资本化 [J]. 施工企业管理, 第 12 期 .
- 屈阳 . 2023. 数据要素的特征与定价方法研究 [D]. 北京：清华大学 .
- 宋冬林、孙尚斌、范欣 . 2021. 数据成为现代生产要素的政治经济学分析 [J]. 经济学家, 第 7 期 .
- 王宏伟、董康 . 2022. 数据要素对企业发展的影响——基于云计算行业 197 家上市公司实证分析 [J]. 东岳论丛, 第 3 期 .
- 许宪春、胡亚茹、张美慧 . 2022. 数字经济增长测算与数据生产要素统计核算问题研究 [J]. 中国科学院院刊, 第 10 期 .
- 张叶青、陆瑶、李乐芸 . 2021. 大数据应用对中国企业市场价值的影响——来自中国上市公司年报文本分析的证据 [J]. 经济研究, 第 12 期 .