



信息资源管理学报

Journal of Information Resources Management

ISSN 2095-2171, CN 42-1812/G2

《信息资源管理学报》网络首发论文

题目： 数据生产:概念、场景、技术与审思할

作者： 胡广伟, 范兆媛

网络首发日期： 2024-08-12

引用格式： 胡广伟, 范兆媛. 数据生产:概念、场景、技术与审思할[J/OL]. 信息资源管理学报. <https://link.cnki.net/urlid/42.1812.G2.20240812.0913.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

数据生产：概念、场景、技术与审思

胡广伟¹ 范兆媛^{1,2}

(1. 南京大学信息管理学院, 南京, 210023; 2. 南京邮电大学社会与人口学院, 南京, 210023)

[摘要] 数字化转型为经济社会发展提供了重大机遇, 同时也面临诸多挑战, 如数据从哪里来、如何持续供给、如何培育数据核心能力及数据生产场景与技术等问题亟待探讨。通过对数据生产的概念、结构、特征、场景与技术的讨论, 希望引起理论界、实践界对数据生产新业态的关注, 促进数字化、智能化、智慧化等新质生产力的发展, 为我国数字化转型与治理能力现代化服务。

[关键词] 数据生产 数据产品 数据业态 生产场景 生产技术 新质生产力

[中图分类号] G201 [文献标识码] A

Data Production: Concepts, Scenarios, Technologies and Reflections

Hu Guangwei¹ Fan Zhaoyuan^{1,2}

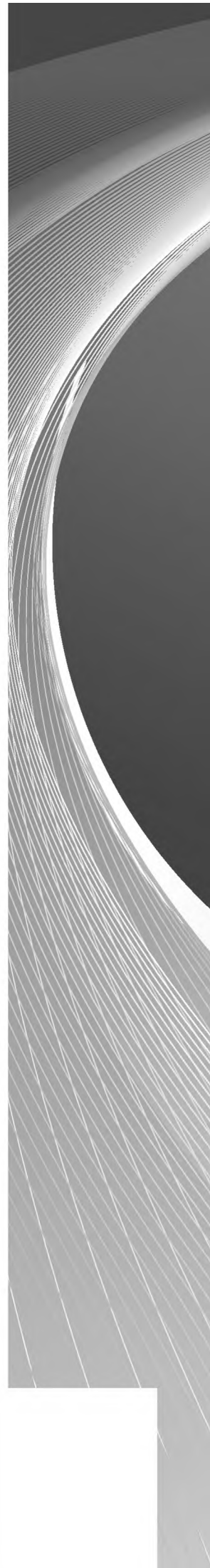
(1. School of Information Management, Nanjing University, Nanjing, 210023; 2. School of Sociology and Population Studies, Nanjing University of Posts and Telecommunications, Nanjing, 210023)

[Abstract] Digital transformation offers significant opportunities for the development of the economy and society while also presenting numerous challenges. Issues such as the source of data, continuous supply, cultivation of core data capabilities, and the urgent need to explore data production scenarios and technologies await discussion. By discussing the concept, structure, characteristics, scenarios, and technologies of data production, we hope to draw the attention of both the theoretical and practical sectors to the new business forms of data production, promote the development of new productive forces such as digitalization, intelligentization, and wisdomization, and serve the modernization of China's digital transformation and governance capabilities.

[Keywords] Data production; Data products; Data formats; Production scenarios; Production technologies; New quality productive forces

[基金项目] 本文为国家社会科学基金重大项目“大数据驱动的城乡社区服务体系精准化构建研究”(20 & ZD154), 国家社会科学基金青年项目“收缩型城市的形成机制与优化路径研究”(23CRK002)的阶段性成果。(This paper is a phase result of the major project, "Research on Precision Construction of Urban and Rural Community Service Systems Driven by Big Data"(20 & ZD154) supported by the National Social Science Fund, and the young project "Research on the Formation Mechanism and Optimization Path of Shrinking Cities"(23CRK002) supported by the National Social Science Fund.)

[作者简介] 胡广伟, 博士, 教授, 研究方向为电子政务、政务大数据、数据智能等; 范兆媛(通讯作者), 博士, 博士后, 南京邮电大学副教授, 研究方向为大数据与城市治理, Email: fzy@njupt.edu.cn. (Hu Guangwei, Ph.D., professor, with research interests in E-government, governance big-data and data intelligence. Fan Zhaoyuan (corresponding author), Ph.D., postdoctoral, associate professor, with research interests in big data and urban governance, Email: fzy@njupt.edu.cn.)



1 引言

随着数字化的发展，“数据”已成为日益重要的经济资源和生产资料，数据的生产和开放利用、数据相关技术及产业创新成为全球经济发展的重要驱动力^[1]。数据不仅本身能创造价值，而且还是剩余价值的重要来源^[2]。社会各领域数据的采集记录、加工处理，实现了原始数据的生产与要素价值的提升，已成为日益重要的经济资源和生产资料，是数字政府、数字经济与数字社会建设的关键生产要素。不同行业、场景、类型的数据蕴藏着丰富的价值，但目前还较少有对数据生产概念、内涵等的系统研究。著者在2017年编著《数据思维》一书的过程中初次思考“数据生产”的问题^[3]，主要缘于数据生命周期的各个环节始终无法绕过“数据产品”的概念。时至今日，“数据生产”常常被实业界所提及，其形式、业态也日益兴盛。因此，希望通过讨论数据生产的概念、场景、技术等，引起理论界、实践界对数据生产新业态的关注，促进数字化、智能化、智慧化等新质生产力的发展，为我国数字化转型与治理能力现代化服务。

2 数据生产的概念、结构与特征

2.1 数据生产的概念

生产是人类创造社会财富的活动和过程。描述特定对象的数据不是天然存在的，而是被生产出来的^[4]。传统数据生产通常通过人工收集、录入、验证、存储和维护，这些数据通常是结构化的，例如表格和数据库中的数据。然而，随着信息技术的发展，特别是集深度学习、人智融合之大成的人工智能生成内容（AIGC）、大语言模型（LLM）等智能化技术的发展，许多传统数据可以实现自动化生产，从而提高效率、减少错误，并允许处理更大规模和更复杂类型的数据，数据生产逐渐融入社会化大生产的洪流。

数据生产在数据生命周期中发挥着越来越重要的作用，例如数据生成、数据收集和存储、数据处理和清洗等工作。数据生成是数据生命周期的开始，可以通过用户输入、设备传感器记录、业务交易等方式生成。数

据采集是数据被收集并存储在某种形式的存储设备中，例如数据库、数据仓库等。数据处理与清洗是将数据清理、转换和整理，以便于后续分析或者深度加工生产的过程。可见，在数据生命周期中，数据生成、采集加工等都是数据生产的形式。数据从多个源头收集、处理、整理和存储，以供进一步的分析和解释，这一过程包含人类劳动与价值创造。因此，数据生产可以归纳为创造新的数据或者以原始数据为基础加工成为新的数据或数据产品的过程，是人类开发利用数据资源来创造社会财富的活动与过程。

2.2 数据生产的层次结构

数据生产是随着电算化、网络化、数字化的发展不断形成的新业态，计算机、（移动）互联网、物联网（数字孪生）、云计算、大数据等为其提供了技术保障，形成了包括原始数据积累到一次、二次乃至n次数据生产的“菱形层次结构”，见图1。最初的初次数据生产是数据的源头，有不同的类型，按参与主体由内而外的不同组合可分为采集生产（自己独立完成）、共创生产（由内外双方共同完成）、众包生产（由大家共同完成）、外包生产（委托外部主体完成），这些过程共同形成初次数据。中间的加工生产是主体，在初次数据的基础上，依据数据流通自由度的不同，通过数据交换、数据共享、数据交易、数据开放形成不同的数据组合生产模式，如聚集、聚合、融合形成更为完善的数据结构，或在清洗标注、分类聚类后用于选择判断、经营管理决策、科学研究等。经过三次或n次生产，数据被高度知识化、智能化与智慧化，形成产业竞争力、新质生产力、国家竞争优势、国家软实力等高层次数据要素。总体而言，数据生产层次由低至高形成数据链，数据生产的参与主体、生产形式、产品形态构成数据产业链与产业生态，在加工生产过程中不断创新产品、服务与商业模式，产生溢出价值，形成数据的创新链。

2.2.1 原始数据生产

按照数据生产过程中参与主体、参与方式、开放程度的不同，原始数据的生产可分

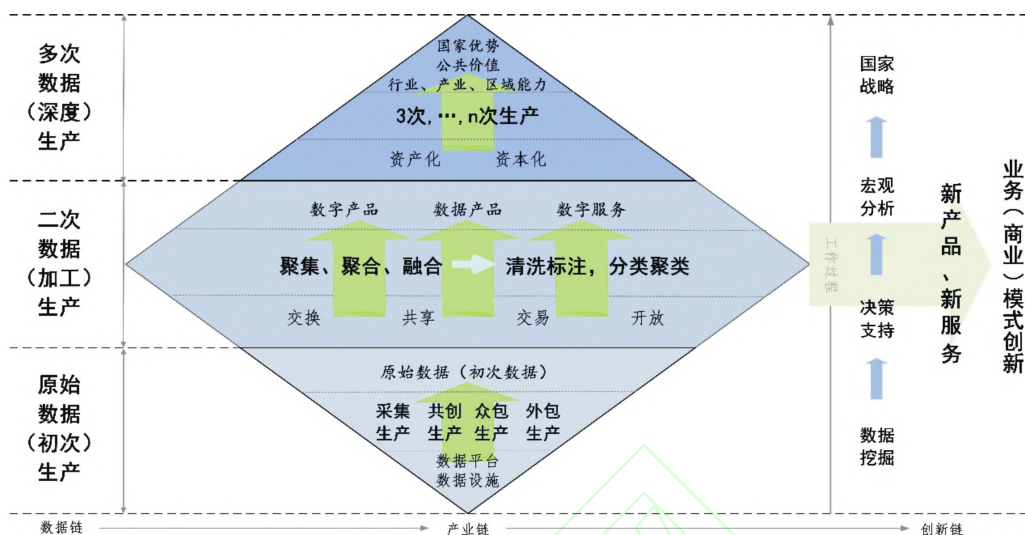


图 1 数据生产的层次结构

Fig. 1 Hierarchical Structure of Data Production

为采集生产数据、共创生产数据、众包生产数据与外包生产数据。

采集生产数据本质上是从源头上生产数据，特别是通过更新数据采集指标体系。数据生产者采集多元、异构、实时、全维度的数据记录，以满足数字化转型或者数据产业化的需求。当下，应特别重视通过更新数据采集指标体系，将统计时代的数据采集理念转变为数据时代、智能时代的数据采集理念，增强数据采集的实时性、全数据、客观性以及 4V 特征（规模大、类型多、速度快、价值高），建立数据平台体系，从根本上革新数据生产的起点。

共创生产数据本质上是内、外部参与主体共同创造产品与服务价值的过程中累积的数据，数据是共创活动的溢出价值。如小米 MIUI（社区）、乐高 LEGO IDEAS 平台、海尔 HOPE（开放创新）平台、华为花粉俱乐部、Kickstarter 众筹平台及社会治理“随手拍”等，让用户的更多想法与创意融入产品迭代与创新过程中，平台运营方与用户共同参与，在提问、回答、评论、激励中生成数据，在发现问题、提出想法、方案设计与价值扩散过程中共同创造使用价值或社会价值，解决复杂问题，创造创新产品、服务和商业模式，同时为平台贡献数据资源。

众包生产数据是一种广泛存在的数据生产

模式，本质上是通过网络平台集合用户资源，利用众多参与者的智慧和力量来共同完成数据生产任务。与其他的生产模式相比，众包生产更加灵活、开放，能够迅速聚集大量的创意和解决方案，从而在社交网络与用户生成内容双重效应中生成数据资源。如知乎、小红书、Facebook、微信、Wikipedia 等，通过动员全球范围内的用户参与使用、贡献内容，这些内容满足用户所需，为平台贡献数据资源，并为平台用户进一步进行数据生产提供原材料。

外包生产数据是通过委托协议由外部主体收集数据的形式。如大量专业的数据采集机构（公司）从事的业务，委托方付费，受托方采用专业的技术、设备采集委托方所需数据。

2.2.2 二次生产数据

二次生产数据是在原始数据的基础上，依据数据流通自由度的不同，通过数据交换、数据共享、数据交易、数据开放形成不同的数据。其中，数据交换与共享是指在系统、应用或组织之间交换与共享数据的过程。如银行、投资公司和监管机构交换或者共享交易数据和风险数据，以进行风险管理和合规审查；政府部门交换或者共享人口、税收和公共服务数据，以提供公共服务和制定政策。数据交易是指公司或机构之间买卖数据的过程，在这个过程中，数据被视为有价值的商

品,可以被收集、处理、打包并出售给数据需求方。如广告商和市场研究公司购买个人的浏览历史、购物习惯、社交媒体活动等在线行为数据。开放数据是指可以被任何人自由使用、再分发和共享的数据,开放数据在很多领域都有应用,例如政府通过开放数据平台,提供各种公共数据用于支撑业务处理、科学研究、社会治理和公共服务等,如人口统计数据、环境数据和交通数据等。开放数据还被用于高德、百度等电子地图应用,支撑诸如“口口安全”“智慧出行服务”“便民菜市场”“城市积水点”等城市动态服务应用的运行。

数据通过交换、共享、交易和开放等途径流通后,通过聚集、聚合和融合等生产方式处理能塑造更加完善的数据结构。这些数据经过清洗标注、分类和聚类处理,转化为数据产品、数字服务和数字产品。具体来说,数据聚集包括从多种数据源(如数据交换、共享、交易和开放)中搜集信息;数据聚合则是将众多数据点或集合融为一体,并通过求和、平均、计数、最大值、最小值、中位数和分位数等方法提取关键特性,简化数据复杂性,优化分析及决策过程;数据融合通过清理、转换和整合不同来源数据,形成统一且完善的数据视图,消除冗余、提升质量,增强一致性和可信度。完成上述过程后,还需进行数据清洗(修正或删除数据集中的错误、不一致性、重复或不完整数据)、降噪(使用滤波、平滑、回归等减少噪声)、分类(将数据归入预定义类别)和聚类(探索数据中的结构或模式)等,以进一步提升数据质量和可用性。基于以上处理得到的数据产品、数字服务或数字产品可用于支撑或辅助研判、决策,或用于AI产品的设计,大大提升决策效率、生产效率和智能化水平。

可见,加工生产数据本质上是将数据资源或者原始数据进行加工生产,变成可用的数据或数据产品的过程。其中,有的数据直接通过挖掘、分析、可视化成为数据成品,比如“知识图谱”“事理图谱”“领导看板”“领导驾驶舱”等;有的经过二次生产成为数据产品,用于后续进一步的挖掘、分析。此外,

数据不仅可以进行生产加工,还可以进行分类包装、质量控制、合规性审查等,以满足不同目的的数据产品需求,这就形成了新的数据产品业态。

2.2.3 多次数据生产

经过深度加工生产的数据产品包括行业咨询报告、产业研究报告、区域竞争力分析报告、各种指数(如股市指数、消费者信心指数、行业发展指数等),以及更为宏观的领导看板、领导驾驶舱、国家经济发展态势分析等形式。其中有些产品可以直接支持决策分析,如行业报告等。然而,还有一些加工后的数据可能仅是半成品,需要进一步的深度加工才能变成完整的数据产品。例如,通过对淘宝行业产品投诉问题进行分产业、分行业、分区域的深入分析,实现数据驱动的多维度管理与治理。再如,通过对“315”消费者权益日的曝光数据进行深入分析用于公共决策,实现了数据的社会价值。在此过程中,数据产品经历了一个数据再加工和使用的循环过程,数据不断地被生成、加工、应用,并再次生成新的数据或数据产品。通过多次数据生产过程,形成新质生产力、国家数据优势、国家竞争优势等,为我国数字化转型与治理能力现代化服务。

2.3 数据生产的特征

与工农业生产不同,数据生产的时间、空间、场景等更加丰富多元,呈现时空、多模态融合与泛在化特征。

2.3.1 数据生产的时空特征

数字时代的数据呈现出实时性、移动性和大空间的特点。随着新兴信息技术的发展,形成了“人机共线”的互联空间,在线用户与联网设备随时随地都在生产数据。以移动设备和传感器产生的流文件数据为代表的实时数据快速生成,从智能手机、智能电器到电网、路网,再到医院、工厂,快速生成的移动数据与实时数据的泛在性、重要性与日俱增、引人注目。数据已成为政府、企业和消费者工作与生活正常运转的关键要素,对数据实时分析、即时可用、融合处理的需求与日俱增。据国际数据公司(IDC)发布的《数据

时代 2025》显示，移动数据和实时数据在未来几年将呈现出强劲增长的势头。数据生产的时空特征在数据生产周期中起着重要作用，为数据生态的构建提供了基础。这些特征对设计高效的数据管理策略、优化数据处理流程、确保数据安全和隐私保护等方面至关重要。

2.3.2 数据生产的多模态融合特征

数字时代产生了大量描述不同时空、不同模态、不同个体行为的互联数据，促进进行互联网、工业互联网、数字孪生生态等新兴领域的发展，为理解数据行为、多重体验与开展业务的随处运营提供了资源。借助于各类时空数据来分析与研究人类行为的时间、空间特征，形成解释社会治理与经济现象的时空分布、相互作用、动态演化过程的理论和方法。比如基于出租车行驶记录、签到打卡等数据获取海量移动轨迹，研究人们在地理空间中的活动和移动范围；基于手机数据获取用户之间的通话联系信息，了解个体之间的社交关系。由于多模态时空大数据包含了海量人群的时间空间行为信息，为研究群体的行为特征，揭示空间要素的分布格局提供了新的机会。

2.3.3 数据生产的泛在化特征

数据的产生、加工与使用呈现出多场景、泛在化的特点。如社交媒体与网络平台依据用户生成的内容与互动数据分析消费者行为、预测趋势并进行个性化推送；电子商务平台依据其浏览记录、购物偏好、下单模式动态进行商品推送，帮助商家了解消费者偏好、优化库存管理，提升用户体验感受；智能设备和物联网（IoT）依据收集到的智能家居、穿戴设备、工业传感器等实时数据，用于监控设备性能，实现远程控制，优化能源管理和支持决策制定。以淘宝评论为例，同样的评论数据在不同场景对不同用户发挥着不同的作用，商品交易完成后用户留下对产品的评论，一方面可以为其他用户提供购买决策参考，另一方面可以为淘宝平台提供服务优他参考，即平台可以通过评论数据分析用户的购物体验、产品体验、预期产品与服务，助力商家改进平台服务。

3 数据生产的场景与技术

随着数字化转型的演进，数据生产场景越来越多样化，数据资源结构也日益丰富。不仅可以采取大量田野调查、实验记录等人工观察和记录的方式来生成数据，还可以借助计算机系统、互联网与物联网设施等来记录、感知、抓取活动对象或观察对象的行为或变化，形成结构丰富的多模态数据。如政府、企事业单位的数据库和数据仓库中的业务数据，社会网络（微信、小红书、抖音等）形成的互联网数据，电商平台服务器日志和各种传感器记录收集的行为数据，二维码与条形码扫描形成的物联网数据等。虽然不同领域都在生产数据，但在信息技术支撑下，互联网、物联网仍是数据生产的主要场所。

3.1 互联网数据生产场景与应用

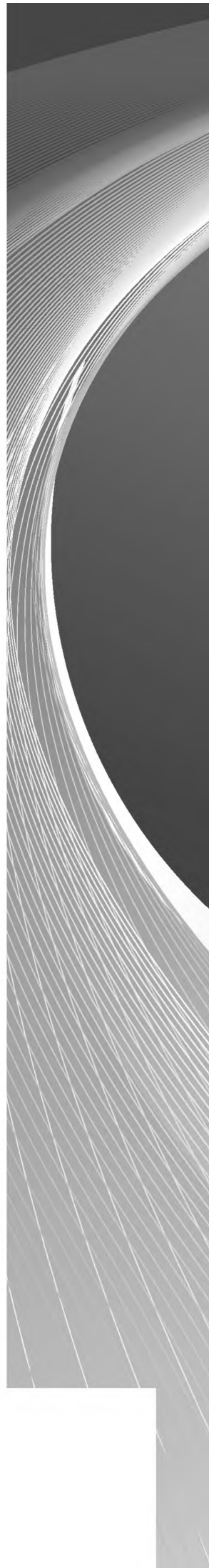
3.1.1 互联网数据

20 世纪 90 年代，随着计算机和数据库系统的广泛使用，许多企业的管理系统存储了大量的数据，用来满足各种业务需求。在这个阶段，数据是被动产生的^[4]。随着互联网的出现，不需要借助于磁盘、磁带等物理存储介质就可以实现数据的快捷传播，网页的出现又加速了网络内容的产生，使得数据量呈现“指数级”增长。然而，互联网数据的真正爆发始于以“用户生成内容”为特征的 Web 2.0 时代。Web 2.0 技术强调自服务，以 Wiki、博客、微博、微信等自服务模式为主。在这个阶段，大量上网用户本身是内容的生成者，因此数据是自动生成的。随着移动互联网和智能手机终端的普及，用户可以随时随地使用手机发微博、传照片，互联网数据量开始急剧增加。

中国互联网络信息中心发布的第 53 次《中国互联网络发展状况统计报告》显示，截至 2023 年 12 月，中国网民规模已达到 10.92 亿，互联网普及率达到 77.5%。规模庞大的网民，在网络上留下多种多样的“使用痕迹”，生产出大量的互联网数据。

3.1.2 互联网数据生产技能与应用

互联网数据生产的技能主要有系统日志生产技术、网络数据生产技术、其他生产技能



(如针对保密单位设置特定的端口)以及人工智能技术。①系统日志是指系统中记录的硬件与软件等信息以及监视的事件,包括系统、安全与应用程序日志。用户可以利用系统日志检查错误原因,寻找系统受到攻击的痕迹。大数据平台可以产生很多高价值系统日志,比如 Hadoop 平台开发的 Chukwa 就是一种典型的系统日志采集方法。目前这种采集技术可以达到数百兆每秒的传输速度,满足数据生产的需求。②网络数据生产技术包括网络数据采集的两种方法:API 和网络爬虫。API 又叫应用程序接口,是网站的管理者为了方便使用者,编写的一种程序接口。目前新浪微博、百度贴吧以及 Facebook 等主流社交媒体平台均提供 API 服务,但是 API 技术受限于平台开发者,为了减小平台的负荷,一般平台会对每天接口调用上限进行限制。因此,网络爬虫技术是更常用的数据获取方法。网络爬虫是按照设定程序自动抓取数据的方法,最常见的爬虫包括搜索引擎,如百度、360 搜索等。③其他生产技能主要针对一些保密单位,以保证数据的安全传递。这些保密单位通常会采用系统特定端口传递数据,降低数据泄露的风险。④人工智能技术,如 AIGC 和 LLM 等技术依赖于互联网的数据传输、存储和计算能力来生成和分发内容,使得基于这些技术的应用(如聊天机器人、内容推荐系统等)能够与用户进行交互。

互联网数据的应用极其多样,服务于商业、科研和社会需求。其应用主要体现在以下四个方面:①个人化推荐:通过利用用户行为数据,互联网平台可以提供个性化的内容推荐,包括购物推荐、音乐或视频推荐等,以提高用户体验和满意度;②社交网络分析:通过分析社交媒体上的互动数据,企业可以了解公众情绪、流行趋势,并据此制定营销策略;③健康监测与预测:利用穿戴设备收集的健康数据,可以监测个体健康状况,预测疾病风险,并提供定制化健康建议;④智能交通系统:通过分析来自传感器和摄像头的交通流量数据,智能交通系统可以优化信号灯控制、减少交通拥堵、提高安全性。

3.2 物联网数据生产场景与应用

3.2.1 物联网数据

随着物联网的发展,人类社会数据量发生了第三次跃升。物联网不仅包含视频监控摄像头,也包含大量传感器,如位移、湿度、压力、温度、光电传感器等。物联网中的这些设备时刻都会自动生成大量数据,相比于人工生产数据方式,物联网通过自动生成数据的方式可以在短时间内根据所有者需要生成合规、全量、密集的数据,从而推向人类社会迅速步入“大数据时代”。随着数字化转型,数据生产发生了翻天覆地的变化,数据形式变得越来越多样化,数据量也呈现爆炸式增长。在此背景下,可以借助计算机系统、网络系统、传感设备等来记录、感知、抓取活动对象或描述对象的行为或变化,形成可用的数据^[3]。如服务器日志、各种传感器收集的数据,二维码与条形码扫描形成的物联网数据。

根据物联网中数据采集和传输的过程,物联网通过感知层、网络层和应用层来获得数据,具体如图 2 所示。感知层由传感器及其网关技术架构而成,包括各种传感和感知终端,其主要作用是识别物体和采集信息,并获取物联网数据。网络层主要是传递和处理感知层所获取的各种信息,具体包含各种私有网络、互联网、有线和无线通信网、网络管理系统和云计算平台等。应用层是物联网和用户(包括人、组织和其他系统)的接口,它与行业需求结合,实现物联网的智能应用。

3.2.2 物联网生产技能与应用

物联网是物与物、人与物之间的信息传递与控制。在物联网数据生产中存在以下关键技术:①传感器技术。由于计算机处理的主要是数字信号,因此需要利用传感器将模拟信号转换成计算机可以处理的数字信号。②RFID 技术。该技术广泛应用在自动识别与物流管理等方面,也是一种传感器技术,将无线射频与嵌入式技术融成一体。③嵌入式系统技术。将计算机软硬件、电子应用、传感器与集成电路技术融为一体,比如 MP3、航天卫星系统。④纳米技术。该技术针对的对象主

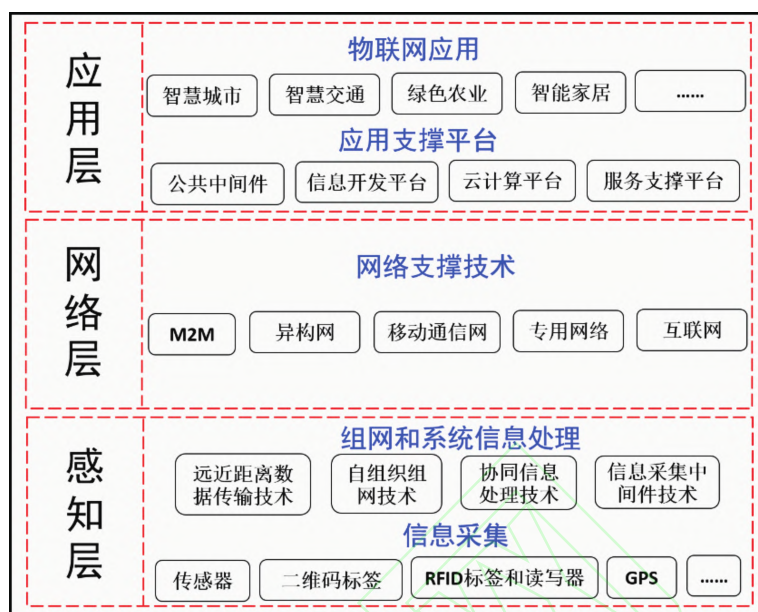


图 2 物联网数据

Fig.2 Internet of Things Data

要是结构尺寸在 0.1nm 到 100nm 的材料，在纳米物理学、材料学、电子学、化学、生物学、加工学、力学等学科得到了应用。⑤人工智能技术。集深度学习、人智融合之大成的 AIGC、LLM 等技术正拓展内容生产、图片生产、视频生产的先河，AIGC 和 LLM 可以通过智能分析和处理设备收集的数据来优化设备性能和服务，以提供智能决策支持、自动化服务和增强用户体验。

物联网设备无时无刻都在生产新的数据。物联网平台可有效提升产业与经济的信息化能力，增强数据分析水平，并高效灵活地利用资源，带动多层次应用发展，其主要应用体现在以下三个方面：①物联网信息化应用。基于物联网平台的“软件上云+简单数据分析”运用在供应链管理与计划资源管理等领域可有效降低企业成本，如 ERP 供应商 SAP、甲骨文、微软、浪潮、金蝶等企业为中小企业提供管理软件 SaaS 服务，降低了中小企业的信息化门槛与风险^[5]。②物联网大数据优化应用。利用物联网平台的大数据优势，采用“模型+深度数据分析”的运行模式，物联网大数据在能耗管理、产品售后、质量管控等多产业场景中得到了大量应用，并取得了显著的经济效益。如美国 Palo Verde

核电站采用 Uptake 提供的数据平台服务，提升资产的性能，每年节省成本 1000 万^[5]。③物联网协同创新应用。部分企业借助物联网平台整合不同产业链资源，如中船集团通过船海智云平台监控船舶与船用设备制造企业，提升企业协同水平^[5]。

4 结论与审思

随着数字化的发展，数据生产业态日益丰富，已成为原始数据生产、加工生产、多次生产等不同层次、场景、技术及其价值空间快速拓展的重要领域，成为推动经济增长、促进社会进步和创新的关键力量。但数据生产仍有诸多问题值得审思：

(1) 原始数据的权属。数据生产仍存在诸多问题，如众包生产的权属问题、共创生产的激励问题、外包生产的质量控制问题等。所有权的模糊导致数据开发利用中的所有权、持有权、收益权、分配权、转让权难以界定，导致数据产业化处于进退维谷的窘境，甚至产生了“数据保守主义”。共创生产数据中，用户协议界定了部分权责关系，但在平台数据增值价值的开发、收益、分配、转让等问题上仍存在争议，例如如何激励各参与主体的共创能效等问题在数据产业化的过程中构

成了进一步发展的障碍，需要行业、政策制定者和参与者共同探讨解决方案。

(2) 数据流通的窘境。随着数据蕴藏的价值为企业和政府熟悉，数据开放与共享取得了一定的成效，但到目前为止，数据资源流通渠道仍未完全打通，“数据孤岛”问题在一定程度上仍然存在。主要表现在：其一，数据跨行业流动仍未真正实现。在各数据主体利益驱使下，部门间和部门内部的数据没有实现真正的互流。其二，数据交易市场的兴起在一定程度上加剧了“数据堰塞湖”“数据孤岛”的形成。以数据销售为盈利模式的新兴企业，在利益的驱使下，必然会提高其收集数据的保密程度，甚至演化为“数据保守主义”。其三，企业对接速度慢、数据更新速

度快，使“数据孤岛”问题更为突出。

(3) 数据智能的伦理。AIGC 与 LLM 等大数据模型技术在应用过程中也面临诸多挑战，比如质量控制、伦理考虑以及保护知识创造中的人类能动性等问题亟待解决。此外，数据霸权、算法公平、数据隐私、信息茧房和技术依赖等问题风险也不容忽视。在训练和使用 AIGC 和 LLM 等模型时，处理大量的个人和敏感数据可能会引发隐私和安全问题。因此，需谨慎验证生成内容的准确性和可靠性，并探索如何避免潜在的负面影响。

以上为著者对“数据生产”业态的观察、思考与整理，诸多内容可能偏颇，希望通过广泛的讨论助推数字化的发展。

参考文献

- [1] 李政, 周希祺. 数据作为生产要素参与分配的政治经济学分析[J]. 学习与探索, 2020(1): 109-115. (Li Z, Zhou X Z. The political economy analysis of the participation of datas as production factors[J]. Study & Exploration, 2020(1): 109-115.)
- [2] 马费成, 熊思玥, 孙玉姣, 等. 数据分类分级确权对数据要素价值实现的影响[J]. 信息资源管理学报, 2024, 14(1): 4-12. (Ma F C, Xiong S Y, Sun Y J, et al. Impact of data classified and graded rights confirmation on the realization of the value of data elements[J]. Journal of Information Resources Management, 2024, 14(1): 4-12.)
- [3] 胡广伟. 数据思维[M]. 北京: 清华大学出版社, 2020. (Hu G W. Thinking in data[M]. Beijing: Tsinghua University Press, 2020.)
- [4] 高富平. 数据生产理论——数据资源权利配置的基础理论[J]. 交大法学, 2019, 10(4): 5-19. (Gao F P. The theory of data production: A fundamental theory of the right allocation of data resource[J]. SJTU Law Review, 2019, 10(4): 5-19.)
- [5] 中国工业互联网研究院. 中国工业互联网产业经济发展白皮书(2021)[R]. 北京: 中国工业互联网研究院, 2021. (China Academy of Industrial Internet. White Paper on the economic development of China's industrial internet industry(2021)[R]. Beijing: China Academy of Industrial Internet, 2021.)