

NATIONAL DATA
RESOURCE SURVEY REPORT

全国数据资源调查报告(2023年)

全国数据资源调查工作组
(国家工业信息安全发展研究中心)

版权声明

本报告版权属于全国数据资源调查工作组（国家工业信息安全发展研究中心），并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：《全国数据资源调查报告（2023年）》”。违反上述声明者，将追究其相关法律责任。

专家委员会

(排名不分先后)

顾问

江小涓 邬贺铨 刘韵洁 梅宏 尹浩 高新民

委员

蒋艳 张向宏 李晓东 姜奇平 王建民 何小龙 廖凯 吴志刚
汤珂 冯海红 黄罡 蔡跃洲 李爱君 欧阳日辉 汤奇峰 王晨
赵刚 傅建平 黄丽华

编写组

组长

陶炜

成员

夏宜君 黄洁 郭晁煜 李清敏 徐福龙 芦钊禾 辛伟涛

前言

习近平总书记指出，要构建以数据为关键要素的数字经济。党的十九届四中全会首次增列“数据”作为生产要素，中国成为首个将数据作为生产要素的国家。

随着《关于构建数据基础制度更好发挥数据要素作用的意见》的发布和国家数据局组建，《“数据要素×”三年行动计划（2024-2026年）》等10余项政策文件陆续印发，我国数据要素化进入新阶段。数据感知采集出现新技术、数据基础设施产生标准化新需求、数据交易机构打开新局面、公共数据启动授权运营、数据开发利用涌现新模式、数据服务商形成新业态、数字化转型大规模铺开，数据生产存储、流通交易、开发利用的各个环节产生新变化，亟需构建一套新的调查方法和指标体系，来全面摸底我国数据资源发展情况。

2024年2月，国家数据局联合中央网信办、工业和信息化部、公安部组织开展全国数据资源调查，国家工业信息安全发展研究中心承担调查实施工作。在前期工作基础上，此次调查扩大调查范围、扩展调查对象、增设调查指标，通过全量测算和抽样调查等方式，调查全国关键数据资源情况，为政府施策、产业发展提供参考。

调查发现，我国数据资源“产-存-算”规模优势基本形成，数据“供给-流通-应用”主体逐渐丰富，海量数据和丰富场景优势潜力亟需释放，数据资源管理和利用整体处于起步阶段。

一是数据生产总量大，但有效供给不足。2023年，数据生产总量¹达32.85泽字节（ZB），同比增长22.44%。5G、AI、物联网技术的创新发展及智能设备的规模应用，推动数据生产规模快速增长。2023年，全国数据存储总量²为1.73泽字节（ZB），新增数据存储量³为0.95泽字节（ZB），生产总量中2.9%的数据被保存。存储数据中，一年未使用的数据占比约4成，数据加工能力不足导致大量数据价值被低估、难以挖掘复用。

¹ 全国数据生产总量指2023年全国数据创建、复制、捕获和消费产生的数据总量。

² 全国数据存储总量指全国累计存储的数据量，包括算力中心存储量，以及政府、企事业单位、个人终端的数据存储量。

³ 当年新增数据存储量包含当年新增存储设备的存储量，以及既有设备中被当年数据覆盖的数据量。

二是算力存力较合理，还需适度超前布局。2023年，全国2200多个算力中心的算力规模约为0.23十万亿亿次浮点运算/秒（ZFLOPS），同比增长约为30%；全国数据存储总空间⁴为2.93泽字节（ZB），存储空间利用率⁵为59%。随着大模型研发应用不断增加，对存力、算力提出更高要求，需保持适度超前布局。

三是数据流通交易需求旺盛，多元流通模式待完善。2023年，全国数据总流量同比增长7.6%，消费领域数据交互活跃度较高。中央企业⁶和平台企业⁷发挥行业枢纽作用，探索数据交互机制。数据交易机构建设加速，场内交易活跃度较低，产品成交率为17.9%。数据供给难以满足旺盛需求，急需建立和完善多元流通模式。

四是数据应用场景加速落地，数据价值有待释放。公共数据成为引领数据开发利用的催化剂，公共数据开放量同比增长超16%，授权运营初步探索。数据多场景应用、多主体复用难度大，96%的行业重点企业⁸已实现数据场景化应用，但实现数据复用增值的大企业⁹仅占8.3%，数据价值有待释放。

全国数据资源调查工作验证了我国发展数字经济具备超大规模市场、海量数据资源和丰富应用场景优势。未来，随着数据技术不断进步，基础制度持续完善，数据生产规模将保持高速增长，存储计算将向按需调度发展，数据流通将向规范有序、多元协同发展，数据开发利用将向需求牵引、智能驱动发展。

报告研究时间较短，不足之处在所难免，敬请各位专家和广大读者指正。

⁴ 数据存储空间包括全国算力中心的存储空间，以及政府、企事业单位、个人终端的存储空间。

⁵ 存储空间利用率=累计数据存储总量/累计数据存储总空间。

⁶ 本报告提到的中央企业，有效样本数量为94家。

⁷ 本报告提到的平台企业，有效样本数量为354家。

⁸ 本报告提到的行业重点企业为东部地区百强企业，中部地区80强企业和西部地区50强企业，有效样本企业数量1202家，不含94家中央企业和768家数据服务商。

⁹ 按照《统计上大中小微型企业划分办法（2017）》（国统字〔2017〕213号）对行业重点企业中的大中小型企业进行划分，报告中的大企业为统计上大型企业。

目录

CONTENTS

一、调查结果	1
(一) 一组数据: 数据“产-存-算”规模优势基本形成	2
(二) 一批成效: 数据“供给-流通-应用”主体逐渐丰富	2
(三) 一些问题: 海量数据和丰富场景优势潜力仍待释放	3
二、工作方法	4
(一) 数据来源	5
(二) 调查方法	7
(三) 结果校验	7
三、结果分析	8
(一) 数据生产规模大, 范围广, 增长速度快	9
(二) 数据存储空间合理, 终端存储高于云存储, 产存转化待提升	10

(三) 算力规模增长快, 区域算力按需布局, 智算能力需求旺	11
(四) 数据流通方式多元, 交互能力有提升, 交易供需不均衡	11
(五) 公共数据开放共享初见成效, 授权运营起步探索, 政企数据融合不断深入 ..	14
(六) 数字化转型激活企业数据, 场景应用提质增效, 数据外溢效应仍不足 ..	16
四、发展趋势	19
(一) 数据生产向高增速和高质量同步发展	20
(二) 数据存储计算向一体化按需供给发展	20
(三) 数据流通向规范有序、多元协同发展	20
(四) 数据应用向需求牵引、智能驱动发展	21
附录	22

数据作为新型生产要素，具有重要基础性和战略性，是发展新质生产力的重要基础。近年来，在政产学研用合力推动下，我国数据规模实现快速增长，数据流通交易架构逐步形成，数据开发利用创新探索，数据要素价值加快释放，为做大做强数字经济、建设数字中国奠定了坚实基础。

一、调查结果

调查发现，我国数据资源管理和利用整体处于起步阶段，数据“产-存-算”规模优势基本形成，数据“供给-流通-应用”主体逐渐丰富，海量数据和丰富场景优势潜力仍待释放。

（一）一组数据：数据“产-存-算”规模优势基本形成

全国数据生产、存储、计算呈现规模大、增速快的特点。

2023年，全国数据生产总量达到32.85泽字节（ZB），同比增长22.44%。

2023年，全国数据存储总量为1.73泽字节（ZB），存储空间利用率为59%。

2023年，2200多个算力中心的算力规模超0.23十万亿亿次浮点运算/秒（ZFLOPS），同比增长约30%。其中，智能算力占比约30%。

（二）一批成效：数据“供给-流通-应用”主体逐渐丰富

数据要素市场化进程中，各类主体不断涌现，为数据“供给-流通-应用”提供有力支撑。

2023年，全国一体化政务数据共享枢纽接入53个国家部门、31个省/自治区/直辖市和新疆生产建设兵团数据，“一网通办”事项不断增加。

2023年，公共数据开放量同比增长超16%。18.6%的平台企业和51%的中央企业在数据开发利用过程中应用到政府开放数据。

2023年，4个国家部门、15个省级数据管理部门探索公共数据授权运营。

（三）一些问题：海量数据和丰富场景优势潜力仍待释放

数据生产存储、流通交易、开发利用过程中的数据量降幅较大，能支撑人工智能训练的数据资源较少，数据价值仍需进一步挖掘。

2023年，全国数据产存转化率¹⁰为2.9%，海量数据源头即弃。

2023年，企业一年未使用的数据占比为38.93%，大量数据被存储后不再被读取和复用。

2023年，交易所需求方是供给方的1.75倍，数据产品成交率为17.9%，数据场内交易供需匹配率低。

2023年，尚未建设数据管理系统的大企业比例为21.9%，实现规划管理的数据资源仍然较少。

2023年，开展数字化转型的大企业中，实现数据复用增值的仅有8.3%，数据价值挖掘任重道远。

¹⁰ 产存转化率=当年新增存储数据量/当年生产数据量。

二、工作方法

近年来，数据资源的基础作用日渐凸显，正在激发数据产业创新活力，不断赋能经济社会高质量发展。2024年2月，为适应数据要素化市场配置改革形势，更好分析我国数据资源结构、利用情况和存在问题，国家数据局、中央网信办、工业和信息化部、公安部四部门发布《关于开展全国数据资源调查的通知》，面向政府部门、高校、协会以及重点企事业单位开展调查。国家工业信息安全发展研究中心在原有工作基础上进行了优化调整，借鉴国际相关机构调查方法，设立了“指标体系-采集机制-分析测算-结果论证”调查方法体系，以期更加充分反映我国数据资源全貌。

(一) 数据来源

本次全国数据资源调查以人、机、物广泛连接产生的电子化数据为调查对象，涵盖数据生产存储、流通交易、开发利用和安全管理全生命周期。在调查项目上，强化对数据资源活跃度、应用场景、投入产出等方面的调查分析。

本次全国数据资源调查标准时点为2023年12月31日，调查结果中的全国数据不包含香港、澳门和台湾地区数据。

表2.1 2023全国数据资源调查有效样本说明

序号	名称	样本说明
1	数据生产	根据主要数据生产/采集设备及系统软件等厂商和协会数据，测算全国数据生产总量。
2	数据存储	根据主要数据存储介质厂商和协会数据，测算全国数据存储总量。
3	数据计算	来自2200多个算力中心的数据。
4	数据流通	来自四大运营商、27家数据交易所的数据。
5	数据应用	来自全国29个国家部门、31个省/自治区/直辖市、5个计划单列市、33家重点高校、14个国家科学数据中心、27家数据交易所、10个国家实验室和2064家样本企业(含94家中央企业和354家平台企业)的数据。

数据生产和**数据存储**两项指标综合国外专业机构研究方法、国内相关专家论证，经过全国22个行业协会推荐，设定数据生产存储样本库。**数据生产**主要来源各类主体在生产生活、交通运输、物联感知、生产制造、数据处理领域，创建、复制、捕获和消费产生的数据总量。**数据存储**主要来源各类存储介质的数据记录和保存总量，以及各类存储介质的存储容量。

表2.2 数据产生和存储样本情况

序号	数据产生和存储分类		
1	人	生产生活	包含手机、PC、智能家居、可穿戴设备等
2		交通运输	包含火车、飞机、船、智能汽车等
3	物	物联感知	包含监控设备、仪器仪表、传感器等
4		生产制造	包含边缘智能设备、工业机器人、数控机床等
5	机	存储计算	包含服务器、内存、硬盘等
6		软件系统	包含资源管理系统、研发设计软件等

数据计算主要来源于2200多个算力中心数据资源调查信息，涵盖通用计算中心、智能计算中心和超算中心。

数据流通主要来源于四大运营商和27个数据交易所数据，覆盖数据交互和交易情况。

数据应用主要来源于中央企业、数据服务商和行业重点企业共计2064家有效样本企业，以及重点高校、国家实验室、国家科学数据中心等数据。

表2.3 数据应用样本情况

序号	分类		
1	政务数据	国家部门	
2		省级数据主管部门	
3		计划单列市数据主管部门	
4	科研数据	高校	
5		国家实验室	
6		国家科学数据中心	
7	企业数据	中央企业	
8		数据服务商	平台企业
9			其他企业
10		行业重点企业	农业
11			工业
12			金融
13			交通物流
14			消费民生
15			科教文卫
16			公共服务

(二) 调查方法

本次全国数据资源调查方法主要采用**全量数据测算**和**抽样数据调查**两种方式。

数据生产和**数据存储**数据结果采用全量测算模型，应用龙头企业和行业协会数据设置模型参数，测算结果反映总体情况。

数据计算、**数据流通**、**数据应用**3项指标采用抽样调查方式。对样本单位数据进行梳理、计算后得出调查结果。

(三) 结果校验

针对数据结果开展三次校验。一是数据之间校验，在数据各环节主体信息开展交叉验证，如协会与设备厂商、云服务商与算力中心等。二是数据结果与国内外研究机构数据校验，如国际数据公司、中国社会科学院等相关报告结果进行交叉验证。三是数据结果专家论证，开展八轮专家论证会，邀请20多位数据领域专家论证数据可靠性。

三、结果分析

我国数据产量、存量、算力规模大，数据流通、交易变化快，公共数据开放共享授权运营加快探索，企业数据多场景落地，为构建数据要素市场提供了坚实基础。

(一) 数据生产规模大, 范围广, 增长速度快

我国数据生产规模持续扩大。2023年数据生产总量达到32.85泽字节 (ZB), 同比增长22.44%, 非结构数据爆发式增长。随着5G、AI技术的快速发展及智能设备的规模应用, 内容创作、影像视听等非结构数据对我国数据总规模增长贡献较大, 消费民生领域新业态不断涌现, 满足了人们不断增长的文化娱乐消费需求。

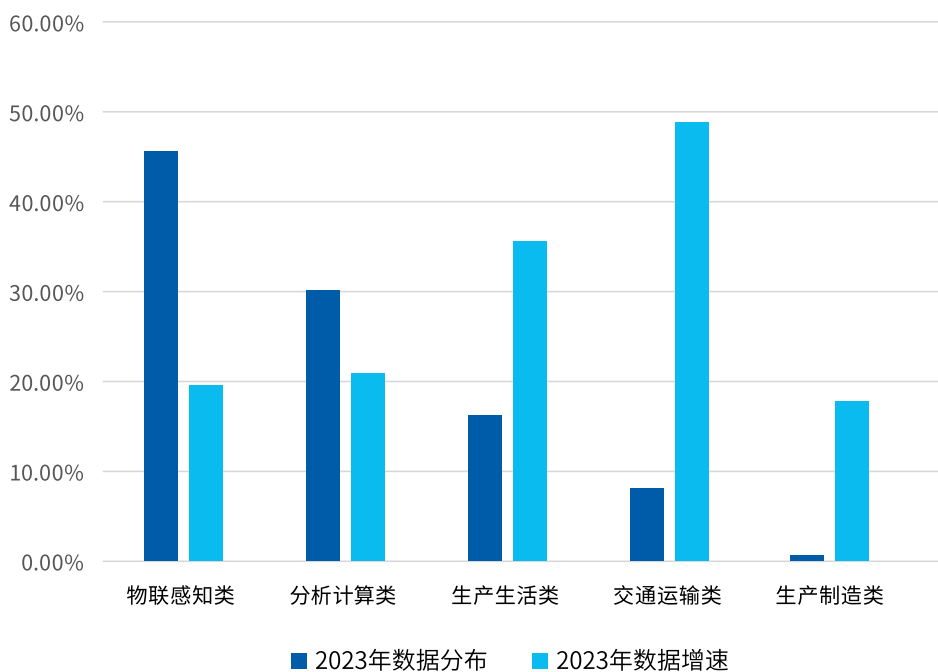


图3.1 数据来源分布情况

数据来源: 2023年全国数据资源调查

智慧城市扩大物联感知范围。公共数据获取范围不断扩大, 设施物联传感数据占比超40%。城市各种基础设施、资源、服务等加速数字化、网络化、智能化升级, 摄像头、传感器、无线设备等相互连接形成城市的“神经网络”, 实时捕捉城市交通、人流等各种信息, 推动海量数据从大街小巷汇入城市大脑。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/235103144310011213>