

CAICT 中国信通院

数据价值化与数据要素市场 发展报告 (2021 年)

中国信息通信研究院政策与经济研究所
2021 年 5 月

版权声明

本报告版权属于中国信息通信研究院，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：中国信息通信研究院”。违反上述声明者，本院将追究其相关法律责任。

前 言

习近平总书记指出，要“发挥数据的基础资源作用和创新引擎作用”，党的十九届四中全会首次明确数据可作为生产要素按贡献参与分配，《关于新时代加快完善社会主义市场经济体制的意见》首次将数据与技术、人才、土地、资本等要素一起纳入改革范畴，《关于构建更加完善的要素市场化配置体制机制的意见》、十九届五中全会等历次重要会议、文件都将数据要素作为重要内容，为加快数据要素市场发展提供了根本遵循、确定了目标、指明了方向。

历史经验表明，每一次经济形态的重大变革，必然催生也必须依赖新的生产要素。如同农业经济时代以劳动力和土地、工业经济时代以资本和技术为新的生产要素一样，数字经济时代，数据成为新的关键生产要素。由网络所承载的数据、由数据所萃取的信息、由信息所升华的知识，正在成为企业经营决策的新驱动、商品服务贸易的新内容、社会全面治理的新手段，带来了新的价值增值。加快推进数据价值化、发展数据要素市场是数字经济的关键。

数据价值化与数据要素市场发展报告建立了数据价值化的“三化”框架，即数据资源化、数据资产化、数据资本化。全球数据的井喷式生产为数据资源化奠定了基础，本报告重点分析数据资源化中的产业发展规律，其中，数据采集、数据标注有望成为撬动产业规模发展的新引擎。数据资产化是数据通过市场流通交易给使用者或所有者带来经济利益的过程，目前全球对数据权属的确定、数据

资产的定价均有一定的探索，数据交易初具规模。以金融衍生产品理论为研究支撑，重点探讨数据资本化中的数据信贷融资及数据证券化，拥有多样化金融工具的数据证券化发展前景广阔。

数据资产化是数据要素市场发展的关键与核心。目前，数据要素市场正在形成包含数据交易主体、数据交易手段、数据交易中介、数据交易监管的“四位一体”发展格局，但仍面临数据确权、数据定价等难点。报告从经济学视角出发，提出数据确权及定价的可操作模型，在数据确权的“三分原则”基础上建立了数据确权路径图，基于数据要素特征基础上，建立了从不完全市场到成熟市场数据定价模型。

针对数据价值化与数据要素市场发展过程中的关键问题、重难点问题，本报告提出解决方案。为推动数据要素流转，深化数据价值，做大做强数字经济提出政策建议。

目 录

一、数据价值化与数据要素市场的概念内涵.....	1
(一) 数据价值化概念内涵.....	1
(二) 数据要素市场概念内涵.....	2
二、数据价值化的现状和进展.....	3
(一) 数据资源化方兴未艾.....	3
(二) 数据资产化加速推进.....	11
(三) 数据资本化点状探索.....	19
三、加快数据要素市场建设.....	21
(一) “四位一体”的数据要素市场格局.....	21
(二) 数据确权“三分原则”及路径实施.....	29
(三) 成熟数据要素市场的“四因素定价模型”.....	38
四、数据价值化与数据要素市场发展建议.....	41
(一) 完善数据要素资源体系.....	41
(二) 构建数据要素市场体系.....	42
(三) 壮大数据要素应用体系.....	44
(四) 建成数据要素安全体系.....	45

图目录

图 1 数据资源化框架图.....	4
图 2 全球每年产生数据量及增速图.....	5
图 3 头部数据标注企业概况图.....	9
图 4 数据标注头部企业基地/工厂区域分布.....	11
图 5 基于区块链技术的数据交易流通体系.....	24
图 6 数据权利束.....	31
图 7 数据确权路径图.....	38



表 目 录

表 1	国内外部分数据交易平台/交易所的定价策略.....	17
表 2	2017—2019 年全球最大的五个数据市场.....	18
表 3	数据需求主体及类型.....	22
表 4	国内数据交易中心.....	26
表 5	省市级大数据管理局.....	28
表 6	部分大数据交易所数据交易规则.....	29
表 7	分类分级数据产权内容.....	36



一、数据价值化与数据要素市场的概念内涵

我们认为：**数据**是对客观事物（如事实、事件、事物、过程或思想）的数字化记录或描述，是无序的、未经加工处理的原始素材。数据可以是连续的值，比如声音、图像，也可以是离散的，如符号、文字。**数据资源**是能够参与社会生产经营活动、可以为使用者或所有者带来经济效益、以电子方式记录的数据。区别数据与数据资源的依据主要在于数据是否具有使用价值。**数据要素**是参与到社会生产经营活动、为使用者或所有者带来经济效益、以电子方式记录的数据资源。区别数据资源与数据要素的依据主要在于其是否产生了经济效益。

（一）数据价值化概念内涵

数据价值化是指以数据资源化为起点，经历数据资产化、数据资本化阶段，实现数据价值化的经济过程。数据价值化重构生产要素体系，是数字经济发展的基础。生产要素是经济社会生产经营所需的各种资源。农业经济下，农业技术、劳动力、土地构成生产要素组合；工业经济下，工业技术、资本、劳动力、土地构成生产要素组合；数字经济下，数字技术、数据、资本、劳动力、土地构成生产要素组合。数据作为数字经济全新的、关键的生产要素，贯穿于数字经济发展的全部流程，与其他生产要素不断组合迭代，加速交叉融合，引发生产要素多领域、多维度、系统性、革命性群体突破。一方面，价值化的数据要素将推动技术、资本、劳动力、土地等传统生产要素发生深刻变革与优化重组，赋予数字经济强大发展动力。数据要素与传统生产

要素相结合，催生出人工智能等“新技术”、金融科技等“新资本”、智能机器人等“新劳动力”、数字孪生等“新土地”，生产要素的新组合、新形态将为推动数字经济发展不断发挥放大、叠加、倍增效应。另一方面，数据价值化直接驱动传统产业向数字化、网络化、智能化方向转型升级。数据要素与传统产业广泛深度融合，乘数倍增效应凸显，对经济发展展现出巨大价值和潜能。

数据资源化是使无序、混乱的原始数据成为有序、有使用价值的**数据资源**。数据资源化阶段包括通过数据采集、整理、聚合、分析等，形成可采、可见、标准、互通、可信的高质量数据资源。数据资源化是激发数据价值的基础，其本质是提升数据质量、形成数据使用价值的过程。**数据资产化**是数据通过流通交易给使用者或所有者带来经济利益的过程。数据资产化是实现数据价值的核心，其本质是形成数据交换价值，初步实现数据价值的过程。**数据资本化**主要包括两种方式，**数据信贷融资**与**数据证券化**。**数据信贷融资**是用数据资产作为信用担保获得融通资金的一种方式，如数据质押融资。**数据证券化**是以数据资产未来所产生的现金流为偿付支持，通过结构化设计进行信用增级，发行可出售流通的权利凭证，获得融资的过程，数据证券化方式较为多样，包括IPO、重组并购、D-ABS（即数据资产证券化）、D-ABN（即数据资产支持票据）等。数据资本化是拓展数据价值的途径，其本质是实现数据要素的社会化配置。

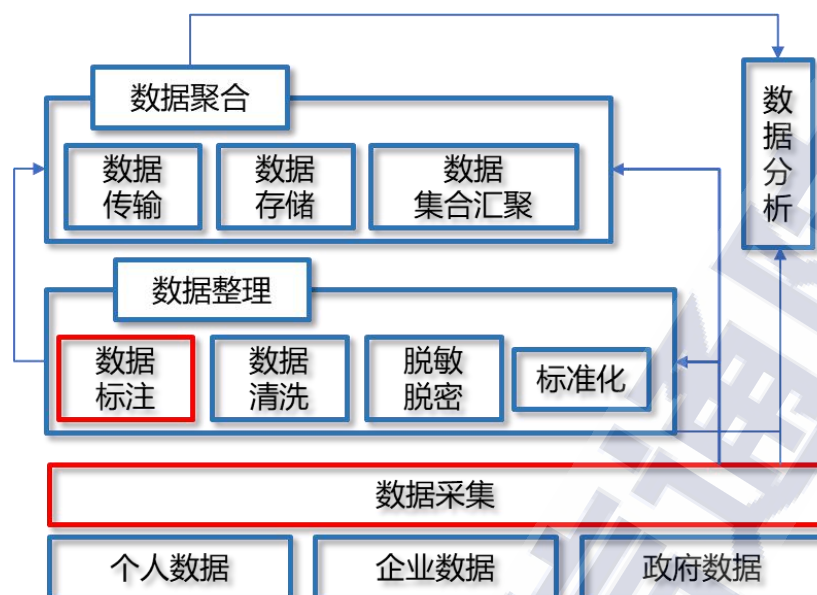
（二）数据要素市场概念内涵

数据要素市场是数据要素在交换或流通过程中形成的市场。要素是指构成事物的必要因素或系统的组成部分，生产要素是生产系统的组成部分，是维持企业生产经营活动所必须具备的基本因素，市场则包含两种含义，其一是交易场所，其二为交易行为的总称。数据要素市场既包括数据价值化过程中的交易关系或买卖关系，也包括这些数据交易的场所或领域。完善数据要素市场是建设统一开放、竞争有序市场体系的重要部分，是坚持和完善社会主义基本经济制度、加快完善社会主义市场经济体制的重要内容。深化数据要素市场化配置改革，促进数据要素自主有序流动，破除阻碍数据要素自由流动的体制机制障碍，推动数据要素配置依据市场规则、市场价格、市场竞争实现效益最大化和效率最优化，有利于进一步激发市场创造力和活力，贯彻新发展理念，最终形成数据要素价格市场决定、数据流动自主有序、数据资源配置高效公平的数据要素市场，推动数字经济发展质量变革、效率变革、动力变革。

二、数据价值化的现状和进展

（一）数据资源化方兴未艾

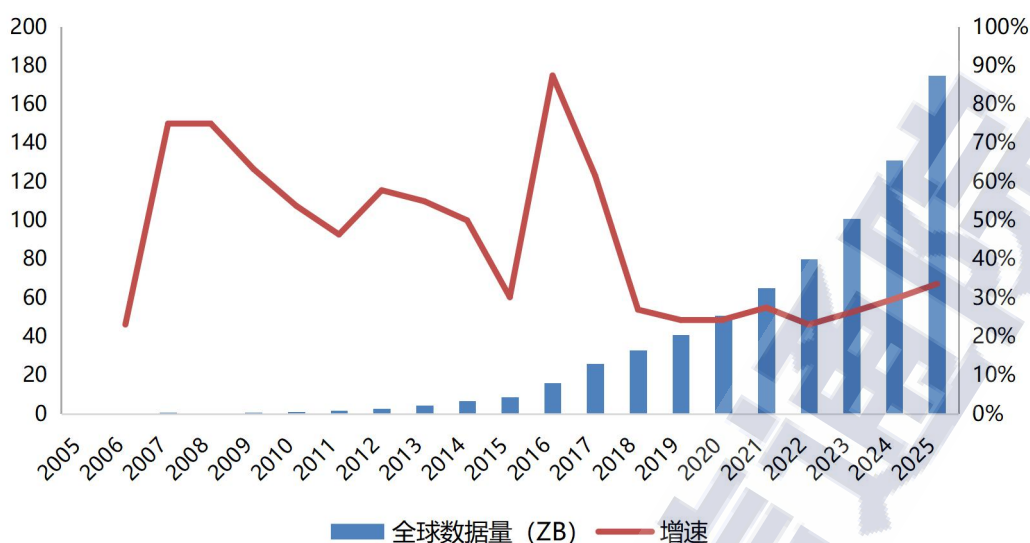
数据资源化是数据价值化的首要阶段，包括**数据采集、数据整理、数据聚合、数据分析**等。数据采集是根据需要收集数据的过程，数据整理包括数据标注、清洗、脱敏、脱密、标准化、质量监控等，数据聚合包括数据传输、数据存储、数据集成汇聚等，数据分析是为各种决策提供支撑而对数据加以详细研究和概括总结的过程。



资料来源：中国信息通信研究院

图1 数据资源化框架图

全球蕴含海量数据资源。根据国际数据公司（IDC）发布的《数据时代 2025》显示，2025 年全球每年产生的数据将从 2018 年的 33ZB（1ZB=10 万亿亿字节）增长到 175ZB，相当于每天产生 491EB（1EB=1.1529e+18 字节）的数据。新一代信息技术的迅速发展及普及、全球数据的“井喷式”生产、数据收集存储和处理成本的大幅下降、机器计算能力的大幅提高，为数据资源化奠定了基础。



数据来源：IDC、Seagate、Statista estimates

图2 全球每年产生数据量及增速图

全球已初步形成较为完整的数据资源供应链，数据采集、数据标注、时序数据库管理、数据存储、商业智能处理、数据挖掘和分析、数据交换等技术领域迅速成长发展。全球看，即使欧美日韩等发达国家，仍处于数据资源化的初级阶段。目前，我国已在数据采集、数据标注环节初步形成了产业体系，数据管理和数据应用能力不断提升。

1. 数据采集产业现状

数据采集是数据资源化的首要环节，是数据标注、数据清洗、数据存储、数据分析等的基础。

数据采集行业产品应用广泛，如电子商务行业通过对商品类别、名称、价格等信息进行数据采集和分析，构建商品比价系统，金融行业通过收集用户的个人交易数据，对用户的征信和贷款进行评级等。数据采集行业主体主要包括采集设备提供商、数据采集解决方案提供商两类。

数据采集设备提供商为数据采集提供传感器、采集器等专用采集设备和智能设备。如工业数据采集通过智能装备本身或加装传感器方式采集生产现场数据，包括设备（如机床、机器人）数据、产品（如原材料、在制品、成品）数据、过程（如工艺、质量等）数据、环境（如温度、湿度等）数据、作业数据（现场工人操作数据，如单次操作时间）等，采集的数据用于工业现场生产过程的可视化和持续优化，实现智能化的决策与控制。

数据采集解决方案提供商通过人工采集服务、系统日志采集系统、网络数据采集系统等方式为客户提供解决方案。

人工采集对象主要包括语音数据、图像数据、视频数据等。语音采集通过采集不同人群的普通话、方言、英文和小语种等各类语音音频，可应用于智能家居、智能设备、智能客服、智慧门店等场景落地。图像采集通过人工拍摄包括人像、商品、汽车、风景等各类真实生活中的图像，助力图像识别模型的训练，可应用于智慧零售、智能设备等场景。视频采集通过人工拍摄指定的物体、人脸、安防等场景的视频，满足多角度、多光线、多场景的多样化采集要求，可在智能安防、智能设备、智慧金融等视觉场景落地。以百度众测为例，其拥有1万名专职外场数据采集员，覆盖40多个国家和地区，遍布全国300+城市，通过其众包平台向数据采集员分配任务，短期内满足客户采集需求。

系统日志是记录系统中硬件、软件和系统问题的信息，同时还可以监视系统中发生的事件，用户通过分析系统日志来检查错误发生的

原因或者寻找设备受到攻击时攻击者所留下的痕迹。互联网公司每天都会产生大量的日志，这些日志一般为流型数据，比如搜索引擎的页面浏览量、查询量，数据量非常庞大。通过进行比对分析和数据挖掘，能够帮助企业更精准的了解用户情况，了解设备的运行情况及安全状态，能够帮助企业提高对用户的服务能力，进而提升营销策略，实现智能运维和统一管控。

网络数据采集包括通过网络爬虫等方式获取数据，对象主要是各类网站，包括新闻类、社交类、购物类以及相应的一些 API、用户接口和一些流型数据。网站 Website、API、流型数据是目前网络爬虫主要爬取的三大类对象，其中 Website 网站数据是网络爬虫的首要对象。

2. 数据标注产业现状

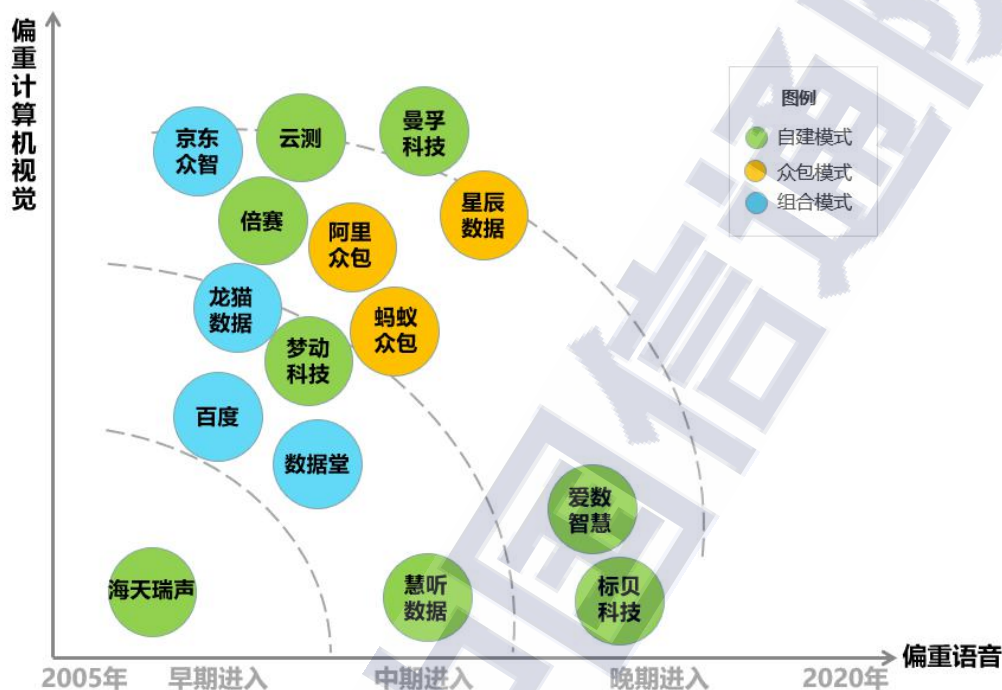
根据美国领先的调查机构 Grand View Research 的一项最新报告，全球（包括美国英国中国等十国）数据标注工具市场规模在 2019 年的收入为 3.9 亿美元，预计 2020 年至 2027 年复合年均增长率将达 26.9%。目前全球已有比较成熟的数据标注企业，如 Appen、iMerit、Infolks，数据标注众包平台如 Playment、Scale AI、Clickworker。现有数据标注以人工标注为主，属于劳动密集型产业，考虑到用工成本，除隐私数据外，欧美国家一般将标注工作转移至第三世界国家，马来西亚、泰国、印度等国家都有欧美数据标注企业分公司。随着机器学习不断完善，自动标注成为大趋势，Google、Microsoft 等互联网公司相继推出了自动标注系统，利用计算机来完成对部分数据的标注。数据标注市场的头部企业通过合作打造新的战略伙伴关系，扩大市场份

额，如 Playment 和 Scale AI 两家提供商合作，为全球高分辨率 LiDAR 传感器制造商 Data 联合开发了高级深度学习标注工具。

中国企业在 2005 年以后逐步涉足标注产业，尤其是 2010 年以后，随着人工智能巨头的崛起，数据标注和采集需求激增，数据标注市场逐渐形成，其提供的标注服务中，文本标注较为基础，多以**语音标注、计算机视觉标注**为主。

从运营模式来看，数据标注企业通过三类模式提供标注服务。众包模式通过搭建众包平台，汇聚数据标注兼职人员力量，成为数据需求方和兼职数据标注员的中介。众包模式有利于节省企业运营成本，但公司对兼职人员管理较为困难，质量难以把控，现有发展较好的众包企业有蚂蚁众包、阿里众包等。自建模式通过自建标注工厂或基地，提供数据标注服务。自建模式有稳定的数据标注员，可以保障专业性和数据质量。但该类数据标注公司大都规模较小，业务承载能力有限，且在项目断档情况下成本压力较大，市场上具有一定规模的专业数据标注公司有 Testin 云测、倍赛、梦动科技、标贝科技等。组合模式将众包模式与自建模式相结合。一方面是互联网公司加入数据标注市场，由于其资本雄厚、自身数据需求强、用户基数大，可凭借自建的标注基地、科学的众包任务分发模式、智能化的数据采集与标注工具，实现规模效应和高效作业。如，百度在山西建立人工智能基础数据产业基地，百度众测推出数据标注开放平台；京东在山东设京东众智大数据标注助残基地，推出 Wise 开放标注平台。另一方面，随专注数据运营的企业规模扩大，可根据项目大小和客户保密要求灵活部署，将

众包和自建模式相结合。如，数据堂成立了合肥数据基地、保定数据基地，运营数加加众包平台；龙猫数据在河南、安徽等十二省建立数据标注基地，运营龙猫众包平台。



资料来源：中国信息通信研究院

图3 头部数据标注企业概况图

从垂直市场来看，数据标注市场可大致分成智能驾驶、智慧家居、医疗卫生、金融服务、新零售、安防和其他领域。近年来，智能驾驶、智慧家居发展迅速，尤其是医疗卫生行业对数据标注的需求显著增长。目前，人工智能技术正被广泛应用于药物开发、基因测序、治疗预测和诊断自动化等领域。数据标注有助于获得支持人工智能技术发展的准确数据，其质量直接影响人工智能应用中算法的准确性和有效性，有望推动医疗卫生行业的智能化发展，同时也意味着未来数据标注市场的门槛会逐步提高，数据标注将由简单标注到复杂标注升级。

从区域分布来看，数据标注已形成以北京为增长极辐射带动三大产业增长带的区域格局。现有的数据标注头部企业有75%总部设在北京，主导了数据标注产业的发展，成为行业增长极，通过技术和业务联系，数据标注产业形成了T字市场结构，产生了三个增长带：环京产业群、环长三角产业群、环成渝产业群。三大产业带发展，又支撑促进增长极中企业的不断壮大，继而发挥示范效应和扩散效应，吸引初创企业纷纷学习、效仿，从而形成一个螺旋上升的循环累积过程。目前，数据标注产业以人工标注为主，企业在华东、华南、华西的一线城市成立分部，管理周边数据标注业务，其标注基地/工厂大都建立在劳动力资源密集省市的小城镇和农村，为当地提供大量就业机会，孵化出新疆和田、河南平顶山、信阳光山县，山东菏泽鄄城县，河北涞源县东团堡乡，贵州百鸟河镇等数据标注村。未来，在垂直市场需求不断精细化趋势下，数据标注产业将催生处更加专业化集聚化的产业集群，数据标注质量和精度也会越来越高。



资料来源：天眼查、数据标注企业官网

图4 数据标注头部企业基地/工厂区域分布

（二）数据资产化加速推进

数据资源化使种类丰富且内容庞杂的数据拥有了使用价值，发展潜力十分巨大。数据资产化，使具有使用价值的数字成为一种资产，在市场上进行流通交易，给拥有者或使用者带来经济利益。数据资产化是构建数据要素市场的关键与核心，包括数据权属的确定、数据资产的定价、数据的交易流通。

1. 全球数据权属探索现状

针对数据确权，全球各国在法律制度上进行了不同探索。欧盟最早进行体系性构建，通过《一般数据保护条例》（GDPR）和《非个人数据在欧盟境内自由流动框架条例》，确立了“个人数据”和“非个人数据”的二元架构。针对任何已识别或可识别的自然人相关的个人

数据，其权利归属于该自然人，其享有包括知情同意权、修改权、删除权、拒绝和限制处理权、遗忘权、可携权等一系列广泛且绝对的权利。针对个人数据以外的非个人数据，企业享有数据生产者权，不过其权利并非是绝对的。

美国依托现有制度，加大对数据隐私的保护。美国并无针对数据的综合立法，而是将个人数据置于传统隐私权的架构下，利用“信息隐私权”来化解互联网对私人信息的威胁。同时通过《公平信用报告法》、《财务隐私法》、《有线通信信息法》、《健康保险携带和责任法》等法律，在金融、医疗、通信等领域制定行业隐私法，辅以包括网络隐私认证、建议性行业指引等行业自律机制，形成了“部门立法+行业自律”的体制。

日本严格界定数据保护范围，并不主张对数据本身另行设定新的排他性私权。经过学界、产业界以及政府部门的多方探讨，目前日本对数据权属问题的处理规则已经比较明确。概括来说，对数据权属以自由流通为原则，特殊保护为例外。即，以构建开放型数据流通体系为目标，不突破现有法律规定和法解释，不对数据另行设置私权限制，以尊重数据交易契约自由为原则，促进数据自由流通。

俄罗斯规定的数据主体的权利与其他国家落脚点不同，多相对于处理人而言。所谓处理人，是指独立或与其他单位合作而处理个人数据，并能确定个人数据的处理的目的、范围的国家机关、主管机关、法人或个人。数据主体的权利包括知情权；更正、中止、删除权；可携带权；防止自动化决策权、被遗忘权和诉权等。

印度《2018年个人数据保护法案（草案）》将数据视为“信托”问题，将每一个决定处理个人数据目的和方法的实体定义为“数据受托人”，并要求其承担主要责任。数据受托人是指单独或者与其他人一起决定处理个人数据之目的和方式的任何人，包括邦、公司、法律实体或个人。

国内看，中央及地方积极探索数据确权，部分地区出台相关文件，建立相关平台，筹划数据确权发展。但整体来看，数据确权尚处于起步阶段。

广东省率先发布较为详细的数权政策文件。以**深圳**为代表发布《深圳经济特区数据条例》，提出探索完善数据产权，着力解决数据要素产权配置问题。创设数据权，明确数据权的财产权属性与数据权的内容，明晰个人数据权属、公共数据权属。**广州**在《广州市加快打造数字经济创新引领型城市若干措施》要求重点在数据确权先行先试，全面开展对数据确权相关法律法规的预研，开展数据确权流通沙盒实验，形成一批实验性成果。

各地也积极进行了数据确权的**实践探索**。2019年9月**工信部**开通了我国首家数据确权平台“**人民数据资产服务平台**”，主要是对数据的合法合规性进行审核，对数据生产加工服务主体、数据流通过程、数据流通应用规则的一系列审核及登记认证。**北京**筹建**北京国际大数据交易所**，要求建立以信息充分披露为基础的数据登记平台，明晰数据权利取得方式及权利范围，建立数据确权工作机制，提供包括数据产品所有权交易、使用权交易、收益权交易在内的数据产品交易服务。

河南省新乡市试点上线数据要素确权与可信流通平台（河南根中心），发出全国首张数据要素登记证书，新乡实施数据要素确权与可信流通平台项目，建立了数权科技研发团队，首创了数据资源规范确权算法等核心技术，基于区块链分布式共识明确了数据要素的拥有权与控制权。贵州省支持建设基于区块链的数字资产交易所，探索数据确权新模式，明确由贵州省大数据局和贵阳市政府作为责任单位，利用贵阳大数据交易所数据交易平台基础，实施“基于区块链的数据资产交易平台”项目。浙江大数据交易中心发布大数据确权平台，通过采用开源大数据分布式计算框架和数据可用但不可见的混淆加密算法对数据确权认证。

2. 不完全市场下的数据定价策略

数据资产无实物形态且具有非货币性，现有数据定价研究多类比无形资产，采用成本法、收益法、市场法。然而由于数据资产重置成本难以确定、价值可持续挖掘、数据使用寿命无法估计、交易市场规模小等原因，传统的数据定价方式并不适用于目前的不完全市场。

不完全市场下的数据定价实践，数据的价格受多因素影响，包括数据量、数据种类、数据深度、数据完整性和数据实时性等。因此平台多通过采取不同定价策略，提高数据供给方参与积极性，满足数据需求方的差异性需求，实现供需双方效益最大化。总体来看，数据定价策略采用了静态定价策略与动态定价策略相结合的方式。

静态定价策略包括固定定价、差别定价、拉姆齐价格。

固定定价是指数据卖方和交易平台根据数据商品的成本和效用，结合市场供需情况，设定一个固定价格在交易平台上出售，最终成交价即为该固定价格。固定定价的优势在于价格固定，节省撮合协调的时间成本和沟通成本；其局限在于适用范围较为狭窄，仅限于批量廉价的数据交易。以 **Data plaza** 为例，该平台采用固定定价方式收费，在 **Data plaza** 上，用户可以通过列表选择需要的数据进行下载。数据在对全部个人信息进行匿名化处理后进行交易。价格因数据量和内容不同而不同，一般在数万至数千万日元之间。根据定价收费依据的不同，分为按数据库范围收费、按数据大小收费、按平台数据计算单位定价、按数据提供方定价收费等。根据收费时间的不同，可分为按年收取、按月收取和现收现付。

差别定价是指以两种或两种以上不同反映成本费用的比例差异的价格来销售一种数据产品或服务。这种差别定价是基于不同的消费者获取数据的愿望不同而实现的。如 **Factual** 的定价方式是通过浮动价格向公司和独立软件开发商出售数据，其定价依据是信息的使用量。对于小规模的数据，**Factual** 提供是免费服务，而针对大型客户，则可能会收取百万甚至千万美元的服务费，包括 **Facebook**、**CitySearch**、**AT&T** 在内的一些大公司都会使用 **Factual** 来获取相关数据信息。

拉姆齐价格是一种高于边际成本的定价，此价格下净收益与净损失的差值最大。这种定价策略主要是针对公共数据服务，它们经济效益不高却极具社会效益，通过设置拉姆齐价格有利于提高效率。政府作为公共部门，其以“福利最大化为目标”是较为理性的。政府与公

司之间形成协同合作：脱敏之后的大数据以高于成本价格的拉姆齐价格对所有公司开放，同时政府以购买服务的形式从公司获得所需的大数据处理结果，这样，政府从大公司获得的收益反过来又被公司所得到，同时社会总福利得到了增加。

动态定价策略包括自动计价、协商定价、拍卖式定价。

自动计价是指交易所针对每一个数据品种设计自动计价公式，卖方和买方在交易系统的自动撮合下成交。根据成交方式，最终成交价分为三种形式：一是自动成交价格。当买方应约价大于或等于卖方挂牌价时，交易系统自动撮合成交，最终成交价为买方应约价。二是卖方选择成交价格。对于不能直接成交的应约，卖方可选择能接受的应约价与其成交，成交价为买方应约价。三是数据分拆成交。当买方仅需要部分数据时，平台将对数据设定拆分原则，系统自动报价，而后自动撮合双方成交。其优势和特点在于定价方式的自动性，即依靠数据交易平台设定的交易系统，完成数据的自动撮合成交。但交易系统的建设成本高，过程复杂，对数据交易平台有极高要求。

当交易双方对大数据价值的评估不一致时，买方和卖方可以直接通过**协商定价**达成对数据商品价值的一致认可。协商定价的基础是各方对数据价值的认可，只是对价值大小的认知程度不一，协商的作用就是尽可能的取得双方对大数据价值的一致性。如湖北一家农业领域的上市公司需要数据帮助其开发新产品，长江大数据交易中心对公开历史数据进行加工，价格由双方协商，一笔是一笔。此外，交易双方可通过反馈进行数据价格的动态调整。买方获取数据产品，使用后及

时反馈，买卖双方对价格进行再调整。其优势在于数据价格在不断反馈、不断调整的循环过程中逐渐趋于合理化，有利于提高数据价格与市场实际的匹配度，促进交易双方的长期合作和市场稳定。其劣势在于现实中反馈者、反馈时间、反馈质量等均不好衡量和把握。

拍卖定价属于需求导向定价，适用于一个卖方和多个买方交易的情形。基于诚信的拍卖定价不仅能够使买卖双方就价格达成一致，还能实现数据的商品价值最大化，有利于促进数据交易流通。例如假设A、B两个公司都在尝试获取一套数据，通过大数据交易平台进行拍卖报价。在该种定价策略下，大数据交易平台会将保留价格（保留价格=（数据卖方固定成本+可变成本）×一个比例，比例由大数据交易平台或数据卖方确定）作为底价，让A、B两家公司对数据进行竞拍，价高者得，最终成交价即为该数据的价格。拍卖定价的局限性在于容易发生买方通过约定实现价格分摊、合伙竞拍的道德风险。

表1 国内外部分数据交易平台/交易所的定价策略

	定价方式	大数据交易平台/交易所
静态定价	固定定价	Quandl、Azure、Oracle、GoodData、Data plaza
	差别定价	Factual
动态定价	自动计价	Qubole、浙江大数据交易中心、贵阳大数据交易所
	协商定价	长江大数据交易中心、上海数据交易中心、华中大数据交易平台、贵阳大数据交易所、浙江大数据交易中心、中关村数海大数据交易平台
	拍卖式定价	上海数据交易中心

资料来源：中国信息通信研究院

不完全市场条件下，尽管国内外已经探索出许多数据定价方法，但远未形成规范有序的数据定价规则，数据定价指标体系仍处于混乱

无序状态。一是价格影响指标尚不统一，不同的交易所对同一数据产品的定价因价格影响指标的不同而不同，阻碍了统一市场的形成。二是数据质量评价指标只能对数据质量进行定性的好坏比较，反映到价格上则只能比较相对价格的高低，无法给出大数据产品的确切价格区间。三是数据历史成交价指标虽具有重要的参考价值，但由于数字时代市场供需变化较快，叠加数据产品的复杂性和数据交易市场的成熟性，其作用难以有效发挥。随着数据要素市场的不断完善，这些问题也将通过数据定价模型逐步解决。

3.数据交易发展现状

随着数据资产化的进程加快和新兴技术的不断融合发展，数据交易呈现稳步发展的态势。On Audience统计显示，2017—2019年全球最大的五个数据市场的市场交易值增长率均在20%以上，规模最大的美国市场交易值在2019年达152.09亿美元。中国数据市场发展迅速，交易值增速在全球遥遥领先，2017年、2018年两年的交易值均接近翻番，2019年也在60%以上，达到23.93亿美元的规模，超过英国的23.55亿美元。

表2 2017—2019年全球最大的五个数据市场

国家	2017		2018		2019	
	市场交易值 (亿美元)	增长率 (%)	市场交易值(亿 美元)	增长率 (%)	市场交易值 (亿美元)	增长率 (%)
美国	97.82	34.9	123.41	26.2	152.09	23.2
英国	14.52	22.3	18.82	29.6	23.55	25.1
中国	7.47	127.2	14.61	95.5	23.93	63.8
加拿大	4.53	30.1	5.88	29.7	7.69	30.6

法国	2.32	56.4	3.41	46.8	4.70	37.8
----	------	------	------	------	------	------

数据来源：On Audience.com, Global Data Market Size 2017—2019

（三）数据资本化点状探索

数据资本化阶段，数据被打包成金融产品进入资本市场，推动资本集聚，促进资源合理配置，发挥数据要素对经济社会发展的乘数效应，实现数据价值的深化。

数据资本化使**数据由货币性资产向可增值的金融资产转化**。其中，数据信贷融资可将数据货币价值迅速变现，获取企业再生产所需资金，数据证券化可将数据货币价值转换为权益价值。如，数据所有人将数据作为财产作价后，以出资入股的形式与其他形式的财产（如货币、实物、土地使用权等）结合，所有权人通过暂缓即时兑现数据价值获得部分股权，其本质是将数据的货币价值转换为股权价值。

数据资本化是**融资者和投资人共同分享数据收益的过程**。从**融资者**角度看，融资者在获得融资的同时也可保留数据自主权。在数据资本化后，发起人仍可保留和管理数据，因此，融资人在实现资金融通的同时，也可对进一步加工和应用数据，持续提升其价值。此外，数据资本化还可为融资者提供较高的融资杠杆，减少资金投入量，快速筹措资金。从**投资者**角度来看，数据资本化产品流动性较好，投资人无需考量发起人的经营状况，可直接投资高价值数据集合，若数据产品盈利，即可与融资人共享数据收益。

目前，国内外已有企业展开数据资本化创新性探索，总的来看，主要有以下四种形式：

一是**数据证券化**，依托数据资产，通过 IPO、并购重组等手段获得融资。如，由于 LinkedIn 拥有高达 4 亿的用户，产生大量行为数据，Microsoft 在 2016 年以 262 亿美元收购拥有海量数据资产的 LinkedIn，溢价超 50%；Facebook 依托自身数据资产以高市值上市，上市市值超财报公布的资产价值逾 14 倍，巨大差额的背后是 Facebook 未在账面体现的数据资产——8.45 亿个月活跃用户及其产生的行为数据。

二是**数据质押融资**。数据质押融资是数据权利人将其合法拥有的数据出质，从银行等金融机构获取资金的一种融资方式。如华夏银行杭州分行推出电商贷产品，通过获取企业经营数据、创建信贷估值模型等方式，分析电商企业数据，短时间内为符合条件的电商企业提供贷款。

三是**数据银行**。数据银行通过吸纳“数据存款”，把分散在个人和集体中的数据资源集中起来，使其易被发现、访问、并具备互操作。如 Science DB，面向学术期刊、科研人员提供数据在线存储、汇交管理、长期保存与获取、共享、出版和引用服务。科研人员通过数据银行储存和出版自己收集整理科学数据，其他研究人员可在数据银行下载和使用共享的科学数据。

四是**数据信托**。中航信托发行了全国首单基于数据资产的信托产品，总规模为 3000 万元。该产品的运作模式分几方面：第一步，数据堂将自己所持有的某一个数据资产包即数据资产作为信托财产设立信托；第二步，信托受益权转让，委托方通过信托受益权转让获得现金收入；第三步，受托人继续委托数据服务商对特定数据资产进行

运用和增值，产生收益；第四步，向社会投资者进行信托利益分配。在这个过程中，既完成了资金的循环，同时也完成了数据资产信托财产的一个闭环。

三、加快数据要素市场建设

随数据的价值化发展，我国数据要素市场格局逐渐明晰，正在形成包含数据交易主体、数据交易手段、数据交易中介、数据交易监管的四位一体市场格局。数据要素市场构建要求数据资产具备三个条件：可控制、可量化和可获益。可控制是指数据应有明确的产权归属，可量化是指数据资产的价值可以用某种数量指标或货币来衡量，可获益是数据资产能带来经济利益，具有交换价值与价值。尽管随数据资产化的加速推进，数据确权及定价有了一定的探索，但仍是数据要素市场发展的重点和难点。加快数据产权界定、建立数据定价机制是培育数据要素市场的关键。

（一）“四位一体”的数据要素市场格局

1. 数据交易主体

从供给端来看，数据交易主体由政府主导向社会多主体共建发展，即由政府指导类、数据服务商类、大型互联网企业三类主体共同参与。

政府指导类是数据交易市场加入最早也是规模最大的参与主体。政府通过指导建设平台、设立数据交易所等方式，推动数据交易，如上海数据交易中心、贵阳大数据交易所。由于这类交易在一定程度上有政府背书，因此具有一定的权威性。政府指导下的数据交易机构主

要分布在西南、华东和华北地区，其分布与地区的基础设施能力及经济发展水平相关。**数据服务商**对数据进行“采产销”一体化运营，盈利性较强，如数据堂、数海、龙猫数据等。其主要特征是向用户直接交付数据产品或服务，包括原始数据、加工处理后的数据以及由多份数据整合后的新数据，分别对应数据产生者、数据加工者和数据整合者的角色。以数据堂为例，其收录了1.8亿+家社会实体信息、90多种维度信息全量实时更新。**大型互联网企业**投资建立的交易平台以服务大型互联网公司发展战略为目标。百度、阿里巴巴、腾讯、360等互联网企业先后推出数据产品服务，抢占数据市场资源。如京东建立的京东万象数据服务商城，主要为京东云平台运营提供支撑。其数据商品涵盖了线上零售、生活服务、企业数据、农业、资源能化等10大类。提供17个API接口、165个数据集、56个数据报告、278个政府开放数据。

数据市场的另一端是数据的**需求方**，包括各类数据分析服务商和行业用户，尤其数据驱动型的公司对数据拥有强烈的需求。需求的数据涉及政府决策、公共服务、影视娱乐、交通物流、医疗健康、金融、零售、广告营销、农业、能源等领域。

表3 数据需求主体及类型

需求主体	数据品种	核心数据类型
医药公司、医疗设备公司等	医疗数据	病历数据、就诊数据、药品流通
银行、小贷公司、互联网金融公司	金融数据	企业数据、个人数据、个体户数量
	企业数据	中小微企业数据、外资企业数据等
能源企业	能源数据	石油、天然气等所有相关的数据
车联网、汽车公司、汽车后市场	交通数据	停车场数据、车辆位置数据等

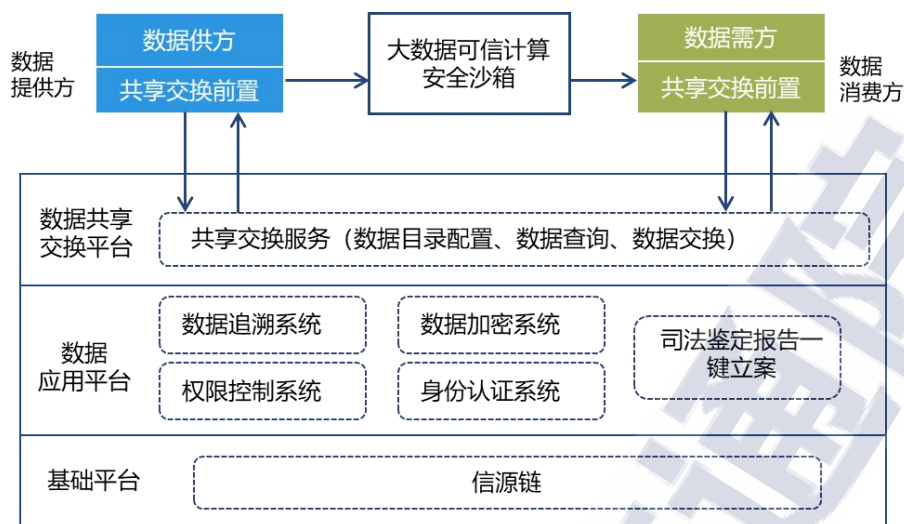
供应链相关企业	商品数据	电子标签数据、商品物流数据等
金融机构、汽车公司、消费品公司	消费数据	个人消费数据、个人征信数据等
教育类机构	教育数据	学习轨迹数据、教育消费数据
政府相关部门	社会数据	与社会管理、政府管理有关的数据
其他（如科研机构等）	社交数据	与社交相关的所有数据
	政府数据	政府统计数据，政府审批数据等
	电商数据	商品交易数据、商品流通数据等

资料来源：中国信息通信研究院

2. 数据交易手段

区块链技术作为数据交易手段被初步运用并逐渐普及。现有的数据交易手段利用 API 接口、数据包交易、EXID 虚拟标识技术等方式。相较于这些技术，区块链具有去中心化、难以篡改、可溯源等特点。它并不是一种单一的信息技术，而是依托于现有技术，加以独创性的整合及创新，创造一种全新的信任方式。随着区块链技术的发展与成熟，“区块链+数据交易”成为企业的选择重点。

如图所示，通过搭建基于区块链技术的平台，促进数据的交易流通。量子加密等技术的开发应用，解决数据交易流通中的安全保密问题，确保数据安全；区块链技术在数据交易流通中的应用，确保数据流通可信、透明、可追溯，解决数据交易流通中数据非授权复制和使用等问题，提高企业参与数据交易的积极性。



资料来源：中国信息通信研究院

图5 基于区块链技术的数据交易流通体系

通过为用户和数据均生成唯一的数字身份ID，用赋码机制确保资产唯一性，为每个数据资产确权。数据提供方将共享数据的元信息登记上链，设置数据目录的访问权限（如公开、指定接收方、授权共享等），授权记录和数据访问记录被永久记录在区块链中。这种方式有利于形成信息闭环，支撑跨机构共享数据的全流程数据要素权属管控，保证连续性和可追溯性。

数据共享交换平台规范数据共享的目录和数据交换标准，实现数据交换的管理，用数字信封加密或者可信计算安全沙箱确保数据隐私安全，可信计算安全沙箱能够保障只输出数据分析结果，数据不落地、不外泄。借助密码学、共识算法和分布式存储等技术，组合出一种新的数据共享方式，通过数据的公开透明、不可篡改与集体维护等措施，让整个系统降低了信息不对称，从而促成新的信任机制。

数据应用平台根据应用场景的需要，可综合使用多个系统，如数据追溯系统、权限控制系统、数据加密系统身份认证系统等。以司法

存证为例，为对接司法业务场景，区块链可以结合很多相关技术，包括电子身份认证、时间戳服务、数据加解密、数据存储、执行智能合约等。依托应用平台，参与数据流通的机构可以根据业务需求快速灵活的构建满足特定业务场景的数据服务，极大降低数据流通的技术对接成本，有助于建立数据开放和数据流动的制度规范。

基础平台主要依托信源链，可为用户提供存在性证明、完整性证明、身份证明、时间戳证明、数据关系证明和凭证登记流转记录等多种服务，确保这些信息具备可验证、可审计、可追溯、不可篡改等特性。

总体来看，通过基于区块链技术的数据交易流通体系，一是能够拉动数据供需双方交易率、提高流通效率，通过建立数字身份和可信数据凭证体系等，使数据交易流通安全可信。二是可以为数据治理方、监管方提供完整的技术管理手段，从而可以确保数据授权和数据交易符合国家政策法规和监管要求，有利于建立开放、透明的数据流动监管体系。三是有利于消除信息孤岛，在不做数据整体物理迁移的情况下完成数据的开放和流通，有效整合分散异构的数据资源，快速消除信息孤岛。

3. 数据交易中介

截至目前，我国大数据交易中介经历了三个阶段：井喷式爆发期（2014年—2016年）、发展停滞期（2017年—2019年）、重现新生期（2020年至今）。

随着大数据写入政府工作报告，2014年成为中国大数据元年。随

后，大数据交易中心数量呈现井喷式增长态势，仅2015和2016两年，就有13家大数据交易中心（平台）成立。然而在2017年以后的几年间，各地的新增数量骤降，已成立的数据交易所实现的数据交易也远未达到预期，数据交易中心的发展踩下了急刹车。2020年4月，中共中央、国务院正式发布《关于构建更加完善的要素市场化配置体制机制的意见》，将数据与土地、资本、劳动力并列为关键生产要素，并提出加快培育数据要素市场的愿景，大数据交易市场才又现生机。各地纷纷加速建立大数据交易中心。8月11日，北部湾大数据交易中心在南宁成立，此外，雄安、北京也在筹建新的数据交易中心。

随着数据交易市场的不断发展，第三方数据交易平台的市场定位出现**综合化、服务化的趋势**，数据交易中介由单一的居间服务商向数据资源综合服务商转型。数据交易中心建立初期，多做一些数据发现、供需撮合、计价清算等简单业务，随着数据交易中心不断创新，完善数据交易服务框架，服务内容多元化、综合化，出现包含数据清洗、数据加工整合、数据分析、数据可视化等数据资源综合服务商。

表4 国内数据交易中心

状态	序号	名称
筹建	1	雄安大数据交易中心
	2	北京国际大数据交易所
已建	3	贵阳大数据交易所
	4	中原大数据交易中心
	5	西咸新区大数据交易所
	6	华东江苏大数据交易中心
	7	哈尔滨数据交易中心
	8	上海数据交易中心
	9	北部湾大数据交易中心
	10	香港大数据交易所

11	华中大数据交易所
12	东湖大数据交易中心
13	长江大数据交易所
14	浙江大数据交易中心
15	钱塘大数据交易中心
16	重庆大数据交易市场
17	中关村数海大数据交易平台

资料来源：中国信息通信研究院

但目前数据交易中介的发展仍然存在一些问题。一是缺乏有吸引力的数据库支撑，数据质量差、数量少、更新慢。与国外基于数据库交易的方式相比，当前已建的数据交易中心并未形成坚实的数据库支撑，提供的数据库质量参差不齐，时效性差，不能满足市场需求。二是供需不对称，数据交易平台盈利状况欠佳。大部分平台处于推广阶段，提供的数据库不能满足买方需求，收取的平台费用及交易佣金较少。三是数据安全保护与数据交易流通矛盾突出。数据交易中心进入发展停滞期的一个重要原因是《网络安全法》的实施和刑法执法力度的加大带来的合规性风险远大于收益。当前，数据安全保护仍是制约数据交易流通的一把枷锁，亟待相关机制予以解决。

4.数据交易监管

对数据交易做好监管，是保障数据交易市场环境公开、公平、公正的基础。现有的数据交易监管主要聚焦数据交易机制的建立与完善，从制度创新、资源融合共享、公共数据开放、应用创新、产业聚集、要素流通、交易监管等角度，规范数据交易行为，增强政府及行业的监督与管理。数据交易的监管主要依托政府，数据服务机构自律为辅助。

目前，全国各地已经相继成立了大数据管理局，监督管理数据交易市场，促进数据资源流通。省级层面已有广东省、浙江省、贵州省等 14 个地区设立了省级的大数据管理机构，省级以下各市、区大数据管理局也达到 12 个，如广州市大数据管理局、贵阳市大数据发展管理委员会等，负责统筹本地区内数据监督管理工作。

表 5 省市级大数据管理局

省级大数据管理局（14 个）		
北京市大数据管理局	河南省大数据管理局	贵州省大数据发展管理局
福建省大数据管理局	吉林省政务服务和数字化建设管理局	浙江省大数据发展管理局
山东省大数据局	广西壮族自治区大数据发展局	广东省大数据管理局
海南省大数据管理局	内蒙古自治区大数据发展管理局	江西省大数据中心
上海市大数据中心	重庆市大数据应用发展管理局	
市级大数据管理局（12 个）		
南京市大数据管理局	广州市大数据管理局	沈阳市大数据管理局
成都市大数据中心	兰州市大数据管理局	保山市大数据管理局
黄石市大数据管理局	咸阳市大数据管理局	银川市大数据管理服务局
昆明市大数据管理局	贵阳市大数据发展管理局	宁波市大数据发展管理局

资料来源：中国信息通信研究院

数据监督方面，地方数据主管部门负责指导、协调和监督本行政区域内的数据交易活动，引导数据供需双方在依法设立的数据交易服务机构进行数据交易，监督数据交易服务机构履行有关规定的情况。数据交易事前阶段，数据管理部门对数据供需双方披露信息进行监督，督促其依法及时、准确地披露信息。在交易实施环节，数据管理部门对交付数据内容进行监测和核验，如发现违法违规事件，及时中断数据交易行为。在数据交易事后监督检查中，对数据交易行为或交易平台存在较大安全风险的，提出改进要求并督促整改。

数据管理方面，政府加强数据管理条例的设置，对数据交易服务

平台及数据交易行为进行管理。2020年7月30日天津市发布的《天津市数据交易管理暂行办法（征求意见稿）》，是我国第一部专门对数据交易进行规范的地方性立法。该办法在探索适合我国国情和数据特点的交易制度方面具有重要意义，一方面可为《数安法草案》的完善提供地方经验，另一方面也会给其他地方数据立法提供借鉴。**数据交易服务机构**通过发布平台交易规则的方式，包含用户管理、交易管理、订单管理、平台管理等制度，增强平台数据管理力度，提高数据交易入门门槛。如数据交易服务机构在数据交易时，与数据供方和数据需方签订三方合同，明确数据内容、数据用途、数据质量、交付方式、交易金额、交易参与方安全责任、保密条款等内容，并对交易订单进行审核管理，确保符合相关法律法规和标准等合规性要求，最终形成完整的交易日志并安全保存。目前我国已有多个大数据交易所设置数据交易规则。

表6 部分大数据交易所数据交易规则

大数据交易所	数据交易规则
贵阳大数据交易所	《大数据交易观山湖公约》、《贵阳大数据交易所 702 公约》
华中大数据交易所	《大数据交易安全标准》、《交易数据格式标准》、《大数据交易行为规范》、《大数据交易管理条例》
上海大数据交易中心	《数据流通禁止清单》及《数据流通原则》
中关村数海大数据交易平台	《中关村数海大数据交易平台规则（征求意见稿）》

资料来源：中国信息通信研究院

（二）数据确权“三分原则”及路径实施

数据确权使得数据资产具备可控制性，有利于加快数据要素的流通速度。通过确定数据产权，稳定数据持有者对未来数据使用权及收益权的预期，减少出现数据纠纷的风险，提高企业参与数据要素市场

的积极性。数据确权加速数据流通，使有更高生产能力和资源禀赋的组织集中数据资源，生产效率较低的个人或组织转让数据以获得收益。这一双向对接过程将不断优化数据资源，提高生产效率，推动数据要素市场化进程。

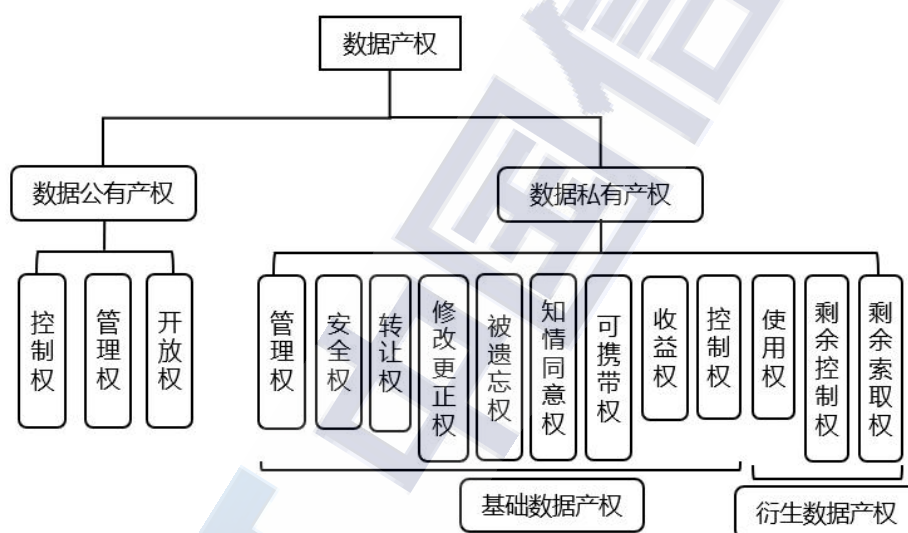
1.数据确权“三分原则”

我国现有的数据确权探索仍处于政策萌芽层面，实际落地需要一定时间。在数据确权初期，可以从数据原则的构建出发，立足数据性质，把握确权方向，进而制定确权路径。

一是分割原则。

数据确权是为了实现不同利益主体激励相容，即平衡数据价值链中各参与者的权益，实现在用户隐私合理保护基础上的数据驱动经济发展。因此，数据确权需要解决的是附着于数据的权益归属而非单纯的所有权归属，核心是确定哪些利益应该受到保护，即数据确权保护的应该是利益而非所有权。所有权的核心功能在于明确财产权利的排他性，产权区别于所有权且具有比所有权更宽泛的范畴。德姆塞茨认为产权是一种社会工具，其重要性在于它能帮助一个人形成他与其他人进行交易的合理预期。因此确立数据产权框架比明确数据所有权更加有利于实现数据资产化，推动数据的交易流通。

数据产权是由多种权利构成的权利束，权利束确定哪些利益应该受到保护。数据产权可以分割，并随着社会经济生活演变而不断扩张。设置权利束的目的是为了实现合理保护消费者隐私的同时，激励企业数据采集以及对数据要素充分开发利用。总体思路是对公共数据强调集体权益不受侵犯而共享收益最大化，对原始数据强化个人信息的隐私保护，对数据企业添附后的衍生数据则突出利益保护。基于此，构建数据独有的数据权利束如下。



资料来源：中国信息通信研究院

图6 数据权利束

对于公共数据，集体对其拥有管理、监督、制约和保护的权利，**数据公有产权**的客体是集体共有的数据。具体而言，数据公有产权主要包括三个方面：一是控制权，即对集体内部数据的安全性、真实性和完整性采取有效措施予以保护，以免数据遭受被篡改、伪造、泄露等危险；二是管理权，即对集体内部数据的生产、加工、流通等进行全生命周期的管辖；三是开放权，即集体将掌握的数据资源根据需要在集体内部公开、共享。

私有数据对应数据的私有产权，包括基于原始数据的基础数据产权及经添附后的衍生数据产权。其中原始数据指的是未加工处理的数据，衍生数据是在原始数据基础上经过算法加工、计算、聚合而成的系统的、可读取、有使用价值的数据，添附有附合、混合、加工等多种类型。

具体来看，**基础数据产权**包含管理权、安全权、转让权、修改更正权、被遗忘权、知情同意权、可携带权、收益权、控制权。管理权是决定怎样和由谁来使用数据的权利。安全权即免于被剥夺的权利，是数据不被他人非法侵扰、知悉、搜集、利用和公开等的一种权利。转让权即将自己的原始数据的合法利益或权利让给他人的权利。修改更正权是指数据主体有权要求数据控制者或管理者对其错误的、过时的个人数据进行修改、更正和补充。被遗忘权是指数据主体有权要求数据控制者或管理者及时删除其个人数据，并通知相关第三方停止利用和传播。知情同意权是指主体在采集或处理个人数据前均须先告知数据主体并征得数据主体同意，告知和同意的内容包括采集数据的目的、用途、数据处理方式、程度等及其后续的变化。可携带权是数据主体有权以结构化、常用和机器可读的格式获得其提供给控制者的个人数据，或有权无障碍地将此类数据从其提供给的控制者那里传输给另一个控制者。收益权是指通过其生产的原始数据产品获取经济利益的权利。控制权即可以根据自己的意志实施对数据未被法律或其他合约所明确约束的权利。

衍生数据产权主要包括三方面：使用权、剩余控制权、剩余索取权。使用权是指数据添附者有权对数据通过独有算法或手段处理过的数据进行各种形式的利用，如进行二次清洗加工等。从激励意义上来说，数据产权安排不能以牺牲价值创造为代价。对数据价值创造贡献最大的一方应该拥有以剩余控制权和剩余索取权，以激励其投资于数据开发利用和大数据驱动的创新。因此，企业对添附后的数据还拥有剩余控制权和剩余索取权，剩余控制权是指数据添附者拥有决定数据及数据产品的运营策略的权利。剩余索取权是指数据添附者通过添附手段，从合法获取的原始数据中得到经济利益的权利。

二是分类原则。

根据数据主体的不同，将数据分为**个人数据、企业数据、社会数据**三部分。

个人数据是指能够识别自然人身份的数据或由于自然人行为产生的数据。个人数据包括自然人独有的特征数据和参与经济活动、社会活动的行为数据，是属于个人的数据，如个人的姓名、电话、住址、职业、学历、偏好、习惯、旅游去过的城市、购物的交易记录、上网浏览的页面等数据。个人数据有明显的敏感性隐私性特征，同时通过添附手段可明显提升数据价值。通过数字化形成的个人数据涉及到个人隐私，因此为避免数据交易成本增加，可将基础数据产权赋予个人。同时，作为数据的主体可以选择将私有数据的部分权利让渡给企业或社会，既为数据主体提供更加优质化的服务，也使数据资源达到最优配置，更有效的降低数据交易成本。此时，个人作为原始基础数据的

拥有者，如果明确同意企业收集其个人信息，那么当企业通过这些数据获得经济或其他利益时，个人也应该部分享有数据收益权，这种收益权不一定以现金或货币的方式体现，企业也可以为用户提供一些除货币之外的免费增值服务，分享方式可灵活多样。

企业数据是企业生产经营活动中产生或合法获取的各类数据。企业数据的组成有企业主体数据、经用户授权的企业数据。企业对主体数据享有基础数据产权，对来源于用户的数据享有部分权利，相比用户，企业更有动力投入人力物力，通过整理、存储、分析数据，挖掘更多价值，应享有部分权利。在征得用户同意后，不侵犯个人数据隐私权的前提下，可以在脱敏脱密后进行下一步的开发利用。添附数据产权可以用契约的方式解决，用户在登录网站或 APP 等通过接受企业的条款将个人信息的数据授权给了企业，相当于用户在将数据的部分权利让渡给企业，而企业依照约定拥有这部分数据的添附产权。

社会数据包含政府及公共机构在开展活动中依法收集的各类数据及其衍生数据，如自然资源数据、经济社会数据。政府及公共机构对社会数据享有数据公有产权。社会数据往往牵涉社会公共利益，其权属的赋予应该将权利配置的重点放在社会要素这一属性上。公共资源、公共财产的属性，使社会数据的价值体现于其所承载的信息，只有使数据在充分流动、共享、交换下才能形成所期望的集聚效应和规模效应。政府及公共机构在公共财政的支持下履职，有责任向社会提供公共产品，创造公共价值。因此，应在不涉及个人隐私安全及国家安全的情况下，向社会公开共享社会数据。

三是分级原则。

按照竞争性和排他性对数据进行不同级别的划分，可以将数据分为私有品、准公共品、公共品。

作为**私有品**的数据，其价值具有竞争性和排他性。竞争性是指当其他人使用该数据时，会减少这个数据的效益或增加生产数据的成本。如两个公司都在尝试获取一套数据，如果这套数据被其中一家公司获取，则可以帮助公司制定有效的策略，带来巨大的价值，而如果这套数据被两家企业所共享，那么这个数据将贬值。排他性是指某项数据具有阻止其他人使用该数据的性质。如企业为使得自身效益最大化，有组织他人使用对自身有利的数据的倾向。

作为**公共品**的数据，则具有非竞争性和非排他性。非竞争性是指当一个人消费或使用某样数据时，并不会减少其他人对这个数据产品的效益。非排他性是指一旦生产出来就不可能把某些人排除在外的数据。一般而言，国家统计局、财政局、税务局等政府官方网站所公开披露的数据属于公共品。一方面，当一个人登录政府官网查询相关数据时，并不会使得其他人使用该数据所得的效益减少；另一方面，当一个人根据需求对相关数据进行处理时，其他公民可以对相同的数据进行操作，不会受到妨碍。一般而言，社会数据具有较强的公共品属性。

作为**准公共品**的数据，具有非排他性和非竞争性两个特点中的一个，另一个不具备或不完全具备，即只具有有限的非竞争性或有限的非排他性，或者虽然两个特点都不具备但却有较大的外部收益产生。

一般而言，企业内部数据、收费数据库的无限使用次数的数据和有限保密的政府数据属于准公共品。对于仅在企业内部公开和共享的数据，具有非竞争性，即增加一人使用并不会减少其他人对该类数据的效益，和有限的非排他性，即仅对企业内部的员工是非排他的，对企业外部员工而言则是排他的。

2. 基于“三分原则”的数据确权路径

数据确权路径确认的标准有两个，一是由易到难，层层推进，二是对有助于实现社会和个人效益更大化的数据优先确权。

对于分类数据而言，个人数据、企业数据、社会数据的性质不同，可确认的产权种类不同，确权难易度也不同。例如，企业数据构成复杂，既有自身运营的原始数据，又有对外部数据添付后的衍生数据，其确权相对于个人数据及社会数据而言更加困难。因此，经过分类分级后的数据包含的数据产权束各不相同，包含的产权束内容越多，数据确权越复杂，越困难。

表7 分类分级数据产权内容

	个人数据	企业数据	社会数据
公共品	公有产权	公有产权	公有产权
准公共品	公有产权 基础数据产权	公有产权 基础数据产权 衍生数据产权	公有产权 基础数据产权 衍生数据产权
私有品	基础数据产权	基础数据产权 衍生数据产权	

资料来源：中国信息通信研究院

对分级数据来说，通过对比公共品、准公共品、私有品的社会和个人效益最大化，可确定确权优先顺序。具有竞争性的数据中，只有

同时具有排他性的私有品数据，可以通过市场机制将其配置给对其评价最高的人来使用，带来社会总福利最大化。同时，私有品数据产权的确认有利于实现最佳的隐私保护和隐私补偿，从这个意义上讲，私有品相对于具有竞争性的准公共品更应该优先确权。

具有非竞争性的数据中，共享再用是最大化数据价值创造的根本。非竞争性商品应该让更多的人接入和使用，因为额外使用带来的边际收益远高于零边际成本，更多的人使用和消费会带来更大的社会总价值。对具有非竞争性且排他的数据，会使数据仅被使用价值最大的那部分所使用，这排除了更多人重复再用同一数据要素的可能，封锁了数据倍增经济增长和实现最大价值的可能。因此先确认公共品的权属，比确认具有排他性的部分准公共品更加有利于实现社会效益的最大化。

综合以上两个标准，得到数据确权路径图。其中蓝色的深浅度表示数据确权的先后顺序，颜色较浅的区域，数据确权更加容易，社会效益和个人收益更大。数据确权可按照颜色由浅及深的顺序进行，即确权路径为 A→B→C→D→E→F。

资料来源：中国信息通信研究院

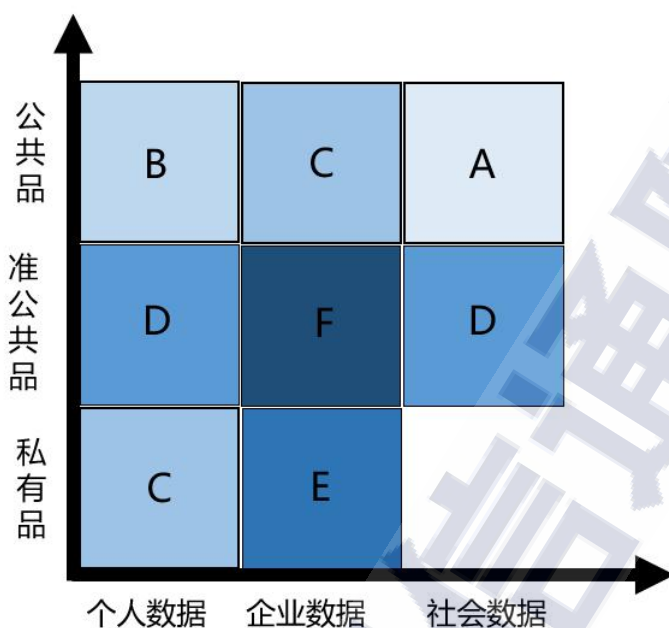


图7 数据确权路径图

（三）成熟数据要素市场的“四因素定价模型”

目前的数据要素市场尚处于初期萌芽阶段，为不完全市场状态，短期内的数据定价侧重于简单可操作，通过动静结合的数据定价策略吸引更多企业加入数据交易市场。未来，随着数据要素市场逐渐成熟，数据交易体量扩大，需稳步形成科学规范统一的数据定价交易体系。因此，成熟的数据要素市场中，数据定价要重视定价模型的设置。

数据具有类似金融资本和人力资本的双重属性，即数据具有金融资本的增值性、风险性，也有人力资本的异质性。故数据价格是其补偿价值、增值价值、异质性价值、风险溢价的集合。具体来看：

补偿价值即为成本，无论是原始数据还是加工数据都需要一定的成本，主要包括固定资产投资成本、无形资产投资成本和管理成本三类。固定资产投资成本指的是数据存储环境成本、相关IT装置成本、

其他必要设备等全部支出之和；无形资产投资是指软件开发费用以及预提的系统改进费用之和；管理成本包括软硬件设备的维护费用以及提供软硬件系统、内容服务、维护服务等所发生的人工费用。

数据的增值价值包含市场价值和社会价值两个方面。数据市场价值是通过经营和管理活动，把数据低投入转换成利润产出。如苹果公司通过 APP Store 建立用户和厂商的连接，运用数据分析，满足用户个性化需求，提供差异性服务，年创造收入在百亿美金。数据的社会价值则主要体现在规模效应上。数据不会随着使用增多而减少，反而可以多次循环使用，并且随着分享范围的扩大而提升价值。由于数据包含的信息密度很低，如果数据规模小、维度少，那么其对生产经营所能起的作用也很小。随着数字数据规模增大及维度增加，人们从数据中挖掘出的价值将呈现指数级的增长。因此，有必要对社会价值贡献主体通过市场方式进行补偿。

数据的异质性价值来源有两个：一是由于数据结构异质、搜集主体各不相同、价值高度依赖使用场景等带来的异质性价值。一个比特数据跟另外一个比特数据包含的生产价值通常是完全不同的，我们几乎无法用某一企业的数据量来衡量这个企业的价值或者进行横向比较。两个同样数据量的视频，一个可能是极有用的信息，另一个则可能是垃圾信息，这种情况在大数据中普遍存在。二是数据交易市场分割导致其具有异质性价值。当前统一的数据要素交易市场仍未形成，数据交易仍处于不规范、不标准的状态，会造成数据生产激励弱化，

导致数据流通渠道阻塞，使得同一数据产品在不同的交易市场上交易价值形成异质性的特点。

风险溢价是指数据投资组合的收益率与无风险收益率的差额，用来衡量数据资产投资者由于承担风险而获得的报酬率。数据价值具有较强的不确定性。这种不确定性主要有三方面原因，一是信息不对称性，数据供给方不清楚自身数据对不同数据使用者的价值，数据使用者在数据交易前也很难完全了解购买数据的实际价值。二是数据误差，误差可能是由于原始数据本来就不准确或是采用了粗粒度的数据集合并，也可能是原始数据是为了满足特殊应用目的或是经过处理缺失值或者数据集成而生成的。三是数据可挖掘性。部分数据由于时效性而导致数据折旧，无挖掘价值，价值下降，而部分数据由于不断挖掘分析，获得价值倍增。

从数据属性出发，我们构建了数据的“四因素定价模型”。数据价值是补偿价值和新增价值的和，补偿价值包含数据生命周期中所需固定资产投资成本、无形资产投资成本和管理成本。而新增价值代表企业数据投资利润，即数据给企业和社会带来的效益。即

$$P = V_i[1 + E(R_i)]。$$

其中， V_i 表示企业为数据资本*i*支付的成本； $E(R_i)$ 代表企业数据资本投资收益率。

$$V_i = \sum_{t=1}^n \frac{C_t}{(1+r)^t}$$

其中， C_t 表示企业在数据资本上的投入， r 为折现率。

然后构建数据资本投资收益率模型。其假设前提为：数据资本的

存在都是为了获取最大的收益；数据资本市场处于完全竞争状态，企业利润决定了数据资本的供需和流动方向；数据资本信息是充分的，其投资的期望收益是可以估计的；该模型中数据资本收益的影响因素是以 β_i 计量的数据投资系统性风险。数据资本投资收益率模型为：

$$E(R_i) = R_{fi} + \beta_i[E(R_m) - R_{fi}]$$

其中： $E(R_i)$ 为数据资本*i*的期望收益率， $E(R_m)$ 为企业数据资本平均收益率，这两个收益率为随机变量； R_{fi} 为数据资本存量的无风险收益率，不同数据资本的 R_{fi} 不同。将 β_i 定义为：

$$\beta_i = Cov(R_i, R_m) / Var(R_m)$$

β_i 系数反映数据资本收益率变动随数据系统风险变动的相关度。 β_i 系数大，意味着系统风险大，从而期望收益率也高；反之， β_i 系数小，系统风险也小，从而期望收益率也低。

此外在数据定价时需要加入异质价值收益率 α_i ，则数据资本的投资收益率模型为： $E(R_i) = \alpha_i + R_{fi} + \beta_i[E(R_m) - R_{fi}]$

至此，数据资本定价模型为：

$$P = \sum_{t=1}^n \frac{C_t}{(1+r)^t} \{1 + \alpha_i + R_{fi} + \beta_i[E(R_m) - R_{fi}]\}$$

四、数据价值化与数据要素市场发展建议

（一）完善数据要素资源体系

1. 加强数据资源采集汇聚

推进数据采集、标注、存储、传输、管理、应用等全生命周期价值管理，打通政府部门、公共机构、企业不同主体之间及不同主体内部的数据壁垒，实现传感、控制、管理、运营等多源数据一体化集成，构建全流程数据链。

构建不同主体的数据采集、共享机制，推动落实不同领域数据标注与管理应用。推动不同设备的数据采集和互通互操作，推广各层级统一的数据交换架构，制定关键设备数据接口标准。

2.推动数据高质量汇聚

引导政府部门、公共机构、企业开展数据资源编目工作，加强数据采集、标注、清洗和预处理，提升数据准确性、完整性、一致性，实现数据资源的可见、可管、可用。建设国家数据采集标注平台和数据资源平台，实现多源异构数据的融合和存储。建立数据质量管理机制，制定规范的数据质量评估监督、响应问责和流程改善方案，积极应用先进质量管理工具，形成数据质量管理闭环。

3.进行数据要素资源体系标准化建设

完善跨部门、跨行业的数据标准体系，建立多层级的数据管理标准。开展数据标准研制工作，建立包含数据基础术语标准、数据交换共享标准、数据安全隐私标准、数据行业应用标准等在内的标准化体系。加强国际数据治理的沟通与交流，推动建立数据资源国际标准。

（二）构建数据要素市场体系

1. 加快建立数据确权机制

建立数据确权基本框架。明确数据权利类型，确定数据权利主体，厘清数据的控制边界和使用范围。加快推动数据确权试点示范工程。建立全国数据统一登记确权体系，分级分类对原始数据、脱敏化数据、模型化数据和标准化数据的权属界定和流转进行动态管理。通过数据登记确权平台及其他手段，实现数据确权和价值变现，促进数据交易和流通。

2. 加快建立数据定价规则

明确数据定价规则，建立准确衡量数据价值和正确评估数据价值的方法，研究开发数据资产价值评估模型，建立完善评估工作机制，推动形成数据资产目录和资产地图，为数据交易提供价值评估和价格依据。

3. 加快建立数据交易市场化机制

研究制定数据流通交易规则，引导培育数据要素交易市场，依法合规交易数据，支持各类所有制企业参与数据要素交易平台建设。

搭建包括数据交易撮合、交易监管、资产定价、争议仲裁在内的全流程数据要素流动平台，建立全社会数据资源质量评估和信用评级体系。整合区块链等新一代信息技术，搭建全社会数据授权存证、数据溯源和数据完整性检测平台。

4. 积极营造便于数据要素流通的市场环境

营造包容、审慎、容错的环境，发掘优秀案例并进行试点示范，营造市场探索数据发展模式的氛围。简化数据市场准入机制和备案制度，降低数据领域新技术新业务和创新型企业的准入门槛。

以数据应用需求为指引，完善数据市场流通环境，精准对接市场需求，坚持多元协同共治原则，充分发挥政府和市场两类资源优势，营造健康可持续的数据市场环境。

（三）壮大数据要素应用体系

1. 推动数据要素全面深度应用

深化数据驱动的全流程应用，提升基于数据分析的工业、服务业、农业的供给与消费。实现不同产业的研发设计、生产制造、产品运维、经营管理的业务全流程综合应用，提升企业研产供销数据应用水平。

2. 开展重点行业应用试点示范

支持农业种植、畜牧、能源、航空航天、建筑、钢铁、化工、工程机械、消费电子、家电、纺织服装、食品追溯、零售、餐饮、酒店、文化、教育、旅游等重点行业企业探索各具特色的数据应用模式。结合重点行业应用示范，梳理遴选重点企业数据应用标杆，加大地方和行业企业对接和推广力度，复制推广典型应用。分行业梳理数据应用路径、模式方法和发展重点，编制数据应用指南，为数据应用提供指导方向。鼓励地方、单位在政策、资金、资源配套等方面加大支持力度，积极推动试点示范项目应用推广。

（四）建成数据要素安全体系

1. 推动数据安全监管体系建设

明确安全主体责任和防护要求，构建形成覆盖数据资源全产业链的安全监管体系。加快推进数据态势感知、测试评估、预警处置等保障能力建设。围绕数据全生命周期的安全保护要求，加快数据安全监测、加密传输、访问控制、数据脱敏等安全技术攻关。

2. 建立数据市场风险防控体系

建立面向企业的数据安全备案机制，提升数据安全事件应急解决能力。建立数据市场安全风险预警机制，提前应对数据带来的就业结构变动、隐私泄露、数据歧视等社会问题，严控数据资本市场风险。设立数据跨境流动风险防控机制，加强跨境数据流动监测和业务协同监管。强化关键领域数字基础设施安全保障，切实加大自主安全产品采购推广力度，保护专利、数字版权、商业秘密、隐私数据。

数据要素的未来未知远大于已知，我们仍在探索之中。

中国信息通信研究院 政策与经济研究所

地址：北京市海淀区花园北路 52 号

邮政编码：100191

联系电话：010-62302667

传真：010-62304980

网址：www.caict.ac.cn

