

数据要素白皮书

(2023 年)

中国信息通信研究院

2023年9月

版权声明

本白皮书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白皮书文字或者观点的，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

2022 年 12 月，中共中央、国务院印发《关于构建数据基础制度更好发挥数据要素作用的意见》（下称“数据二十条”），这是我国首部从生产要素高度系统部署数据要素价值释放的国家级专项政策文件。“数据二十条”确立了数据基础制度体系的“四梁八柱”，在数据要素发展进程中具有重大意义。

随着“数据二十条”等一系列政策措施相继出台，数据要素市场培育进展加速，畅通数据资源大循环的方向愈加明确。尤其是在人工智能快速迭代、大模型与大数据相得益彰的发展态势中，数据要素战略地位进一步凸显。各地方、各部门、各大企业纷纷加快数据要素领域布局，从体制机制、市场流通、产品研发、标准规范等多层次、多角度开展落地方案的深度探索，涌现出数据要素价值释放新热潮。

本白皮书在《数据要素白皮书（2022 年）》的基础上，进一步探讨数据要素理论认识，聚焦过去一年来数据要素探索过程中不断涌现的新模式、新业态、新热点，重点关注资源、主体、市场、技术四大方面的发展：资源方面，公共数据授权运营、企业数据资产入表、个人数据权益保护是各类数据发展的新趋势和着力点；主体方面，企业和政府逐步提升能力，双向发力推进数据要素可持续探索；市场方面，场外场内各有突破，相互促进、共同发展，推动数据资源的最优配置；技术方面，基于业务需求支撑数据要素发展的技术体系正加速创新。期望本白皮书的研究成果能为社会各界进一步参与数据要素实践探索提供有价值的参考。

目 录

一、数据要素再认识	1
（一）国家战略全方位布局数据要素发展	1
（二）人工智能发展对数据供给提出更高要求	3
（三）数据要素概念聚焦于数据价值释放	5
二、资源：分类推进数据要素探索已成为共识	7
（一）不同类别数据资源面临不同关键问题	7
（二）授权运营促进公共数据供给提质增效	11
（三）会计入表推动企业数据价值“显性化”	15
（四）权益保护仍是个人数据开发利用主线	18
三、主体：企业政府双向发力推进可持续探索	21
（一）企业侧：数据管理与应用能力是前提	21
（二）政府侧：建立公平高效的机制是关键	26
四、市场：场内外结合推动数据资源最优配置	29
（一）数据流通存在多层次多样化形态	30
（二）场外交易活跃，场内交易多点突破	33
（三）多措并举破除数据流通障碍	35
五、技术：基于业务需求加速创新与体系重构	37
（一）数据技术随业务要求不断演进	37
（二）数据要素时代新技术不断涌现	38
（三）数据要素技术体系重构加速	42
六、趋势与展望	42
参考文献	46

图 目 录

图 1 数据基础制度体系	2
图 2 公共数据、企业数据、个人数据的复杂关系	8
图 3 北京公共数据专区运营模式	13
图 4 数据要素分级授权体系示例	20
图 5 DataOps: 敏捷协同的一体化管理	24
图 6 地方数据相关条例出台情况	28
图 7 数据要素流通的多种形态	31
图 8 数据要素流通技术流程图	39
图 9 数据要素重构技术体系	42

一、数据要素再认识

数据的爆发式增长和规模化应用不断催生新产业、新业态，对生产力和生产关系的发展和变革具有重要影响。将数据增列为生产要素意味着对数据要素价值释放提出更高目标，需要通过深度研究、广泛实践，反复认知和领会数据要素战略布局、时代背景与理论内涵，不断推进数据要素发展，有力支撑数字中国建设。

（一）国家战略全方位布局数据要素发展

我国数据要素政策进入体系化构建阶段。自 2014 年大数据首次写入政府工作报告以来，在关于数据的系列政策布局推动下，数据与实体经济融合程度不断加深，数据技术、数据产业、数据应用、数据安全等方面都取得长足发展。2019 年，十九届四中全会首次将数据增列为生产要素，关于数据资源整合共享、开发利用、安全治理、市场化配置等方面的数据要素体系化顶层设计正式启动。四年来，《关于构建更加完善的数据要素市场化配置体制机制的意见》《“十四五”数字经济发展规划》《关于构建数据基础制度更好发挥数据要素作用的意见》《数字中国建设整体布局规划》等文件相继出台，数据要素政策体系架构初步形成，擘画出数据资源大循环、数据要素价值充分实现、全体人民共享数字经济发展红利的宏伟蓝图。

“数据二十条”为推动数据要素发展筑牢政策基础。习近平总书记指出，数据基础制度建设事关国家发展和安全大局，要维护国家数据安全，保护个人信息和商业秘密，促进数据高效流通使用、赋能实体经济，统筹推进数据产权、流通交易、收益分配、安全治理，加快

构建数据基础制度体系。2022 年 12 月，“数据二十条”的出台明确了数据基础制度体系基本架构（如图 1），提出建立保障权益、合规使用的数据产权制度，建立合规高效、场内外结合的数据要素流通和交易制度，建立体现效率、促进公平的数据要素收益分配制度，建立安全可控、弹性包容的数据要素治理制度。以“数据二十条”为指导，各地各部门将制定数据要素相关细则规定，围绕“数据二十条”不断丰富完善数据要素各方面制度体系和配套政策，打造“1+N”数据基础制度体系。



来源：国家发展和改革委员会

图 1 数据基础制度体系

数字中国建设引领数据要素价值释放方向。建设数字中国是数字时代推进中国式现代化的重要引擎，是构筑国家竞争新优势的有力支

撑。2023 年 2 月，《数字中国建设整体布局规划》指出，畅通数据资源大循环是数字中国建设的两大基础之一，要构建国家数据管理体制机制，健全各级数据统筹管理机构，推动公共数据汇聚利用，释放商业数据价值潜能。规划提出的“五位一体”总体布局为数据要素价值释放指引了方向，数字技术与经济、政治、文化、社会、生态文明建设的深度融合将带动数据要素在各场景发挥独特作用，从而充分激活数据要素内在价值，全面赋能经济社会发展。

数据要素统筹管理、协调发展的体制机制进一步完善。2022 年 7 月，国务院批准建立由国家发展改革委牵头，中央网信办、工业和信息化部等 20 个部委组成的数字经济发展部际联席会议制度，强化国家层面数字经济战略实施的统筹协调。2023 年 3 月，《党和国家机构改革方案》提出组建国家数据局，负责协调推进数据基础制度建设，统筹数据资源整合共享和开发利用，统筹推进数字中国、数字经济、数字社会规划和建设等工作。国家数据局的组建有利于破除“九龙治水”的数据治理环境，平衡数据要素安全发展的辩证关系，从而进一步推动数据要素的开发利用，推进多层次数据要素市场建设，促进数据要素、数字经济与实体经济的深度融合。

（二）人工智能发展对数据供给提出更高要求

2023 年，以 ChatGPT 等为代表的 AIGC 技术应用火遍全球，大模型技术取得的突破使人工智能技术发生了深刻的变革，而这个突破离不开高质量数据的发展。可以说，数据已成为未来人工智能竞争的关键要素，人工智能正在从“以模型为中心”加速向“以数据为中心”

转变。

人工智能发展驱动数据要素市场需求爆发。伴随着大模型时代的到来，通用人工智能（AGI）产业正迎来爆发期，更加需要大规模、高质量、多样化的数据集提升模型效果和泛化能力。大模型训练使用的数据集规模持续增长，例如根据公开资料显示，2018 年 GPT-1 数据集约 4.6GB，2020 年 GPT-3 数据集达到了 753GB，而 2021 年 Gopher 数据集已达 10550GB，2023 年 GPT-4 的数据量更是 GPT-3 的数十倍以上。

当前，主流大模型预训练数据主要来源于公开数据集、合作数据分享、大规模网络数据以及通过数据众包方式获取的数据。然而，我国人工智能领域高质量数据集缺乏、数据供给的产业生态不健全、企业数据资源获取成本高等问题依然严峻。一是国内人工智能领域高质量数据集缺乏。虽然我国已有部分中文开源数据集，但在数量上远远少于国际英文公开数据集，在数据质量方面参差不齐、部分内容十分陈旧。由于高质量数据集的缺乏，部分国产大模型采用“英文数据集+翻译软件”的方式生成中文语料库，导致训练结果出现巨大的文化冲突。二是人工智能领域数据供给的产业生态不健全。由于国内数据要素市场发展尚处于初级阶段，数据流通规则和数据供需对接机制未有效建立，目前国内尚未形成高效完整的人工智能数据产品供应链。三是企业数据资源获取成本高。在模型训练过程中，通常 80% 的工作是数据构建和准备高质量数据，人工智能企业需要花费大量的人力和物力进行数据集采集、清洗和标注，成本极高。同时，人工智能企业

通常难以获取行业高质量数据集，常陷入“寻数无门”的困境。

对此，各类主体通过数据要素市场积极应对上述问题。部分地方和行业推出一系列举措加强高质量数据供给，为大模型成长提供充足“养料”。例如，2023年5月印发的《深圳市加快推动人工智能高质量发展高水平应用行动方案(2023—2024年)》提出，“建立多模态公共数据集，打造高质量中文语料数据”。2023年8月，近50家单位成立“开放算料联盟”，围绕高质量中文训练数据和多模态训练数据，协调数据要素、数据治理、训练数据、数据标注等相关标准制定，协助数据交易所增加大模型相关的新品类和新专区。

此外，合成数据也成为模型训练中的重要类型，为数据要素市场带来了新需求。据专家预测，模型训练中必不可少的语言数据将于2030-2040年耗尽，其中能训练出更好性能的高质量语言数据将于2026年耗尽，而视觉数据恐将于2030-2060年耗尽。未来，合成数据将成为模型训练的关键数据。根据Gartner的预测，2024年用于训练大模型的数据中有60%将是合成数据，到2030年大模型使用的绝大部分数据都将由人工智能合成。这是否会对数据要素市场带来结构性的变化，还有待未来观察。

（三）数据要素概念聚焦于数据价值释放

数据要素概念的内核是提高生产效率与资源配置效率。生产要素是对某一时期经济发展中所需重要资源的科学抽象，是对生产过程中所投入成本的高度凝练。作为一种理论视角下的概念，从外延角度看，数据要素固然包括根据特定生产需求汇聚、整理、加工而成的计算机

数据及其衍生形态¹，但数据要素这一概念不只是对各行业各领域各类数据的指代，更是对数据所蕴藏巨大价值的强调。数据支撑业务贯通、推动数智决策、流通对外赋能的三次价值²是挖掘、释放数据要素价值的主要手段，而激活数据要素的根本目的是将数据以多样、创新的方式投入于经济社会发展全过程，通过数据开发利用增加生产经营活动的投入产出比，促进跨领域活动过程中资源的高效流动，从而全面提高生产效率与资源配置效率。

业界对于数据要素阶段划分有诸多探讨。为推动数据要素价值释放，可将具体过程进行分解。例如，按照供应链可分解为数据供给、数据流通、数据应用、数据安全等阶段；按照数据价值增值的阶段性目标，可分解为数据资源化、数据资产化、数据资本化或产品化等阶段。每个阶段均可进一步细分，例如，不同意义层次上的数据资产概念突出了数据要素价值释放的不同需求：在经济意义上，凡能产生价值的数字资源都可以用数据资产强调，从而提升组织对数据管理、数据应用的重视程度；在会计意义上，参照我国《企业会计准则——基本准则》第二十条“资产是指企业过去的交易或者事项形成的、由企业拥有或者控制的、预期会给企业带来经济利益的资源”定义，数据资产需具备组织拥有或控制、能够产生经济利益等性质，这对数据的确权、流通提出了要求；而在资产负债表列报意义上，参照我国《企业会计准则——基本准则》第二十二条“符合资产定义和资产确认条件的项目，应当列入资产负债表；符合资产定义、但不符合资产确认

¹ 《数据要素白皮书（2022 年）》

² 《数据要素白皮书（2022 年）》

条件的项目，不应当列入资产负债表”要求，数据资产若要入表，还应满足“有关经济利益很可能流入企业”“成本或价值能可靠计量”两大资产确认条件。总之，类似的分解有利于数据要素政策、目标等落地，各阶段的活动职能成为实现数据生产要素价值的必要环节。

二、资源：分类推进数据要素探索已成为共识

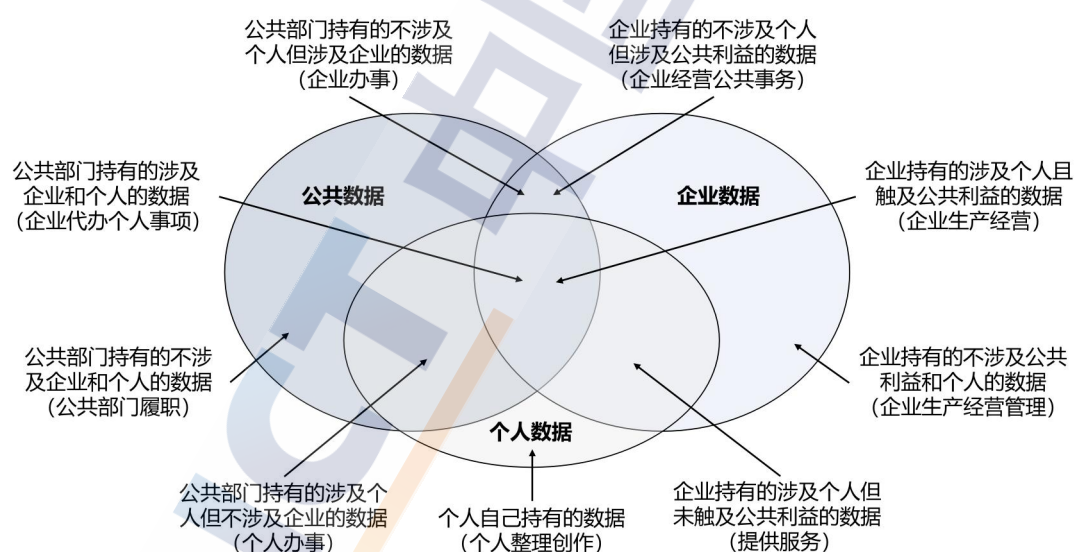
数据资源是释放数据要素价值的“原材料”。随着数字经济的迅猛发展，各类主体在生产、经营、管理、服务、消费等过程中产生大量数据。在推进数据开发利用、释放数据价值过程中，不同类型数据面临不同的重点任务与关键问题，分类推进数据要素探索已成为当前共识。

（一）不同类别数据资源面临不同关键问题

数据分类存在多种维度和多种方法，适用于不同场景。按照数据资源存储的维度，可分为基础层数据、中间层数据、应用层数据等，不同层次对数据的集成性、灵活性等要求不同。按照对数据资源加工程度的维度，可分为原始数据、衍生数据、数据产品等，数据加工者在其中的劳动和贡献存在差异。按照数据安全的维度，可分为一般数据、重要数据、核心数据等，通过数据分类分级降低数据安全风险。数据分类的维度和方法多种多样，反映出数据在存储、加工、应用、安全等过程和场景中的复杂性。

国外主要根据公共利益或个人权利划分数据类型。例如，美国《开放政府数据法案》等按照数据持有者类型的不同，将数据分为公共数据、非公共数据两类，以区分数据是否具备公共属性及相应的流通导

向与策略。美国《开放政府数据法案》确立公共数据可以被所有人公开使用、非公共数据需通过许可协议获得使用的原则，规定美国的所有政府部门都要向公众开放非敏感类的政府数据，从而加强公众、企业或其他组织对政府公开数据的利用。欧盟《通用数据保护条例》（GDPR）按照数据所描述对象的不同，将数据分为个人数据、非个人数据两类，凸显出因数据来源主体实际掌控数据能力不同，需采取差异化的数据相关权利措施。GDPR 在个人数据方面赋予用户知情权、被遗忘权、携带权等权利。针对非个人数据，欧盟制定《非个人数据在欧盟境内自由流动的框架》《数据法案》等推进非个人数据的自由流动与跨境传输。



来源：中国信息通信研究院

图 2 公共数据、企业数据、个人数据的复杂关系

我国“数据二十条”在总述探索数据产权结构性分置制度时，提出“建立公共数据、企业数据、个人数据的分类分级确权授权制度”，按照数据相关权益归属的不同，将数据分为公共数据、企业数据、个

人数据三大类型，突出不同类型的重点关切，探寻相应的突破方向。

从数据实际生成与持有角度看，三种类型的划分尽管存在复杂交叉（如图 2），但有利于根据不同类型数据的特性建立细化的分类标准与流通使用规范，也有助于建立不同主体的权责利动态调整机制，促进相关探索向“深水区”推进。

公共数据的概念与范围仍处于不断讨论和迭代中。“数据二十条”总体描述了公共数据是在各级党政机关、企事业单位依法履职或提供公共服务过程中产生的数据，但并未明确划定公共数据的范围和边界。当前，来自政务体系的数据（部分文件中称“政务数据”）和来自公共事业的数据（包括科研、教育、文化、供水、供电、公交等公共事业）由于受公共财政支持或在履行公共管理或服务职能中产生，归属于公共数据基本不存在争议。但是，另有一类数据是企业经营公共服务性质业务时收集、产生的关乎公共利益的数据，其是否属于公共数据仍存争议。在企业经营的复杂过程中，数据是否因提供公共服务而产生不易区分，**讨论该类数据时务须审慎判断企业所持有数据与公共服务、公共利益的关系。**逐步厘清公共数据的范围，有针对性地引导和促进公共数据开发利用，对提升公共治理与服务水平、发展壮大数字经济具有重要意义。对于来自政务体系的数据和来自公共事业的数据，一般认为，其归国家或全民所有，管理、开放等职责由政府或其他公共部门代为行使。本着“取之于民、用之于民、造福于民”的原则，在保障国家安全、商业秘密安全、个人信息安全的前提下，按用途加大供给使用范围将使公共数据价值更好回馈全社会。因此，如

何加大供给规模、推动供给提质增效成为公共数据发展的关键问题。

企业数据来源广泛，有丰富的生成方式。由于企业的生产、经营、管理链条较长，企业的数据需求多元复杂，对数据需要进行不同形式和程度的加工、整理与分析，这一过程中形成的多种衍生数据也丰富了企业数据的表现形式。按照企业数据生成方式的不同，可将企业数据划分为三种主要类型：一是企业自行采集、记录客观现象所得到的数据；二是企业在生产经营活动中，采集与用户的交互记录所得到的数据；三是企业基于已产生的数据，在赋予数据全新价值过程中所得到的数据。在不同类型企业数据的生成过程中，企业尽管付出的资源和劳动有所不同，但为数据赋予了独特的业务价值，需要得到有效地评估与认定。因此，**如何认定企业数据的业务贡献，促进数据价值“显性化”**成为企业数据的关键问题。

个人数据大多由公共部门和企业实际持有。对于描述或标识特定自然人信息的数据，如自然人的姓名、身份证号码等，其承载的信息具有一定的客观性，即这类信息不依赖于数据持有者所搭建的业务系统或应用软件，但数据持有者掌握这类信息后，有可能出现隐私泄露、滥用等风险。对于自然人与数据持有者交互产生的描述行为痕迹信息的数据，其所承载的信息对人来说可读性较弱，一般依赖于数据持有者设计或搭建的表格、系统和软件。数据持有者汇集大量个人痕迹数据后，经数据挖掘与分析可将数据价值不断放大，但也可能出现“大数据杀熟”等风险。此外，自然人创作的各类信息也形成了大量数据。个人创作的作品一般用于其他目的，而随着数据挖掘的深入，尤其是

人工智能大模型的迅猛发展，个人作品被数据持有者收集汇聚，成为特定场景中独具价值的数 据。但这一过程中，个人相关权益可能被侵犯。总体来看，由于个人对不同类型个人数据的掌控能力均有限，保障个人的隐私、人类道德伦理乃至人的主体性等已成为大众关注的焦点，如何在加强相关个人权益保护的基础上开发利用仍是个人数据的关键问题。

（二） 授权运营促进公共数据供给提质增效

各地各部门积极开展公共数据开放实践。近年来，各地方政府、部分行业主管部门以建立公共数据开放平台为抓手，积极推进公共数据开放，逐步完善公共数据开发利用体系。截至 2022 年 10 月，我国已上线 208 个省市公共数据开放平台³。基于开放平台，各级政府开放大量公共数据集，覆盖市监、工商、交通、生态、公共设施等多个领域，并且多个地区建立了开放数据需求反馈机制，以便满足社会对公共数据的应用需求。此外，行业主管部门也通过建设平台开放高质量数据。例如，中国气象局通过中国气象数据网开放气象数据，在 2023 年 2 月、7 月先后发布两批《基本气象数据开放共享目录》，共包含 106 项气象数据产品，在天气预报、农业指导、物流运输等场景发挥着极为重要的作用。

由于责任与激励机制缺乏，公共数据高质量供给受到制约。总体而言，我国公共数据仍然存在供给质量不佳、开发利用程度有限等问题。当前公共数据分布较为分散，同主题公共数据汇聚整理不便。公

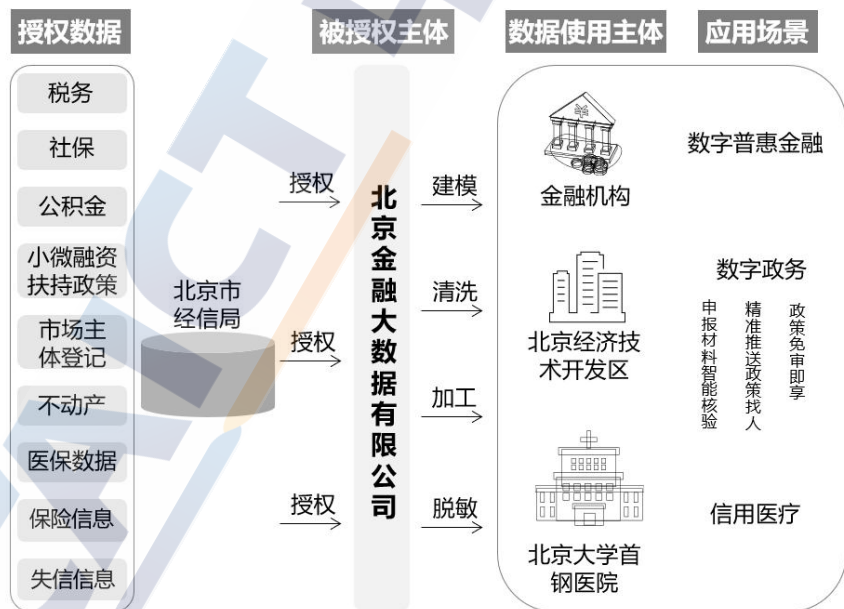
³ 数据来源：复旦大学数字与移动治理实验室，中国地方政府数据开放报告——省域（2022 年度）

共数据开放范围仍然不足，且开放平台数据更新频率整体较低。各地组织公共数据创新利用活动持续性不强，有效成果数量有限。主要原因在于，公共部门履职以“三定”职责作为核心依据，大部分公共部门承担数据安全风险，但对公共数据进行治理和开放的职责则不明确，推动公共数据开发利用的激励机制不完善。因此，公共数据开放对于公共部门来说没有明确的动力，特别是对于从事非信息化领域的部门来说，也往往不具备足够的力量。以上原因导致仍有大量高价值公共数据“深藏闺中”，高质量供给受到制约。

各地、各行业开始探索公共数据授权运营，引入社会化力量进行开发利用，促进公共数据供给提质增效。当前，授权运营在公共数据领域的实践探索中具有关键意义，可以充分发挥社会力量，解决公共数据开发利用中开放质量不佳、供需匹配不足、应用挖掘不够等问题，是大量高价值、高敏感的公共数据开发利用的核心路径。国内各地区、各行业积极开展授权运营实践。各个先进地区纷纷落实试点举措，在机制、平台等方面均取得成效。

在机制探索方面，各地方推进探索各具特色的授权运营机制，以释放公共数据价值为核心，逐步形成了几类不同的发展模式。一是**集中1对1模式**，以浙江、安徽、贵州、成都、青岛等地为代表。地方政府集中统一授权某一机构承担该区域平台建设、数据运营、产业培育等公共数据运营相关工作。一方面，集中授权具有权威性，有利于通过地域数据整合实现价值最大化；另一方面，集中授权存在市场效率不够高的问题，单一运营主体响应市场需求变化的速度可能有限。

二是分行业的 1 对 N 模式，以北京市的金融公共数据专区为代表（如图 3）。地方政府授权不同的行业属性机构，按照行业特点开展公共数据运营工作，专业性更强，有利于充分发挥行业属性作用，但统筹协调的难度也随之增大。2023 年 7 月，《北京市公共数据专区授权运营管理办法（征求意见稿）》发布，在金融场景的实践基础上推进专区制度体系建设，并深化交通、位置、空间、信用等各专区建设和应用。三是分散的 1 对 N 模式，以广东、上海、武汉等地为代表。地方政府根据不同数据与不同机构特点进行匹配，授权各类型市场主体分别开展公共数据运营工作。分散授权的灵活性更好，有利于发挥市场主体主观能动性作用，但一定程度上也容易出现混乱，对监管的要求更高。同时，由于地域数据未能完全整合，数据可发挥的价值有限。



来源：中国信息通信研究院

图 3 北京公共数据专区运营模式

在平台运营方面，部分地方通过建立统一的地域性公共数据运营平台探索公共数据产品或服务的应用与流通。一类是公共数据运营服务平台，例如贵州省的云上贵州平台，在明确授权统一平台对全省公共数据进行汇集、存储、共享、开放的基础上，依托全省资源打造数据产品及服务体系，与贵阳大数据交易所协同推动构建贵州大数据产业生态。成都市政府授权成都数据集团搭建公共数据运营服务平台，已于 2020 年上线。平台直接接入公共数据并作为供需对接桥梁，基于应用需求，经协调获取相应数源部门授权后，与应用方共同打造数据产品并接受数据使用监督。另一类是兼具公共数据加工与数据交易的综合服务平台，例如海南省的“数据产品超市”，以授权运营的思路搭建集数据归集、管理、加工、交易为一体的公共数据平台，将各类型参与主体纳入平台，由应用主体对公共数据进行加工增值后以数据产品的形式开放给市场。

此外，众多行业主管部门的公共数据也得到有效开发利用，通过授权运营的模式引入社会化力量推进数据价值的充分释放。相较于地方的公共数据，行业主管部门持有和控制的公共数据是纵向的数据归口，在行业领域内具有相对完整性和全面性，推进行业主管部门公共数据的应用与流通也是公共数据授权运营的关键工作之一。例如，司法数据作为来自各级司法体系履职和提供服务过程的公共数据，包含法条、立案、审判、裁判文书等多方面的司法相关数据，具有非常独特的应用价值。当前，司法数据由最高人民法院信息中心下属中国司法大数据研究院统一推进开发利用，应用主要集中在为政府及公共部

门提供决策支撑、为金融领域产品提供服务依据、为企业或个人主体提供司法大数据服务等场景。此外，人力资源和社会保障部将社保数据授权金保信社保卡科技有限公司运营，人民日报社将人民日报历史数据授权人民网科技（北京）有限公司运营，中国气象局将气象数据授权北京天译科技有限公司等相关机构运营，这些均是行业公共数据授权运营的典型案例。

（三） 会计入表推动企业数据价值“显性化”

近年来，虽然“将企业数据资源视为资产进行管理”逐渐成为众多企业的共识，但这里的“资产”往往只是经济意义上的资产，体现了企业数字化过程中数据资源的重要性。而对于数据能否真正成为会计意义上的资产、能否计入企业资产负债表，始终是各界讨论的焦点。今年8月21日，财政部正式发布《企业数据资源相关会计处理暂行规定》（以下简称《暂行规定》），并规定自2024年1月1日起施行。《暂行规定》的出台肯定了数据资源可具有资产属性，是数据要素市场发展的重要里程碑。作为报表意义上的“资产”，企业数据的市场价值与业务贡献将在财务报表中得以“显性化”。

《暂行规定》明确了企业数据资源在财务报表中进行会计确认和计量的思路。出台《暂行规定》的主要目的是，为消除相关企业实务中对“数据资源能否作为会计上的资产确认”“作为哪类资产‘入表’”等疑虑提供指引。因此，《暂行规定》在不改变现行企业会计准则的基础上，给出了将数据确认为无形资产、确认为存货和不确认资产三条入表思路。具体来说，《暂行规定》规定：企业使用的数据资源，

符合《企业会计准则第 6 号——无形资产》（财会〔2006〕3 号）规定的定义和确认条件的，应当确认为无形资产；企业日常活动中持有、最终目的用于出售的数据资源，符合《企业会计准则第 1 号——存货》（财会〔2006〕3 号）规定的定义和确认条件的，应当确认为存货；企业出售未确认为资产的数据资源，应当按照收入准则等规定确认相关收入。

企业数据资源“入表”有利于内外部发现并确认企业数据的价值。

在将数据确认为无形资产或存货进行会计计量的过程中，企业需要明确取得和持有数据资源过程中产生的各项成本或增减值，企业数据的产生成本或应用价值将得到更加完整、准确的列示，基于企业数据资产产生的收入进一步明确。继而，对于从事数据产品（或数据服务）供应的企业来说，有助于企业确定更为市场所认可的数据产品定价，促进企业参与市场化的数据交易流通；对于数据密集型企业来说，有助于投资者更好发现和理解企业整体的价值，提升企业的融资机会，扩大企业的发展空间。同时，针对企业如何贯彻实施《暂行规定》，财政部会计司表示，企业应当主动按照相关披露要求，持续加强对数据资源的应用场景或业务模式、原始数据类型来源、加工维护和安全保护情况、涉及的重大交易事项、相关权利失效和受限等相关信息的自愿披露，以全面地反映数据资源对企业财务状况、经营成果等的影响。因此，在实施“入表”的过程中，通过对数据资源相关情况的盘点和披露，企业将更加了解、重视和明确数据资源的分布、特点和应用价值，有利于企业进一步挖掘数据应用，释放数据要素价值。

需要注意的是，《暂行规定》还不是真正意义上的“数据资产会计准则”。《暂行规定》目前只给出了“满足资产确认条件且价值确定的数据资源如何计入报表”的解决思路，但是没有解决“数据价值如何确定”的问题。由于数据具有未来经济利益难确定、经济寿命不确定、价值易变性等区别于传统要素的特性，成本法、收益法、市场法等估值方法均有其局限性。一些企业探索综合三种方法，基于收益法建立数据资产估值体系，通过评价数据资产业务经济贡献值指导形成数据资产价值基准。例如，浦发银行提出的数据资产价值评估框架，围绕数据资产潜能预测、效能评估和收益测算，将数据资产价值分为内在价值、成本价值、业务价值、经济价值以及市场价值五类，并给出了每类价值的价值因子及计算公式。2023年9月，中国资产评估协会印发《数据资产评估指导意见》，对资产评估机构按照成本法、收益法和市场法开展数据资产评估提供了进一步指导。然而，企业数据估值问题是个复杂的难题，当前全社会仍未完全取得共识，还需要企业、专业研究机构、资产评估机构、会计和审计事务所等进一步探讨。

总的来看，《暂行规定》的出台迈出了我国为发展数据要素而建立相关会计核算制度的第一步。《暂行规定》是在现行企业会计准则体系下的细化规范，在会计确认计量方面与现行无形资产、存货、收入等相关准则是一致的，不属于国家统一的会计制度要求变更型会计政策。现阶段，《暂行规定》只是将既有的会计处理规则迁移到“符合相应条件的数据资产”上，而并非从数据资产的特点出发来设计针

对性的会计规则，也就是说，数据资产暂时将按照与传统资产同样的方法进行会计计量。从形式上看，只是在企业资产负债表的“存货”“无形资产”和“开发支出”三个报表项目下分别设置了“数据资源”子项目；对应的企业核算时将会在相应总账会计科目下面设置“数据资源”二级科目。但是，只按照这样的会计方式，数据要素区别于传统要素的特殊性或将无法从报表中得到充分体现，无法被适应。“数据入表”目前还只是一个试点的过程，真正建立适应数字经济发展需求、符合数据要素特殊属性的会计核算方法，未来仍将有很多复杂、挑战性的工作有待完成。下一步，如何与国际会计准则相协调，如何建立相适应的审计制度，如何厘清企业数据资产与信息资产的划分，企业是否会通过数据资产实现增值，银行是否会普遍接受数据资产作为质押等，这些问题仍然需要时间的检验。

（四）权益保护仍是个人数据开发利用主线

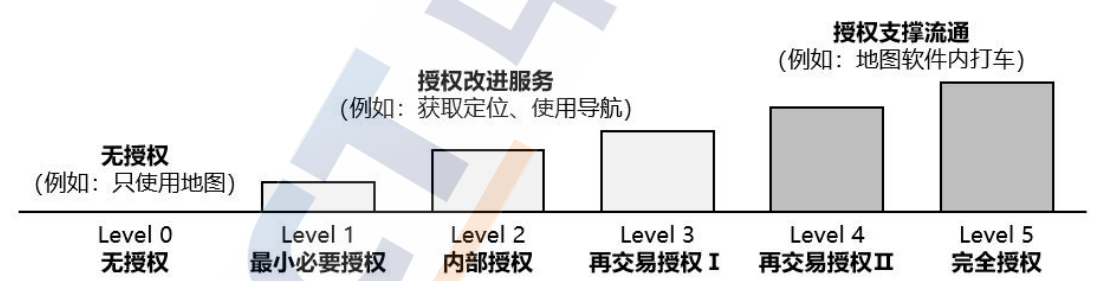
个人数据方面，《个人信息保护法》针对个人信息数据确立了保护原则。《个人信息保护法》是中国首部规范个人信息处理活动的专门立法，旨在保障公民个人信息安全，遏制个人信息泄露、滥用等现象的发生。该法规定了个人信息处理全过程中应遵循的原则和要求，明确了相关主体的权利和义务，健全了个人信息保护工作的体制机制。该法既通过个人信息处理规则、敏感个人信息处理规则、个人信息对外提供规则、单独同意规则等保障个人权益，又通过匿名化处理、个人信息处理者不需取得个人同意可以处理个人信息的若干场景等方面的规定保证个人信息数据可得到合理的开发利用。国际横向比较看，

我国《个人信息保护法》在规则严厉程度上基本对标欧盟 GDPR，两部法律文本均赋予个人全面、细致的权利，均对信息处理者施加了较高的合规义务。同时，我国关于个人信息违法犯罪相关罪名适用主体广泛，入罪门槛较低。因此，我国已建立起较为严格的个人信息数据保护原则。

个人对个人数据掌控能力有限，权益保护落地实践未达预期。由于个人数据主要由公共部门和企业实际掌控，个人数据相关的权利在具体场景和问题中又较为复杂，因而世界各国在操作层面都还未能完全适配个人信息、个人数据保护的 legal 要求，个人信息泄露时有发生、个人数据越权滥用仍然存在。例如，国外许多网民发现自己在网上分享的照片成为 Flickr 数据集的素材，自己此前既不知情，又很难将自己创作的照片数据从数据集中删除。2020 年 7 月以来，工业和信息化部开展纵深推进 APP 侵害用户权益专项整治行动。截至 2023 年 7 月，工信部已发布关于侵害用户权益行为的 APP（SDK）通报 30 批。通报并责令整改的 APP、SDK 中，违规收集个人信息、强制频繁过度索取权限、违规使用个人信息等问题仍然突出。

通过分级授权细化个人数据采集使用规范已有理论探索。个人数据开发利用所需的数据多样、场景多变，个人理应可以根据场景和需求，分级、分步骤进行个人数据的授权，也可以按照意愿撤回相应级别、步骤的授权。个人数据处理者则应严格遵守个人对数据开发利用场景、范围等要求，保护个人数据权益。在《个人信息保护法》施行前，大量收集、处理个人数据的过程属于“一揽子授权”，个人在使

用服务过程中可能需要让渡许多与服务无关的数据。《个人信息保护法》规定了数据收集与处理的最小必要原则和单独同意原则，支付宝、微信、美团等平台逐步增加了新业务首次使用单独授权同意、广告权限管理、个性化推荐权限管理、第三方授权管理、清除历史行为等功能，实现了初步的个人数据分级、分步骤授权模式。“数据二十条”再次强调数据处理者应按照个人授权范围依法依规采集、持有、托管和使用数据，不得采取“一揽子授权”、强制同意等方式过度收集个人信息。学界和业界正在探索更加细化、完备的数据要素分级授权体系（如图4），以更简洁清晰的方式划分数据开发利用场景，用户既可以选择授权必要数据来使用基础服务，又可以授权更多数据享受改进服务乃至支撑数据流通，从而以较低成本实现个人对数据处理范围的控制。



来源：《数据要素论》

图 4 数据要素分级授权体系示例

通过专业数据托管服务机构降低个人数据行权门槛已有国际实践。“数据二十条”提出探索由受托者代表个人利益，监督市场主体对个人信息数据进行采集、加工、使用的机制。个人数据管理机构作为受托人，可以根据用户的授权委托，代理大量用户进行个人数据集

中存储与管理应用，在遵循相关法规和用户意愿、保护数据隐私和安全的基础上，以忠诚义务和专业水平帮助个人行使个人数据权利。通过受托人提供的平台，个人可以集中化管理自己的数据，自主决定每个机构、平台对自己数据的收集、使用和共享情况，自行携带、转移自己的数据。因此，个人数据统一托管可向个人赋予更多控制权，有效降低个人行权门槛。个人数据委托在国际范围内仍属于探索早期。韩国 MyData 服务于 2021 年开始落地，MyData 运营商通过开展个人数据汇总整合工作，向个人提供一站式查询和管理服务，但这种模式的可持续性仍有待观察。

三、主体：企业政府双向发力推进可持续探索

企业和政府构成推进数据要素发展的核心力量。其中，企业是冲锋在前的创新主体，政府则主要发挥有序引导和规范发展的作用。在数据要素市场建设过程中，企业和政府需着力提高自身能力，在扮演好各自角色的基础上守正创新、双向发力，共同推进数据要素发展的可持续探索。

（一）企业侧：数据管理与应用能力是前提

企业作为数据要素的创新主体，在丰富的生产经营活动中，积累了大量宝贵数据有待挖掘使用，同时也是数据要素市场的主要需求方。结合业务发展目标构建数据管理体系、在业务场景中深度应用数据是企业整体数据能力提升的必要环节，构成企业数据价值释放的基础。数据管理规范了数据采集、加工、使用过程，是企业丰富数据应用、参与数据要素流通的前序基础。数据应用将数据真正转化为生产力，

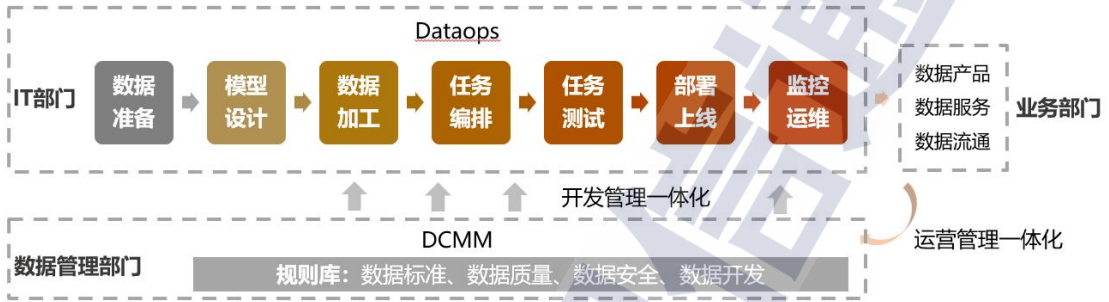
是释放数据要素价值的最后一公里。在不断夯实自身数据能力的基础上,企业才能进一步探索参与数据要素市场、构建核心竞争力的方式。

企业不断通过 DCMM 贯标评估提升数据管理能力。自 2020 年国家标准《数据管理能力成熟度评估模型》（简称 DCMM）贯标评估工作启动以来,随着全国范围内的持续宣贯,截至目前已有十余个行业的千余家企业参与到了 DCMM 评估中,显著提升了全社会数据管理意识,使得贯标企业数据管理能力得到明显增强。以评估贯标为契机,企业在战略规划、组织架构、技术工具等方面推动开展数据管理相关工作:在战略规划方面,通过建立全景数据管理工作视图,确立数据管理中长期目标和管理活动优先级,统筹、协调各层面工作,明确所需资源投入总量和分配机制,监督、评价和优化企业数据管理工作的执行。在组织架构方面,企业通过成立专职团队增强数据管理执行效率,成立统一数据管理归口部门,普遍形成了决策、管理、执行的数据管理三层架构组织,并设置专门的数据管理管理岗位,建立了数据部门、技术部门、业务部门协同机制。在技术工具方面,企业通过构建统一的数据管理技术平台,将各分散的单一功能型技术工具进行集成,消除数据管理协同难点,提高数据管理效率。例如,中国联通集团以《中国联通集团数字化转型行动计划(2022-2025 年)》为战略纲领,明确了数据管理工作的总体目标、发展路径及保障体系,形成以数字化转型领导小组为领导,数字化转型推进办公室统筹落实推进,总部数据治理组、省分数据治理组、地市数据治理组纵向联结协同的数据管理组织架构,依托联通云底座,打造“湖仓一体、批流

融合”的数据中台，实现全域数据统一采集加工，强化数据管理能力。同时也应看到，DCMM 标准聚焦在数据管理体系的建设，它的贯彻落实能帮助企业建立数据能力的基座，但无法直接解决企业数据应用场景不丰富、业务价值赋能不充分的问题。为此，需要在贯彻 DCMM 标准的基础上，结合数据资产价值评估、会计入表、交易流通等实践，以价值为导向，推动数据资产运营能力建设，建立常态化的数据资产运营体系，为数据价值持续释放提供坚实保障。

数据研发运营一体化（DataOps）在数据管理的基础上，解决开发管理“两张皮”的问题，有效提升数据管理与应用的协作效率。随着数据应用场景日益丰富，数据分析需求快速变化，数据加工链路日益复杂，数据工程师、数据管理员、报表开发人员、运维工程师等各类数据管理角色逐渐增多，导致企业数据交付任务难度大大增加。DataOps 倡导协同式、敏捷式的数据管理，建立清晰通畅的数据管道，明确数据管理的流转过程及环节，采用先进的数字化技术推动数据管理自动化，缩短数据项目的周期，提高各数据管理相关人员的数据检索、获取和应用效率，并持续改进数据质量，降低管理成本，加速数据价值释放（如图 5）。例如，通过标准设计、模型设计指导数据开发，前置化数据质量管理，并建立 SLA 开展数据资产运维，实现开发与管理的协同；数据管理成果通过被业务分析人员、数据科学家等角色自助使用，支撑业务运营，同时运营结果反向指导数据管理工作，实现管理与运营的协同。中国信息通信研究院与多行业头部企业于 2022 年共同成立 DataOps 标准工作组，发布能力框架模型和实践指

南，致力于将敏捷、精益等理念融入数据开发过程，打破协作壁垒，构建集开发、治理、运营于一体的自动化数据流水线。目前，国内已有超百家机构开展 DataOps 相关实践，我国企业对于 DataOps 已从概念启蒙引入阶段演进至规模化落地阶段，各行业高质量、高价值数据的汇聚、融合与应用工作也得以大大加速。



来源：中国信息通信研究院

图 5 DataOps: 敏捷协同的一体化管理

在管理数据的同时，企业正在同步推进业务需求梳理与挖掘，数据应用场景持续扩充。早期互联网、金融等数据密集型企业为提高营销、风控等业务能力，通过对自身数据的挖掘和分析，洞察用户需求、优化运营、提供个性化服务，从而提升自身竞争力和用户满意度。例如，电子商务企业可以根据用户的购买历史和浏览行为，推荐符合其兴趣和喜好的产品；普惠金融机构可以根据用户的资质、历史行为等数据评价信用分数与风险等级，实现对个人、小微企业等主体的精准放贷。随着数字经济的发展，数实融合成为产业发展与经济增长的新动力。数字化水平较高的头部企业进一步挖掘业务链条中的数据需求，通过自有数据的挖掘和分析，为生态伙伴提供更加科学化的决策依据，以实现其生产、流通等环节的高效运作。例如，头部电商自营

品牌京东京造，利用其丰富的用户数据和数据赋能的强大供应链管理
能力，基于“你做工厂，我做市场”的分工定位与品牌方深度合作，
2022 年总销售额同比增长 60%、100 个以上品类年均销售额增长超过
300%、新品开发成功率超过 90%、帮助工厂平均降低 30 天库存周转。

传统企业愈加重视数据要素的沉淀、挖掘和应用能力提升。随着
信息技术的持续发展和普及，传统企业也越来越重视数据与生产、调
度、调控、营销等业务的融合应用，以实现设备级、工厂级和企业级
的互联互通，从而提高生产效率、降低成本、改善产品质量和创新能
力。例如，在制造业中，工业互联网可以实现设备的远程监控和故障
诊断，提高生产线的稳定性和可靠性；可以通过数据分析和预测技术
进行生产计划和库存管理，实现定制化生产和个性化营销；可以通过
供应链的数字化和智能化，实现物流协同和合作，降低成本。

**在夯实数据能力的基础上，企业积极挖掘并输出自身数据能力特
色，通过提供高质量数据服务创新数据要素市场参与方式。**数据要素
供应链条长，在供给、流通、应用、安全等环节均有细分的市场需求。
**对于专业数据能力突出的中小企业，可借助集聚优势，寻找市场定位，
塑造服务品牌。**例如，人工智能的发展催生了数据标注、清洗等基础
数据服务需求。山西省大同市积极布局基础数据服务产业，并于 2022
年 11 月印发《大同市促进数据呼叫（标注）产业发展的若干政策》，
明确了坐席、场地等补贴标准。上海润迅、重庆人和、华顺金服等
20 多家基础数据服务企业先后落地大同，利用大同市人力资源、场
地、电力等成本优势，不断强化在数据要素市场中的竞争力。**对于数**

据资源富集、数据综合能力强的大型企业，除供给数据资源外，利用数据能力提供综合性数据产品与解决方案也是启动数据要素市场的有效手段。在数据要素市场发展的初级阶段，利用自身数据积累优势和服务能力优势，深入调研数据需求方的核心业务痛点，提供针对性的数据产品与解决方案，进而形成可复制推广的数据产品与能力架构，可有效实现数据流通所带来的价值倍增。例如，国家电网向银行出售电力大数据金融风控等服务，打造“电力数据看”新模式，拓展电力数据应用场景。中国移动打造了生态开放、技术先进、数据全面、安全可控的“梧桐”大数据品牌，通过梧桐大数据门户实现“产品订购入口、生态合作入口、技术分享入口”的统一，全面对外输出中国移动的数据产品与能力。招商银行整合发布“企业数智金融”品牌服务，突出从产品视角向客户视角的转变，从客户的销售场景、采购场景和员工服务中研究客户的需求和痛点，提供整体的数智化解决方案。

（二）政府侧：建立公平高效的机制是关键

政府及其他党政机关作为国家治理的核心力量，在数据要素发展过程中扮演着有序引导和规范发展的关键角色。除向社会供给高价值公共数据外，提升治理能力，创新治理手段，建立公平高效的数据要素发展促进机制，守护数据要素可持续发展环境是更好发挥政府在数据要素领域作用的关键。

面对数据要素这一新兴事物，各地方在中央宏观战略指引下，加紧制定数据条例与行动方案，更加灵活、主动地探索推进数据要素发展的实施规范乃至细化的操作流程。自 2015 年实施国家大数据战略

以来，我国已有 22 个省级行政区、4 个副省级市出台数据相关条例共 30 份（如图 6）。作为地方人民代表大会或其常务委员会通过的法律文件，这些条例为地方数据要素的发展提供了法律基础。截至 2023 年 8 月，全国已有 27 个省（自治区、直辖市）设置了专门的省级大数据管理机构。省级大数据专职管理机构的不断扩充，有效推动数据资源整合和开发利用效率与效益提升。此外，部分地区发布发展规划、行动方案等政策文件，将数据要素的产业集聚、流通交易、数据驱动的经济高质量发展等作为重点，致力于营造公平高效的数据要素发展环境。例如，上海市《张江数据要素产业集聚区建设三年行动方案（2023-2025 年）》《立足数字经济新赛道推动数据要素产业创新发展行动方案（2023-2025 年）》等文件谋划了上海数据要素发展的体系性布局，在维护数据安全、强化数据安全产业的基础上，基于要素融合和生态构建，着力打造高水平的数据要素产业集聚区。深圳市发布的《深圳市数据交易管理暂行办法》《深圳市数据商和数据流通交易第三方服务机构管理暂行办法》规范了数据交易范围和市场主体行为，并成立数据交易监管机制专责小组，率先以制度的形式明确建立跨部门协同监管的机制。先行地区的制度与机制体现了数据要素领域有效市场和有为政府相结合的创新成果，对中央和其他地方引导和推进数据要素发展均有重要的借鉴意义。

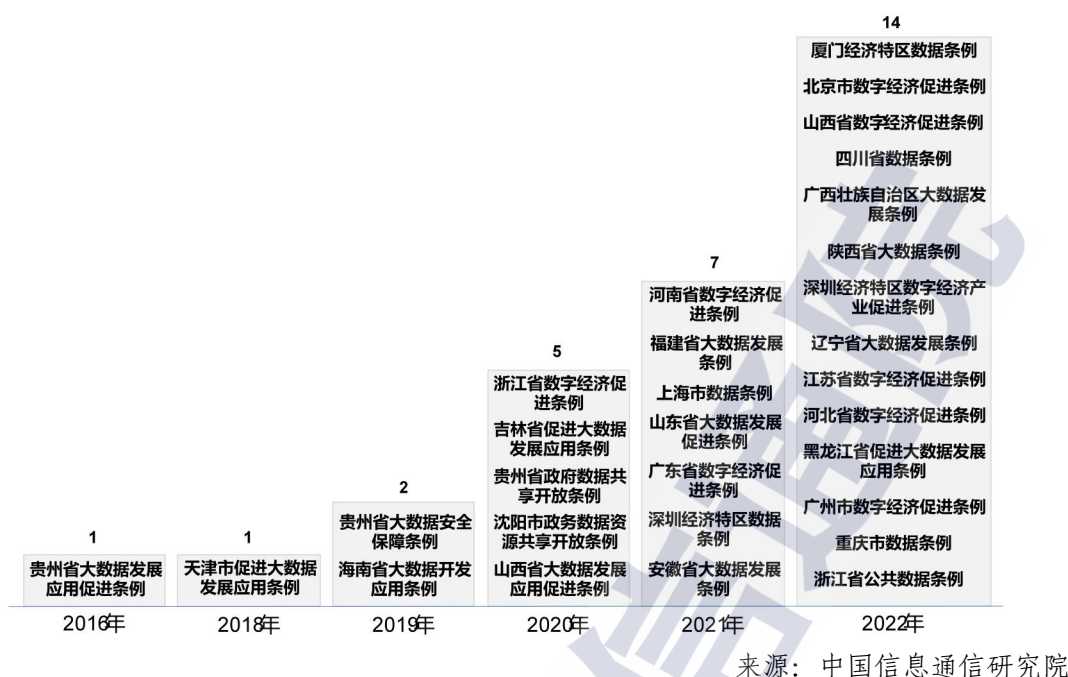


图 6 地方数据相关条例出台情况

加快完善数据领域央地协同和条块协同机制是当前的重要命题。

在地方纷纷成立大数据管理机构承担数据领域统筹协调与资源整合共享职能的背景下，国家数据局的成立顺应了各界对国家层面组建专职管理部门的期待。然而，国家数据局尚处于初建阶段，如何理顺国家和省级数据管理机构关系、平衡国家各部委及地方各委办局间职能，需要尽快探索方案并落地执行。一方面，各地大数据管理机构的单位性质和隶属关系并不相同，主要可分为省政府直属机构、省政府办公厅管理机构、省行业主管部门管理机构三类，且多有承担推进数字政府建设、数据安全治理、大数据产业管理等当前国家数据局所规划职责之外的工作。国家和省级机构履行职能过程中仍会存在“一对多、多对一”等多头管理问题。加快理顺管理关系，完善央地协同机制，是推进构建数据要素可持续发展环境的必要前提。另一方面，国家和地方数据管理机构与其他行业主管部门各有职责划分，既包括产业发

展、资产管理等业务领域的区分，又包括人社、交通等垂直行业的区分。加快理顺国家各部委、地方各委办局间职能，完善条块协同机制，促进数据要素跨部门、跨行业、跨地域互联互通与开发利用，是促进数据要素公平高效发展的重要保障。

提升治理能力、营造兼顾效率与公平的数据要素发展环境仍应成为政府工作重点。纵观各地数据要素相关政策制度，尽管产业格局、发展目标、数据供给、流通、应用等重点发力方向等基本明确，但落地抓手相对有限，还不能完全适应数据要素的新特性和新趋势。“数据二十条”提出，要充分发挥政府有序引导和规范发展的作用，守住安全底线，明确监管红线，打造安全可信、包容创新、公平开放、监管有效的数据要素市场环境。守住数据安全底线、推进数据要素可持续探索、促进数据价值释放应成为各地治理的共识。一方面，要提升数据安全纵深防护与综合防御能力，健全数据要素市场监督管理体系，在明确监管对象、范围、标准、程序等规则基础上，加强合规、违约、欺诈等重点领域的执法司法，维护数据要素市场的公平环境。另一方面，要适应数据要素新特性，不断创新治理手段，通过建立市场准入第三方评估机制、负面清单动态调整机制、合同约定与争议仲裁机制等，提高市场主体合作互信程度和自发探索的活跃程度，避免对市场主体的不必要干预，为数据要素市场的高效实践留足空间。

四、市场：场内外结合推动数据资源最优配置

数据要素市场是实现数据要素价值第三次飞跃的关键。数据在市场中流通使数据流向更需要的地方，让不同来源的优质数据在新的业

务需求和场景中汇聚融合，实现双赢、多赢的价值利用⁴。在此基础上，不同的数据流通形态串联起各类主体，推动场内外数据要素市场活跃探索，引导数据要素在供需关系与价格机制的作用下实现最优配置，创造更大的经济效益。

（一）数据流通存在多层次多样化形态

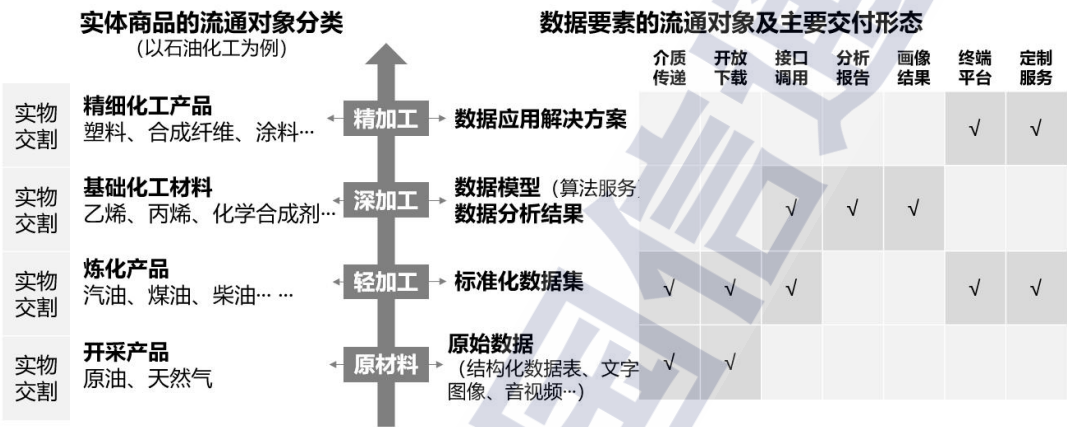
推进数据要素充分流通，首先应正确认识数据流通的表现形式，从不同角度分析数据流通的类型，包含参与主体的供需关系、流通对象的形态和交付方式等。

从流通主体间的供需关系来看，数据流通存在开放、共享、交易三种供需模式，区别在于数据需求方获取数据所需要支付的对价关系不同。数据开放是指数据提供方无偿提供数据，需求方无需支付对价的数据单向流通形式。由于数据提供方无法通过数据开放直接获得收益，因此数据开放的对象往往是公共数据。数据共享是指参与主体互为数据供需方，不强调货币媒介参与的数据双向流通形式，但共享过程往往涉及复杂的相互博弈，相较于一对一的两方共享，政府间或行业间的多方数据共享更容易形成规模和持续开展。数据交易则是指数据提供方有偿提供数据，需求方通过货币等形式支付对价的数据单向流通形式。相较于数据开放和共享，数据交易更容易激发市场参与主体积极性，成为数据要素市场化流通的主要形式。

此外，近两年来关于“数据运营”“数据信托”等新概念、新探索的讨论十分火热。我们认为，站在供需双方流通闭环的角度，这些

⁴ 《数据要素白皮书（2022 年）》

新模式并未改变双方的供需关系，不属于与开放、共享、交易并列的全新流通模式，而是在供需对接过程中引入了新主体，由数据经纪商、数据运营服务商和数据托管服务商等各类数商提供了中介性功能，在促进数据资源供给、降低数据获取成本、消除供需方之间资源错配等方面发挥了重要作用。



来源：中国信息通信研究院

图 7 数据要素流通的多种形态

从流通对象来看，流通中的数据包括从原料到衍生服务的多种形态（如图 7）。就产品形态而言，流通中的数据产品可以参照实体商品的加工程度来进行区分。类比石油化工领域，原油、天然气等是一经开采即可得到的原材料，在此基础上进行不同程度的炼化和加工得到不同类型的下游消费产品，一类如汽油、煤油等成品油，是对原材料进行炼化得到的轻加工产品；一类如乙烯、丙烯等基础化工材料是炼化后经裂解得到的深加工产品；一类如塑料、合成纤维等精细化工材料是对基础化工材料进一步加工合成得到的精加工产品。相应地，在数据流通领域，原始数据、标准化数据集、数据模型或分析结果、数据应用解决方案可大致对应不同加工程度得到的数据产品和服务。

就交付形式而言，为满足不同的应用需求和安全保护要求，数据产品的流通也呈现多样化的交付形式。与实体商品不同，数据产品的流通并非简单的转移交割。在原材料层级，原始数据主要通过介质传递、开放下载等方式直接复制或转移，但是，出于对数据可控性和可用性的考量，实践中以原始数据形态进行的数据流通极少。在轻加工层级，标准化数据集主要通过介质传递、开放下载、接口调用、终端平台、定制服务等方式交付，其中介质传递或开放下载多适用于内容固定的静态数据集，接口调用或终端平台多适用于随时间或需求不断更新的动态数据集，此时流通的数据仍主要存储在数据提供方，但使用方可以按需使用数据。在深加工层级，数据模型或分析结果可以通过接口调用、分析报告、画像等形态交付，数据提供方主要交付数据加工分析的结果，基于隐私计算的数据核验、数据查询、联合建模等也属于这一类。在精加工层级，数据应用解决方案主要通过终端平台和定制服务的形式交付，针对不同应用方的个性化需求提供定制化的数据产品。

综合以上分析可见，数据流通的对象往往不是标准化、同质化的，数据流通的表现形式也不是简单的数据产品买卖或转移。事实上，数据流通并非数据上升为生产要素后的新生事物。长期以来，已有大量的数据流通活动深度融合在政府和企业日常运转中，只是并未被标记为一种单独的业务形态。例如，征信公司从不同金融机构收集同一客户的金融行为数据，汇聚整合后形成对客户信用情况的评判，出具征信报告，虽然这被称为“征信服务”，但其背后是大量数据在不同主

体之间被传递和使用。同理，互联网广告、资讯服务和企业供应链管理等营销、风控领域的业务运转均依赖于海量数据的流转、整合与应用，也存在大量的数据流通活动。

（二）场外交易活跃，场内交易多点突破

场外交易仍是数据交易主要形式。如前所述，数据流通交易并非新生事物。早在 1803 年，一群伦敦裁缝互相交换不能偿清债务的客户信息，成为征信机构益博睿（Experian）发展的基石。1841 年，邓白氏集团（Dun & Bradstreet）率先将各种商业信息收集汇聚，构建信用评定的数据指标体系，数据服务商的雏形初现。长期以来，大量机构收集多方数据，创新业务模式，打造竞争优势，满足数据需求，场外点对点数据交易始终活跃。例如，查询服务行业中，天眼查、企查查等企业通过非人工方式从全国企业信用信息公示系统、中国裁判文书网等公共数据开放平台以及国家知识产权局等部门的官方网站收集、聚合公开数据，运用独有技术与算法，提供独具价值的企业信息查询服务，以会员制的方式实现数据价值的转移。金融行业中，万得、同花顺等企业研发金融终端，对接、采购各类交易所、券商、资讯公司等方数据，汇聚研发形成标准、实时、全面的数据库，供市场各大机构购买使用。综合查询服务、金融、征信、广告、人工智能等各个行业的场外数据交易情况，2022 年场外数据交易规模约为 1000 亿元，是场内交易规模的 50 倍⁵。

部分行业外部数据采购需求强烈。部分企业期望通过内外部数据

⁵ 数据来源：黄丽华，2022 年全球数商大会

的相互融合获得更加精准的市场分析、用户画像等能力来解决业务开展过程中的核心问题，金融机构、互联网平台、广告公司等企业的外部数据采购规模呈稳定增长趋势。以金融机构为例，通过对招投标等公开信息的整理，众多银行每年都会以招标形式采购大量外部数据，涉及反洗钱名单数据、资信数据、司法数据、供应链数据、行业经济数据等种类，用以全面提升本行的风控与营销能力。截至 2022 年，金融行业的数据供应商可达 2200 多家，其中注册资本金在 1000 万元以上的占 67%，公司成立时间不超过 10 年的接近半数⁶。

场内数据交易多元化探索不断取得突破。中国信息通信研究院根据公开信息统计显示，2023 年 1 月至 8 月，国内新成立 5 家数据交易机构。目前，全国已先后成立 53 家数据交易机构，其中活跃的数据交易机构已上架数据产品超 12000 种。各数据交易机构从不同角度发力，在拓展业务模式、强化权益保障等方面积极创新，提升数据进场交易吸引力。2023 年 1 月，贵阳大数据交易所发起招募数据首席地推官，集结银行、交通、医疗、时空数据、数据交付等多个行业和领域的专家，破题数据供需信息差，并率先探索“数据专区”运营模式，打造全国首个气象数据专区和电力数据专区，官网显示截至 2023 年 7 月，两专区交易额分别达到 3500 万元和 1 亿元。2023 年 2 月，深圳数据交易所提出构建动态合规体系，引入“信用”工具，打造动态信用评级，力图降低企业合规门槛，构建可信交易环境，实现数据交易的包容审慎监管。2023 年 7 月，北京市发布数据要素市场建设

⁶ 数据来源：上海数据交易所研究院，金融业数据流通交易市场研究报告

领域十二大落地创新成果，其中九项由北京国际大数据交易所主导完成，包括跨境征信报告核验项目、数据登记业务互认互通、数据资产抵押授信等，为数据跨境、数据服务业务互通、数据要素金融服务等方面探索提供宝贵经验。

所商分离背景下，数商生态不断完善。在广泛意义上，“数据二十条”所鼓励的数据商和第三方专业服务机构，均可纳入“数商”范畴，其核心功能是为数据供需方提供各类数据服务与市场服务。各类数商在地方政府、数据交易所支持下，积极主动寻找市场定位，推动数商生态不断完善。上海数据交易所打造数商生态为特色，围绕完善交易制度、扩大产品供给、完善基础设施建设、提升交易活跃度四项内容，初步构建数商生态体系。2023年6月，上海数据交易所上线“数商生态”服务平台，通过多元化途径协助数商开展业务，将服务渗透到数据交易的各个环节，调动数商参与数据要素市场的积极性。数据经纪人是数商生态中的重要角色，承担数据价值洞察、撮合、代理、评估等职责。2023年7月，由数交数据经纪（深圳）有限公司牵头发起，全国近五十家数据交易场所、科研院所、数据商以及数据交易中介机构共同起草的《数据经纪从业人员评价规范》团体标准正式发布，培育并规范更多专业的数据经纪从业人员。

（三）多措并举破除数据流通障碍

面对流通规则不完善的问题，各地加快健全数据要素市场制度与规则。当前，数据流通的激励和权益保护规则仍未健全，市场主体合规顾虑仍然较大。为此，各地在“数据二十条”指导下，制定符合当

地实际的政策，推进制度与规则的探索。例如，2023年7月，北京印发《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》，提出推进数据产权结构性分置、完善数据收益合理化分配等具体举措，布局推进公共数据专区授权运营、推进数据技术产品和商业模式创新等探索方向。未来，细化“数据二十条”提出的“谁投入、谁贡献、谁受益”原则，建立维护数据资源资产权益、兼顾效率与公平、突出激励导向的数据收益分配制度，完善数据要素市场准入制度和市场竞争框架，建立尽职尽责、容错免责机制等制度与规则仍为各地政府探索的重点。

面对权利归属难界定的问题，各地推进数据产权登记新方式。数据的来源广泛，涉及到的相关主体较为复杂，数据权属不明及错配始终是数据交易操作中的“拦路虎”。“数据二十条”提出研究数据产权登记新方式。各地、各数据交易服务机构针对场内交易，积极探索数据登记制度，通过对每一个进入流通的数据产品赋予唯一的产品编码或标识，发放数据登记凭证，为流通市场提供权威信息，助力保障数据产品流通的安全合规性。2023年5月，浙江省知识产权局、北京市知识产权局分别制定发布了数据知识产权登记相关管理办法。2023年6月，深圳发改委发布了《深圳市数据产权登记管理暂行办法》，从登记主体、机构、行为与监督管理等角度，制定了覆盖数据产权登记全流程的管理办法。未来，数据确权本身仍将存在大量争议，构建以促进产业发展为导向的数据产权登记框架，推进权责明确、保护严格、流转顺畅的数据产权登记体系建设需加快探索。

面对支持力度仍有限的问题，还需进一步鼓励数据流通创新探索。

数据要素市场作为一项新兴事物，有着独特的属性和规律，培育数据要素市场既缺乏先前经验借鉴，又缺少国外经验参考。当前我国数据要素市场培育仍具有“重安全、轻发展”的特点，对数据流通创新探索的支持仍然有限。未来，需着力统筹发展和安全，加大对创新探索的支持力度：积极开展数据特区、数据委托运营、行业数据市场、多级数据市场等创新试点。加大财税综合支持，扶植具有创新技术应用或商业模式的代表性企业，鼓励自由竞争的市场化流通交易。鼓励企业加强数据产品研发力度，主动创新商业模式，主动创造应用场景，主动寻找市场机遇。大力培育多样、专业的数据服务机构，主动对接数据创新成果、分析数据应用需求、撮合数据流通。研究建立数据资产价值评价指标体系，从行业、场景等方面切入，开展数据估值定价试点。进一步探索数据要素金融服务体系，提高相关企业的融资效率。

五、技术：基于业务需求加速创新与体系重构

技术发展持续推动着产业进步，业务需求也对技术提出新的要求。随着数据规模爆炸式增长、数据类型日渐丰富，传统大数据处理技术面临着诸多挑战，以满足业务需求为导向的数据技术体系将会不断变革创新。

（一）数据技术随业务要求不断演进

数据技术（Data Technology）伴随业务要求发展，数据要素三次价值释放不同需求也推动着数据技术不断演进。当前，第一代数据技术（DT1.0）、第二代数据技术（DT2.0）体系已基本成熟，第三代数

据技术（DT3.0）逐渐兴起。

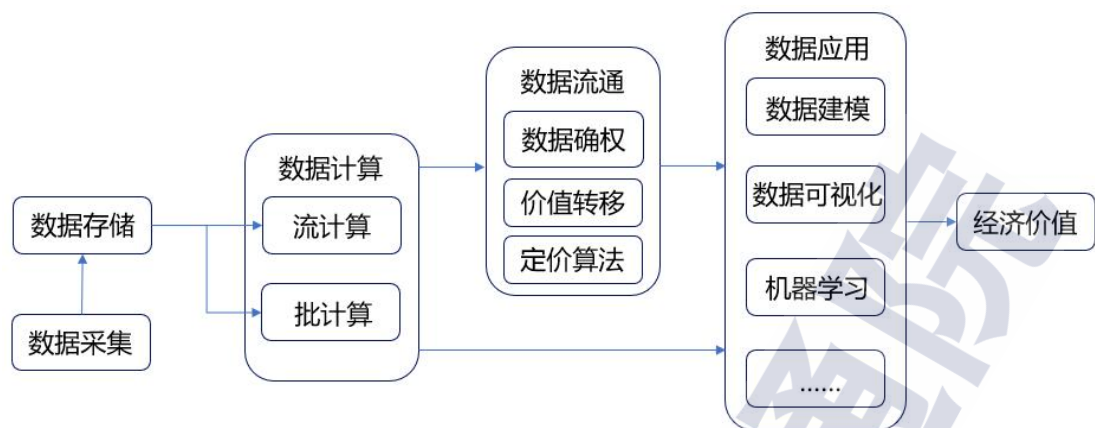
第一阶段（DT1.0 时代），数据技术支撑业务贯通。在此阶段，数据主要来自于业务运转，在不同的业务系统中进行共享及流通，在此阶段数据技术主要支撑数据的事务处理，以文件系统、数据库等技术为代表。

第二阶段（DT2.0 时代），数据技术推动数智决策。数据经过挖掘、清洗、筛选并嵌套入相应场景，实现业务的智慧化、智能化决策。在此阶段数据技术以数据仓库、数据湖以及湖仓一体等技术为代表，以支撑数据的分析、治理等工作。

第三阶段（DT3.0 时代），数据技术进入可信流通对外赋能阶段。在这一时期，数据不仅在企业内部流转，也将会通过流通发挥更大价值，实现多方共赢。数据技术将推动各行业从“有数可用”到“数尽其用”，全场景智能、跨领域协同、数据流通跨域安全管控成为新阶段的发展目标，推动数据要素价值不断向更多应用场景拓展。技术方面以相对匿名化、隐私计算、区块链、全密态数据库以及防篡改数据库等技术为代表，以支撑数据要素可信流通。

（二）数据要素时代新技术不断涌现

数据要素在强调通过大数据处理投入生产的同时，进一步突出了数据在交易流通过程中产生经济价值（如图 8）。可控、可计量、可流通是数据要素对数据技术提出的新要求，新技术不断涌现为数据要素价值释放保驾护航。



来源：中国信息通信研究院

图 8 数据要素流通技术流程图

以云原生、软硬协同以及湖仓一体等技术为代表的数据处理技术持续助力用户降本增效。云原生技术通过存储计算分离架构，实现资源池化和极致弹性，具备高扩展性、高可用性、跨地域规模、低成本等优势，可为用户提供真正具备秒级智能弹性扩容能力、随需而动。软硬协同技术为软件技术带来了新的机遇，一些企业陆续发布 GPU 数据库、数据库一体机等产品，以满足业务规模不断扩张的需求。硬件技术的发展一方面促进了数据处理技术性能提升，另一方面也推动了数据处理技术与其他新兴技术的融合，使得技术体系的安全性和智能性得到提升。在湖仓一体方面，数据仓库（Data Warehouse）和数据湖（Data Lake）两项技术在不断演进过程中逐渐融合形成湖仓一体（Data Lakehouse）技术架构。湖仓一体集数据湖的灵活性、可扩展性优势以及数据仓库的数据结构和数据管理功能于一体，能够降低数据冗余、减少存储成本，提升数据处理时效性。当前，湖仓一体技术落地应用速度持续加快，应用领域主要集中在互联网、电信运营商以及金融等国内数字化程度较高行业。由于数据要素具有规模经济性，企业将趋向于持续积累数据，能够助力企业降本增效的技术将会进一

步受到关注。

以人工智能、隐私计算、区块链、及图技术等为代表的新兴技术持续护航数据要素安全流通。AI 技术与数据处理技术相结合降低技术使用门槛，2023 年 AIGC 技术的跨越式突破发展，不仅使大语言模型进入公众视野，更扩展了数据基础设施与 AI 融合的发展空间。以数据库为例，生成式 AI 在数据库结构设计、架构设计、数据分析挖掘等方面可以不同程度简化人员操作，提高开发、运维、分析的效率，降低用户使用门槛，更好助力数据流通。**数据库与隐私计算相结合落地的全密态数据库**能够解决数据全生命周期的隐私保护问题，使得无论在何种系统环境下，数据在传输、运算以及存储的各个环节始终都处于密文状态。**数据库与区块链技术相结合形成的防篡改数据库**，能够很好地保障云上数据可信运维。防篡改数据库可以内部构建可信索引结构，确保数据库数据的整体完整性，与区块链相结合形成的解决方案，能够形成链上链下链路数据安全可信，解决多方业务间数据安全可信流通难题。**隐私计算与区块链相结合实现数据密态流转**。在控制面以区块链为核心构建数据流通管控层；在数据面以隐私计算为核心构建密态数联网。数据流转过程中以密态形式流通，保障其流转、计算、融合直到销毁的全链路安全可控，同时将数据要素持有权和使用权分离，实现数据可用不可见、使用可控可计量、以及使用权跨域管控。**图联邦技术能够打破数据孤岛**，图联邦技术是为了解决数据孤岛、隐私保护和数据安全问题提出的概念，在保护用户隐私和公司数据的前提下，更好地发挥数据价值。未来，新兴技术将在应用过程中不断融合以适应不同场景下的技术需求，向着更加高效、安全的方向不断发展。

向量数据库、图分析技术、时空大数据平台以及时空数据库等技术有效支持新兴业务场景下的数据要素价值释放。向量数据库能够支撑 AI 技术赋能数据要素价值释放。一些非结构化数据需要通过机器学习算法从中提取出以向量为表示形式的“特征”，向量数据库的兴起便是为了解决对这些向量进行存储与计算的问题。图分析技术助力洞察数据连接新价值。数据要素时代，数据规模将会不断增大，数据量及数据自身丰富度不断增加，图分析技术能够有效分析数据之间的关联性以及处理数据之间的复杂关系。时空大数据平台或时空数据库能够实现海量时空数据管理、查询、统计与分析。在现实世界中有超过 80% 的数据都和地理位置（空间）相关，而所有数据均含有时间属性。因此，实际业务场景中许多数据需要通过时空大数据平台或时空数据库来处理。近年来，时空大数据平台以及时空数据库等技术的发展有效应对北斗时空大数据服务、数字孪生、智慧城市等新兴数据应用业务场景下对于时空数据处理的需求。数据在不同应用场景下得到更好地应用是数据要素价值释放的重要体现，适应应用场景的数据处理技术将在融合创新中满足各项需求。



来源：中国信息通信研究院

图 9 数据要素重构技术体系

（三）数据要素技术体系重构加速

数据要素时代，围绕数据的采集、存储、计算、管理、流通、安全各个环节，技术体系不断革新（如图 9）。**数据采集**方面，传统数据采集通常采用外部数据源购买、人工检索以及搜索引擎数据等手段。未来，数据采集技术将向采集途径多样化、数据类型多样化以及数据结构多样化等方向持续发展，尤其是物联网技术的发展将极大扩展数据采集的范围。**数据存储**方面，数据存储技术将向更高性能、更低成本方向升级，从以存算耦合为代表的传统存储架构向存算分离的新型架构不断演进，从而提高资源共享性和伸缩性。**数据计算**方面，新型计算平台持续涌现以满足不同类型数据处理需求，计算实时性、交互性将会不断提升。**数据管理**方面，传统数据管理技术人工参与多、效

率低，随着 AI 技术的不断发展将会极大程度代替人工从而提升整体效率、压缩成本和周期。**数据流通**方面，在传统的数据流通过程中，由于明文数据拷贝成本很低，数据资源持有权很容易失控。未来，对于数据流通过程中“数据可用不可见”“数据可控可计量”“可溯源存证”等技术理念将不断落地。**数据安全**方面将由传统的防护边界安全技术向保护全生命周期的内生安全技术逐步演进。在以上环节构成的技术体系不断革新完善的基础上，数据要素基础设施、可信数据空间等综合性技术框架将逐渐落地成型，成为系统解决数据共享流通瓶颈问题、安全可信类问题和数据内容保护问题的必要技术底座，确保数据要素在供给方、使用方、服务方、监管方等主体间的通力合作中释放价值。

六、趋势与展望

我国数据要素发展处于活跃探索期，突破方向将逐渐显现。各地、各部门、从多个角度积极布局数据要素发展，各市场主体积极寻找在数据要素发展中的定位和角色，寻找新的业务增长点。尽管目前尚未形成成熟的方法论和推广模式，探索的路径尚未形成体系，但随着探索的深入，新视角、新方式不断涌现，新方向和新热点层出不穷。未来一段时间，公共数据授权运营可能进入大规模落地探索阶段，授权运营的制度、平台、标准等将不断完善，高价值公共数据的高质量供给有望在数据要素市场中率先“突出重围”。各大企业有望结合自身数据资源、数据能力等各方面优势，将各自功能有机联系，形成数据要素生态体系，带动市场各参与主体有序运转，形成“飞轮效应”。

个人数据相关权益保护仍将持续加强，同时基于个人权益保护的个人信息开发利用技术、模式也将不断创新。此外，面对国际数据治理新格局，北京、上海、深圳等地将依托自身跨境合作优势，进一步探索数据跨境的落地实施方案。

数据要素仍是各国战略布局与发力的重要领域。世界主要经济体虽然也未找到数据价值高效流动、充分释放的成熟路径，但也在积极布局数据战略，加快数据制度建设，以平衡数据保护与利用。例如，美国于 2023 年 3 月发布《促进数据共享与分析中的隐私保护国家战略》，将从国家战略的高度支持以隐私计算为核心的保护隐私数据共享和分析（Privacy-Preserving Data Sharing and Analytics, PPDSA）技术体系发展，从而在确保用户隐私、秘密安全的前提下抓住数据发展机会、充分激发数据潜力。2023 年 7 月，欧盟委员会通过了“欧盟-美国数据隐私框架”（EU-U.S. Data Privacy Framework）的充分性决定，标志着欧盟与美国的数据跨境流动重启，全球数据治理格局、数据资源配置体系面临新的变化。

面向未来，数据要素与时代相伴发展，不断衍生出新的问题等待破解。生成式人工智能为数据流通带来了新的可能模式，以内容生成代替数据本身、以模型部署代替数据流动是否可能颠覆现有数据流通模式，需要持续跟踪。推广数据文化、倡导数据利他主义的潮流既是促进数据管理、数据应用、数据流通的理念，又是对数据的经济、社会、文化特性的全新阐释。从面对未知的风险角度看，数据伦理的讨论也日益增加。数据伦理的本质是人的伦理，构建由个人及至家国的

秩序，关怀数字时代的“弱势群体”，持续推进科技向善也是数据要素发展的应有之义。

“大鹏一日同风起，扶摇直上九万里”。历史经验表明，生产要素的扩充代表着生产模式、经济社会运行模式乃至人的生存状态、文化形态、价值追求的变革。从“结绳记事”到大模型，数据伴随社会发展变迁，承载人类认知世界的努力与探索。当前，我国正从“数据大国”向“数字中国”全速前进，数据要素与新兴技术交织融合，数据价值与行业发展相互交错，数据引擎正在被注入巨大动能。面向未来，我们相信，在落实数据基础制度和建设数字中国的新篇章中，数据要素的价值将被更加充分的挖掘，我国数据要素的发展也将迎来新的突破。

参考文献

- [1] 中国信息通信研究院. 数据要素白皮书（2022 年）[R]. 2023.
- [2] 中国信息通信研究院. 大数据白皮书（2022 年）[R]. 2023.
- [3] CCSA TC601 大数据技术标准推进委员会. 数据库发展研究报告（2023 年）[R]. 2023.
- [4] 北京邮电大学经济管理学院数据资产研究中心. 个人成为本人信息数据的主导者——数据要素市场新范式[R]. 2023.
- [5] 戎珂, 陆志鹏. 数据要素论[M]. 北京: 人民出版社, 2022.
- [6] 清华大学金融科技研究院. 数据要素化 100 问[M]. 北京: 人民日报出版社, 2022.
- [7] 高富平, 冉高苒. 数据要素市场形成论——一种数据要素治理的机制框架[J]. 上海经济研究, 2022(09): 70-86.
- [8] 黄丽华, 窦一凡, 郭梦珂等. 数据流通市场中数据产品的特性及其交易模式[J]. 大数据, 2022(03): 3-14.
- [9] 李红光, 王磊, 李颖. 数据资产化视角下企业增信机制研究——基于深圳的实践探索[J]. 价格理论与实践, 2023(04): 33-37.
- [10] 李金璞, 汤珂. 论数据要素市场参与者的培育[J]. 西安交通大学学报(社会科学版), 2023(04): 78-89.
- [11] 李三希, 李嘉琦, 刘小鲁. 数据要素市场高质量发展的内涵特征与推进路径[J]. 改革, 2023(05): 29-40.
- [12] 吕艾临, 王泽宇. 我国数据要素市场培育进展与趋势[J]. 信息技术与政策, 2023(04): 2-8.

- [13] 王宝珠, 王朝科. 数据生产要素的政治经济学分析——兼论基于数据要素权利的共同富裕实现机制[J]. 南京大学学报(哲学·人文科学·社会科学), 2022(05): 21-36+165.
- [14] 王丹阳, 侯宁. 数据要素市场建设中的企业数据合规分析[J]. 信息通信技术与政策, 2023(04): 42-47.
- [15] 王建冬, 李嘉瑜. 构建数据要素“双循环”新发展格局的政策思考[J]. 数字图书馆论坛, 2022(10): 13-16.
- [16] 王伟玲. 中国数据要素市场体系总体框架和发展路径研究[J]. 电子政务, 2023(07): 2-11.
- [17] 徐偲骥. 数据要素红利全民共享机制——基于一、二次分配相结合的探索[J]. 学习与实践, 2023(06): 20-29.
- [18] 严宇, 孟天广. 数据要素的类型学、产权归属及其治理逻辑[J]. 西安交通大学学报(社会科学版), 2022(02): 103-111.
- [19] 张斯睿, 闫树. 数据要素市场建设的关键突破口: 公共数据授权运营[J]. 信息通信技术与政策, 2023(04): 22-26.
- [20] 张新宝, 曹权之. 公共数据确权授权法律机制研究[J]. 比较法研究, 2023(03): 41-55.

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62302928

传真：010-62304980

网址：www.caict.ac.cn

