

# 加快我国大数据技术创新发展 着力打造产业高地优势

## ——《“十四五”大数据产业发展规划》解读

清华大学大数据系统软件国家工程研究中心  
总工程师 王晨  
清华大学软件学院院长 王建民  
中国工程院院士、清华大学大数据系统软件  
国家工程研究中心主任 孙家广

随着第四次工业革命的深入展开，大数据日渐成为宝贵的战略资源，成为推动产业数字化、网络化、智能化发展的关键生产要素。为此，应坚决贯彻习近平总书记对科技创新“四个面向”的战略部署，以创新驱动发展和自主安全可控为使命，以国家《“十四五”大数据产业发展规划》为指导，推动自主技术发展，激发数据要素潜能，积极促进大数据应用落地，形成我国大数据产业面向高端的供给能力，推动我国大数据产业高质量发展。

### 加快发展自主大数据系统软件技术体系，打造高端产品链

近年来，随着新一代信息技术的高速发展，大数据技术日新月异，新概念、新技术、新架构层出不穷，技术和产品迭代速度明显加快。国际上，以 Hadoop、Spark、Flink 等为代表的大数据系统软件，依托 Apache 等国际开源社区，其开放性使得技术与产业生态均得到了快速的发展。

我国在大数据领域的布局较早，无论是政府还是民间都投入巨大，在企业、高校、研究机构等各方的共同努力下，我国大数据技术和产品取得了长足的进步，但是与国外相比仍然存在着一定的差距。系统软件层面的核心技术和产品虽然不存在“卡脖子”的情况，但依然无法避免核心代码依赖国外开源大数据软件的问题，对开源社区及其技术的发展方向缺乏控制力，开源创新不足业已成为制约我国大数据产业发展的瓶颈。

针对我国大数据产业创新不足的问题，应面向世界科技前沿，从大规模多源异构数据一体化管理、交互式异构数据分析框架、数据可视化与智能数据工程等多方面开展技术研发和工程化，研发一系列基于自主技术的国际领先的大数据核心技术与产品，突破创新驱动产业升级转型和关键领域自主可控中的技术瓶颈，着力打造自主可控的大数据技术体系，建设国际一流、开放协同的产学研用大数据技术创新团队，增强对国家重大战略方向、重点工程的大数据系统软件保障能力。

与此同时，面向全球化、开放化、开源化的大数据技术发展态势，积极鼓励我国大数据技术企业“走出去”，通过广泛参与、领导国际开源项目、海外企业并购等方式，加强技术话语权与领导力，形成一批我国主导的大数据开源项目，培养一批具有世界级影响力的大数据技术人才。鼓励我国大数据技术企业积极拥抱开源软件“订阅”模式、云服务模式等新兴业务形态，低成本、高效率地开拓海外市场，推动我国自主的大数据产品扬帆出海。

### 推动大数据技术在各行业深化应用，优化行业价值链

我国大数据与产业的深度融合仍然不足。一方面，我国大数据软件产品在实体经济领域的技术成熟度和易用性距离实际场景下的应用要求仍有差距，总体发展水平滞后于企业的需求和应用水平，多数企业仍然依靠第三方公司来进行大数据平台的管理和大数据应用的开发；另一方面，企业对自主大数据产品的带动孵化缺位，大胆创新仍然不足，应用的深度和广度明显不够，供需互促的良好发展格局仍有待形成。

为此，应围绕工业互联网等国家战略，着力满足我国实体经济发展需求，瞄准国家重大战略需求，努力为国家战略任务和重点工程实施做出贡献。以工业企业为主体，面向航空航天、船舶、电子、石化等重点行业，充分发挥龙头企业的“链主”作用，依托重大工程研发任务，面向数字化和智能化的实际需求，组织产学研用各方力量，多创新主体共同参与，联合研究、共同开发，研制行业/专业领域国际领先的大数据应用产品，探索大数据应用实施工程方法，将实际需求和技术溢出转化为现实导向和目标，加速推动企业数字化转型与高质量发展。

突出需求侧和供给侧之间的良性互动，加速推动企业形成的先进技术创新成果向大数据系统与平台软件开发企业的溢出，创新供需双侧融合驱动的技术攻关与产品协同发展模式。一款好的大数据产品是“用出来的”，支持和鼓励企业试用、使用国产大数据产品，实现产业需求带动技术发展和技术发展促进企业革新的双向正反馈，在使用过程中逐渐迭代完善国内产品。扩大试点示范项目，促进优秀产品的推广应用，实现大数据产品的质量提升，并带动传统产业转型升级。

### 构建开放融合的开源社区 推进大数据技术创新

平台级的大数据系统软件技术与产品，在大数据产业的价值分布中位于核心地位，其范围涵盖数据采集、存储、治理、分析的数据全生命周期，与行业大数据应用相比技术难度大、开发周期长、商业回报率低，企业自主研发投入大，因此我国现有的大数据系统软件产品大量采用国外的开源技术，但对于开源软件的贡献

相对不足。

开源具有天生的“开放基因”，凭借其开放性、灵活性、低成本和技术创新性等特点而迅速走向成熟，逐步发展成为一种主流模式，成为软件多元化的重要来源之一，日益改变着信息产业和大数据产业的发展轨迹。开源是解决我国大数据技术瓶颈的有效途径，通过大数据开源社区建设，构建自主大数据开源社区及支撑平台，形成良好的运营机制，孵化由我国主导的大数据开源项目，同时鼓励将各类专项成果汇聚到开源社区之中，通过开源的方式向全社会共享，强化“产学研用”深度融合，引导构建自主可控的大数据生态，解除桎梏和枷锁，实现我国大数据技术和产业从“并跑”向“领跑”的跃进。

为此，应通过建立大数据开源生态社区，将关键技术以开源的方式对社会免费开放使用，快速触达真实用户形成社区，从而获得有价值的反馈，吸引高水平开发人才加入共同研发，保证大数据系统软件的开发朝着代表广大用户真实需求的方向迭代发展，形成正向反馈，逐步产生示范效应扩大用户群体，吸引更多高水平的伙伴深层次合作，从而促进形成国内一流的产业环境。

培养和汇聚大数据系统软件高端技术人才，主动承担国家和行业重大科研项目，取得一批关键技术成果并成功转化，构建大数据领域自主知识产权体系，形成可持续的产学研协同创新机制，为推动我国大数据的技术进步和产业发展提供技术支撑。

### 加速大数据标准与测评体系建设，实现标准引领

我国大数据相关标准仍有较多缺失，亟须通过标准引领和规范大数据技术与

产品的发展。从标准类别分布看，平台/工具及行业应用类标准化需求最为旺盛，与我国大数据产业应用场景多、应用模式丰富等特点相一致。从标准研制路径看，标准化工作重心逐渐由基础类标准、数据类标准向平台/工具类标准、治理与管理类标准转移，由通用领域标准向垂直应用领域标准延伸。从新增标准化需求看，标准需求从通用应用系统向场景化应用系统转变（如图数据库系统、批流融合计算系统）。

围绕 DCMM 的工作推进，治理与管理类标准面向产业供需双方需要，标准化方向主要覆盖数据治理、数据服务、数据资产评估等领域，大数据治理标准体系初现雏形，需要结合后续评估工作不断优化完善。行业应用类标准持续关注工业大数据标准化，在生态环境、电力等垂直应用领域也出现了数据体系、数据分类等标准需求，数字化转型支撑相关标准化预研工作逐步开展。围绕“十四五”大数据国家标准，要不断夯实大数据标准体系建设，加快重点领域标准研制，健全标准符合性评测体系，推进国际标准化工作。

另外，我国大数据产业发展的公共服务能力和综合保障能力仍然不强。目前我国缺少统一的评测机构开展对大数据软件的评价工作，开发企业在软件测评时、用户企业在软件选型时缺乏第三方测评机构的评估和认证支撑。应搭建由多方参与的测试验证平台和安全保障平台，通过统一科学的标准和专业的测试环境，对大数据软件产品的质量和安全性等进行独立客观的测试、验证和评价，其结果可为企业的选型和应用提供指导。同时，应不断提升综合保障能力，为大数据产业提供政策引导、知识产权保护、数据资产保护等服务。

# 回眸“十三五”展望“十四五” 我国大数据产业蓬勃发展潜力巨大

## ——《“十四五”大数据产业发展规划》解读

中国电子信息产业发展研究院副院长  
张小燕

2021年11月30日，工业和信息化部发布《“十四五”大数据产业发展规划》(以下简称《规划》)，全力推动大数据产业高质量发展。各地方为贯彻《规划》要求，推动《规划》落地实施，有必要系统梳理“十三五”时期产业进展和优势短板，分析研判“十四五”时期产业趋势和方向路径，为新时代推动大数据产业发展开好局、起好步打下坚实基础。中国电子信息产业发展研究院自2017年开始，连续五年开展中国大数据产业发展水平评估和趋势研判工作，对各地“十三五”时期大数据产业进展阶段、发展特点、模式路径以及重大趋势进行了深入研究，以期为推动我国大数据产业发展提供有益参考。

### 回眸“十三五”，我国大数据产业蓬勃发展，整体呈现梯级分布、路径多样、集聚发展三大特点

2017—2021年间，按照《大数据产业发展规划(2016—2020年)》提出的“建立大数据发展评估体系”的要求，我们聚焦基础环境、产业发展、行业应用3个关键领域，经过5年迭代优化，形成了由3个一级指标、13个二级指标、40余个三级指标组成的中国区域大数据产业发展水平评估工具箱。在5年的评估工作过程中，评估工具箱保持一级、二级指标的延续性、稳定性，并根据年度产业发展的新形势、新特点、新趋势对三级指标进行优化调整，确保既从主体方向上体现产业发展的持续性和可比性，又从细分领域引导产业发展新方向和新任务。

通过5年评估，从总体看，我国大数据产业发展水平整体呈现从东南沿海向中西部内陆地区梯级下降的特点，先进省市从环境优化、产业发展到行业应用全面发力，后发省市立足独特优势加快特色赶超，推动全国逐步形成了以8个国家大数据综合试验区为引

领、多区域集聚的良好发展格局。

#### （一）梯级分布特征明显，东部地区整体领先，中西部追赶势头强劲

“十三五”期间，我国东部地区大数据产业整体发展水平最高，5年来总指数和平均指数排名“双优”，北京、广东、江苏、上海、浙江、山东、福建7个省市一直位居全国前10名；中西部地区发展势头强劲，四川、湖北、贵州、重庆、安徽等亦多次入围前10榜单。

“十三五”期末，通过对各省（自治区、直辖市）的大大数据发展水平及其当年的GDP进行关联分析可以看出，目前各地区的大大数据产业发展状况分为四类。一是蓬勃发展型，这类地区经济基础与大数据产业发展水平平均好于全国平均水平；二是潜力型，这类地区经济基础良好，超过全国平均水平，大数据产业发展虽相对落后，但具备赶超发展的潜力；三是特色引领型，这类地区经济基础相对落后，但大数据产业发展水平好于全国平均水平；四是起步型，该类地区大数据产业发展水平和经济基础均弱于全国平均水平。其中，东部地区整体表现亮眼。此外，西部地区的四川、中部地区的湖北表现优异，也形成了产业发展与经济基础并强的局面。

#### （二）发展路径丰富多样，第一梯队均衡发展，第二梯队特色赶超

“十三五”期间，各省（自治区、直辖市）立足本地产业基础和资源禀赋，积极探索大数据产业发展模式，逐步形成了均衡发展型、环境带动型、产业驱动型、应用牵引型四类模式。

第一类均衡发展型，地方经济基础雄厚、理念先进，致力于推动基础环境、产业发展和行业应用协同全面发展。以“十三五”期末数据为例，我国大数据产业发展第一梯队的省市主要属于均衡发展型，其中尤以江苏、上海和四川为突出代表。

第二类环境带动型，以优化产业发展环境为切入点，围绕完善政策环境、健全组织机构、夯实基础设施、强化集聚示范等维度，以良好环境吸引企业主体、资金、人才等产业要素集聚。如贵州、河南在“十三五”期末

大数据产业发展指数排名虽位列全国第11和第13，处在第二梯队，但其大数据产业发展环境较好。以贵州省为例，“十三五”期间，贵州省以完善基础环境为立足点，依托国家大数据综合试验区相关政策，建立健全组织机构，发挥集聚示范作用，不断推动大数据产业发展迈上新台阶。

第三类产业驱动型，立足本地产业优势，大力引培优势企业、不断提升创新能力和水平，加快数据资源体系建设，以产业发展驱动大数据整体进步，其中以北京市最为典型。“十三五”期间，北京市大数据产业的产业发展指数始终位于全国前列，在企业数量、企业竞争力、创新能力等方面具有显著优势。

第四类应用牵引型，围绕本地产业特色和发展需要，积极拓展行业应用，推动产业数字化转型，实现大数据与传统行业深度融合，进而牵引本地大数据产业发展。如重庆、安徽大数据发展指数排名虽位列全国第12、第14，处在第二梯队，但其大数据行业应用较强，已逐步形成应用牵引型发展模式。

#### （三）地区集聚效应凸显，综合试验区引领作用显著，国家重大战略区域逐渐成为大数据产业集聚发展重要载体

“十三五”期间，以8个国家大数据综合试验区为引领，多区域集聚发展的格局日益凸显。从5年的评估结果看，8个国家大数据综合试验区示范引领作用明显，在全国大数据产业发展总指数中总体占比均接近40%，广东省、北京市、上海市等更是长期占据综合排名前5，在全国大数据发展总指数中占比平均超过15%。受益于区域一体化发展等国家重大战略的深入实施以及数字产业集群建设加速推进，以北京、广东、浙江、江苏等省市为引领，以京津冀、长三角、珠三角等国家重大战略区域为依托的大数据产业集聚发展步伐加快。各集群以释放数据要素价值为核心，在数字基础设施建设、大数据技术产品研发、大数据与实体经济融合等方面协同发力，推动大数据产业发展模式和区域发展格局持续优化。

### 展望“十四五”，我国大数据产业潜力巨大，将着力推动五大转变

展望“十四五”，我国大数据产业将有可能在数据要素市场、产业基础能力、产业链优化升级、行业融合应用、产业规范治理等方面实现“五大转变”。

#### （一）数据要素市场建设提速，市场化配置从理念设计走向落地实践

随着数据要素市场顶层设计不断完善，以及广东、江苏等地方实践的深入开展，数据要素市场化配置将在制度规范、运营模式、交易机制等方面形成一批地方特色模式、特色路径，数据要素规范有序流通、高效公平配置、充分开发利用的发展格局将初步形成。数据产权界定、价值评价等基础理论研究加快推进，将有可能在技术支持、动态衡量、场景驱动等条件下取得阶段性进展，数据要素价值评价将在金融、能源、政务等基础较好的领域率先实现突破。

#### （二）产业基础能力稳步提高，基础设施建设从分散同质迈向错位协同

在全国一体化大数据中心体系建设部署下，以数据中心为代表的数字基础设施将加快突破，在区域协同、数据流通、算力普惠等方面形成合力，进一步增强我国大数据产业发展基础优势。

数据中心统筹推进、一体建设态势明显，(超)大型数据中心规模化建设布局持续优化，京津冀、长三角、粤港澳大湾区等8个全国一体化算力网络国家枢纽节点领先发展，在区域数据中心集群建设、算力协同调度机制突破、市场化运营模式创新等方面率先突破。“东数西算”工程全面实施，在突出地区优势、补足发展短板的原则下，将推动东西部地区数据中心差异化、协同化发展。

#### （三）产品服务体系加快完善，产业链发展重心从建链补链转向优链强链

经过5年的努力，我国已经初步形成了贯穿数据全生命周期的大数据产品和服务体系。

在以大数据为代表的数字技术主导权竞争日益激烈的背景下，拥有关键技术能力的大数据企业将更加注重基础平台、数据存储、数据分析等产业链关键环节的自主研发，数据采集、存储、处理、应用等部分领域关键技术短板将有所突破。面向大量数字化转型需求，企业将更加注重提升定制化、体系化的综合解决方案能力，强化系统运营增值服务，拓展价值创造空间。受数据要素市场政策红利拉动，数据标注、治理、流通、交易等新兴服务业态快速兴起，大数据产业链将持续优化完善。

#### （四）融合应用市场加速繁荣，场景挖掘将从边缘浅层向核心深层拓展

随着大数据融合应用场景挖掘持续向核心业务环节、更高层次拓展，应用广度、深度和强度将大幅提升，推动我国大数据融合应用市场加速繁荣。

从应用场景覆盖环节看，大数据与实体经济的融合将从管理、服务等边缘环节加快向生产、制造等核心环节深入拓展，通过行业知识、业务、流程的数据化、算法化、模型化，赋能“产业大脑”系统，形成服务于产业发展的动态感知、敏捷分析、全局优化、智能决策等强大能力。

从应用场景覆盖层次看，大数据应用将从描述性、预测性等浅层应用向决策指导性更深层次分析应用延伸，企业运营管理决策优化、供应链精准管理、基于数字孪生体的制造执行、跨系统跨产业链的综合性分析等深层次应用将加速涌现。

#### （五）产业生态日趋规范有序，行业治理从无法可依进入有章可循

随着《数据安全法》《个人信息保护法》发布实施，细分领域以及地方的条例规定将加速出台，为大数据产业规范有序发展提供重要遵循。数据安全、数据采集、隐私保护、个人信息处理等方面的治理规制将持续完善，大数据企业无节制采集用户信息、泄露个人隐私数据、大数据杀熟、算法歧视等问题将得到集中整治。大数据行业治理的技术手段、监管机制等正加快成熟，符合大数据产业发展特点和需求的治理体系将逐步形成。