

开放性创新赛道（等级奖）

一等奖

AI+自动化技术赋能药物研发数据要素构建及应用

生物医药行业面临高质量数据匮乏、获取成本高、负样本缺失及跨域异构严重等痛点，制约药物研发效率与原始创新能力。深圳市晶泰科技有限公司创新性提出“AI+自动化”数据生成范式，构建集 AI 智能体、AI 自主实验室、仿真计算于一体的数据基础设施，实现药物研发数据规模化生产、采集与治理。



图 1 深圳市晶泰科技有限公司 AI 操作平台

一是汇聚多域多类药物研发数据，全面支撑研发全环节需求。汇聚多领域、多类型药物研发数据，包括公开域的全球专利数据，私域的高通量 AI 自主实验过程数据、实验结果数据及行业稀缺的负样本数据，以及基于物理模型的微观分子结构与力场仿真计算数

据，全面支撑药物研发全环节需求。

二是多渠道标准化采集研发数据，闭环治理保障数据迭代优化。针对公开数据，通过自主研发的 AI 智能体平台实现千万级规模数据高效挖掘、清洗；自有数据方面，依托 300 余台机器人工作站集群，7×24 小时工业化采集标准化真实实验数据；利用高算力仿真平台生成计算数据。通过数据校准和标准化形成高质量数据集，构建“数据—算法—实验”闭环治理流程，保障数据持续迭代优化。

三是数据融合赋能药物研发全核心环节，模型反哺提效降本增成功率。数据深度融合应用于药物发现、晶型预测、活性筛选、构效分析等核心研发环节。基于高质量数据训练垂类大模型与专用小模型，模型反哺指导实验设计与预测，自主推进实验，形成自进化系统。数据生产效率较传统方式提升数十倍，药物早期研发周期显著缩短，研发成本降低、成功率提升。

四是聚数盈利带动产业增效，破垄断促创新惠及民生。在经济效益方面，累计积累数千万条高质量数据，自身实现盈利，带动产业链上下游降本增效；在社会价值方面，打破国外在药物研发软件与基础数据领域的垄断，保障国家生物医药数据安全，加速创新药上市进程，造福广大患者；在产业创新方面，推动科研范式变革，拓展至新材料等领域，驱动相关产业创新发展。

二等奖

汽车行业可信数据空间探索与应用

汽车产业数据体量大、应用广，但长期面临数据类别广模态多导致“加工难”、数据来源分散导致“协作难”、行业交易机制缺失导致“变现难”三大痛点。中国汽车工程研究院股份有限公司联合多家单位，基于跨主体研产销服数据，构建数据智能治理工具、安全可信流通平台、数据全链协同机制，推动数据在 50 余家主体间高效可信流通。

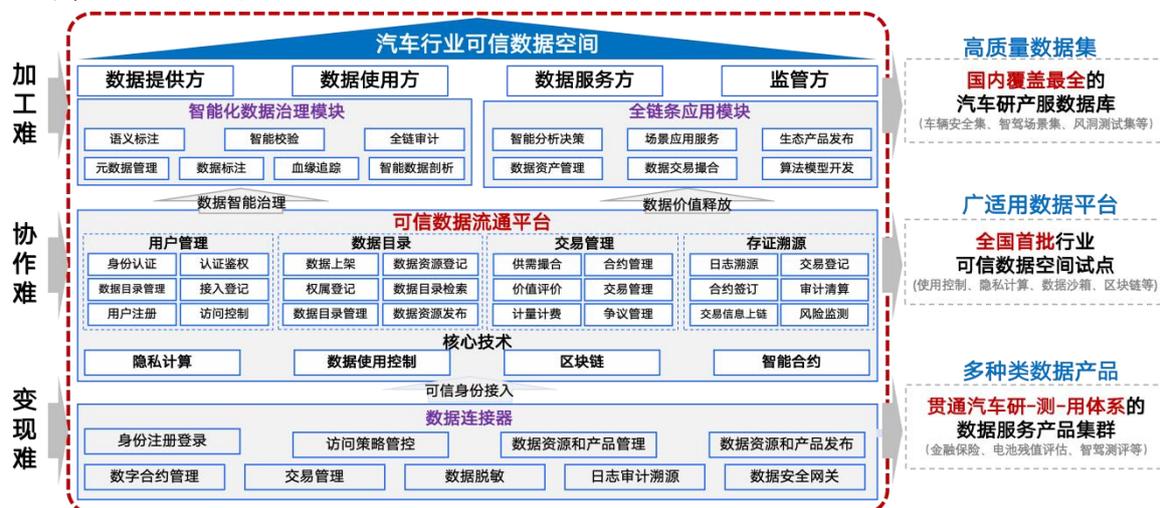


图 1 总体架构图

一是汇聚全周期数据集，全面覆盖汽车产业全场景。汇聚“三横三纵”高质量数据集，总量超 18PB，包括重型车排放、智能新能源汽车安全监管等监管数据，汽车智驾、风洞、安全等研发检测数据，以及新能源汽车热失控、标准法规、失效及事故调查等特色数据库，全面覆盖汽车产业全生命周期。

二是多元模式实现常态数据汇聚，规范商业闭环赋能产业流通。

通过数据授权运营、联盟多方联合建库与场景共建等方式实现常态化数据汇聚。建立规范化商业模式：上游加工端实施协商分配、自主定价模式，提升数商积极性；下游产品端实施充分竞争、市场竞价模式，发挥市场调配能力。数据流通采用授权调用、平台直连等方式，服务车企、保险和监管机构，同时通过西部数据交易中心开展数据产品上架、撮合与交付。

三是多场景落地数据产品，研运两端赋能汽车产业升级。18个数据产品落地五大高价值场景，包括保险及理赔、二手车流通、汽车金融服务、汽车跨境流通服务、智驾大模型训练。上游研发端，基于百万公里自然驾驶数据集与6000余例交通事故数据集，支撑智能网联汽车算法训练与验证，服务多款车准入试点及出口；下游运行端，基于千万辆新能源汽车运行数据集，支撑电池安全预警算法研发，成功预警300余高危车辆，算法水平国际领先。

四是企监双向提效赋能治理，以产业创新支撑品牌全球化。**企业侧**，提升车企、保险、能源等行业研发与测试效率，促进产业链协同与模式创新；**监管侧**，提高安全监管、车辆合规管理与风险预警精度，助力智能交通与城市治理；**产业侧**，赋能智能网联汽车技术迭代创新，支撑30余家车企、100余款车型研发测试，服务民族品牌走向全球，提升我国汽车产业国际竞争力与话语权。

二等奖

格兰德全球企业跨境数据集：企业出海加速器+AI大模型新基座

我国企业出海面临西方数据垄断、跨境交易“信用裸奔”风险，同时 AI 大模型存在“数据荒”、全球数据“汇治用”难题。青岛格兰德信用管理咨询有限公司建成国内最大国产化全球企业数据集，覆盖 230+国家/地区、6 亿家企业、超 300 亿条数据，打造“查全球”平台，并为 AI 训练提供数据支撑。

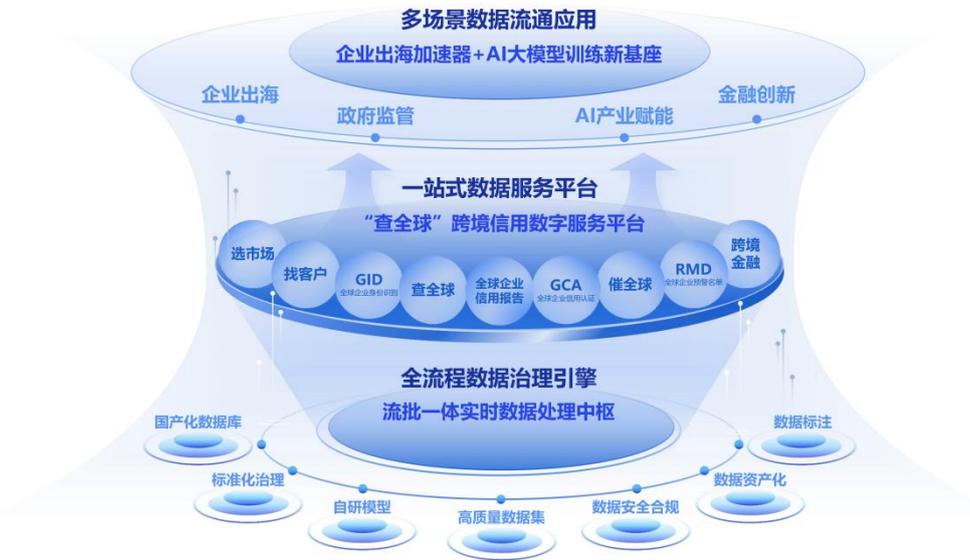


图 1 总体结构图

一是汇聚企业全生命周期全球数据，全面支撑出海全流程需求。汇聚全球企业全生命周期数据，包括企业注册与核心识别信息、股权与高管信息、财务与经营状况、全球海关数据、知识产权与司法诉讼信息、新闻舆情、上下游供应链数据，以及多国官方制裁与风险名单数据，全面覆盖企业出海全流程需求。

二是多元模式汇聚可持续数据，以智能治理赋能合规高质应用。采取“多元采集、生态共建”模式，通过全球公开数据汇集、国际合规采购、行业数据交换、企业自主提供及自有业务沉淀五大方式汇聚数据。与合作伙伴建立数据交换与收益共享机制，确保数据可持续更新。构建独创的智能治理范式，将海量原始数据转化为可直接驱动 AI 大模型训练的高质量数据资产，同时通过标准化处理保障跨境数据合规流通。



图 2 “查全球”服务平台

三是双核心方向赋能数据应用，出海服务与模型训练双向提效。数据应用聚焦两大核心方向：企业出海服务方面，依托“查全球”平台，以“全球企业身份识别编码（GID）”实现企业精准匹配，融合“RMD 全球风险预警名单库”实时筛查，为市场开拓、风险管控、贸易融资提供全链条服务；AI 大模型训练方面，将清洗、结

构化后的多语言、多模态商业数据作为高质量语料库，优化大模型商业分析、报告生成、跨语言理解等能力。

四是数据助力跨境贸易提质增效，赋能开放合作。**在经济效益方面**，保障企业跨境贸易额超 6000 亿元，服务 20 万用户，助力信贷 10 亿元，挽回应收账款损失超 1 亿元；**在社会价值方面**，落地“信用上合”“信用澜湄”跨境信用平台，支撑“一带一路”及上合组织建设；**在产业创新方面**，实现全球企业数据国产化替代，参与国际规则制定，提升我国在全球数据治理与跨境信用服务体系中的话语权。

二等奖

油田土地智慧沙盘 × 数据生态共同体，构建数实融合“双石油”价值创造新范式

石油行业面临高投入、高风险、数据流通难等痛点，一口陆上深井成本高昂，油气勘探成功率不足 30%，能源行业敏感性导致数据流通壁垒。新疆石油管理局有限公司创新“双石油”范式，以地下石油开采流程为蓝本，构建覆盖“勘探—钻井—炼化—集输—销售”全生命周期的“数据石油”开采体系，降低行业成本与风险。



图 1 “数据石油” 开采解决方案示意图

一是汇聚石油产业全链条多源数据，构建全生命周期数据集。汇聚生产经营全链条多源异构数据，包括 70 年时间序列数据、油田生产空间位置信息等空间序列数据，以及油田物联网、5G、卫星、无人机等业务序列数据。构建 28 个专业高质量数据集，总规模达 3.83PB，全面覆盖石油产业全生命周期。

二是标准引领建高质量数据集，生态共建促数据交易落地。以数据标准为治理基础，通过油田物联网、5G、卫星、无人机等技术开展高质量数据集建设。与武汉大学、新疆大学等构建产学研数据生态共同体，采用数据产品交易分成、收益共享模式，确保数据可持续汇聚。依托北京国际大数据交易所开展数据交易，上架地面设施数据产品和区域资源数据产品，已有 60 余家单位意向采购，完成首单交易。

三是融智赋能石油产业多场景应用，破题增效实现数据合规流通。融合多类型数据，基于可信数据空间，依托数字孪生、mm-yolo 算法、GIS、隐私计算、区块链等技术，经多环节处理后，应用于设备巡检、异常识别、生产监控、用地管理、工程规划评估、数据产品交易等场景。破解原油处理状态不可视、异常响应慢、人工巡检难、隐患漏检等难题，实现数据要素合规安全流通。

四是降本增效带动产业链增收，赋能发展树立行业标杆典范。**在经济效益方面**，对内年降本增效 1.1246 亿元，对外带动上下游产业链间接创造经济效益 1.8 亿元/年；**在社会价值方面**，为新疆“数字丝绸之路”建设提供支撑，带动 300 余家相关企业，创造 10000 余个就业岗位，2025 年调整 139 口井井位避让生态红线与基本农田；**在产业创新方面**，作为中石油数据资产领域首个标杆案例，形成显著示范效应，开创数据要素价值创造新路径。

三等奖

面向国际月球科研站等深空探测任务的全球云化数据平台

国际合作深空探测任务面临数据格式不一、标准异质性、协作效率不高、权属关系不明确、数据主权无法保障等问题，制约全球协同探索。深空探测实验室（天都实验室）搭建全球云化数据平台，构建国际化深空探测全链条数据生态体系，形成高质量数据集，支撑月球科研站建设与全球科技治理。

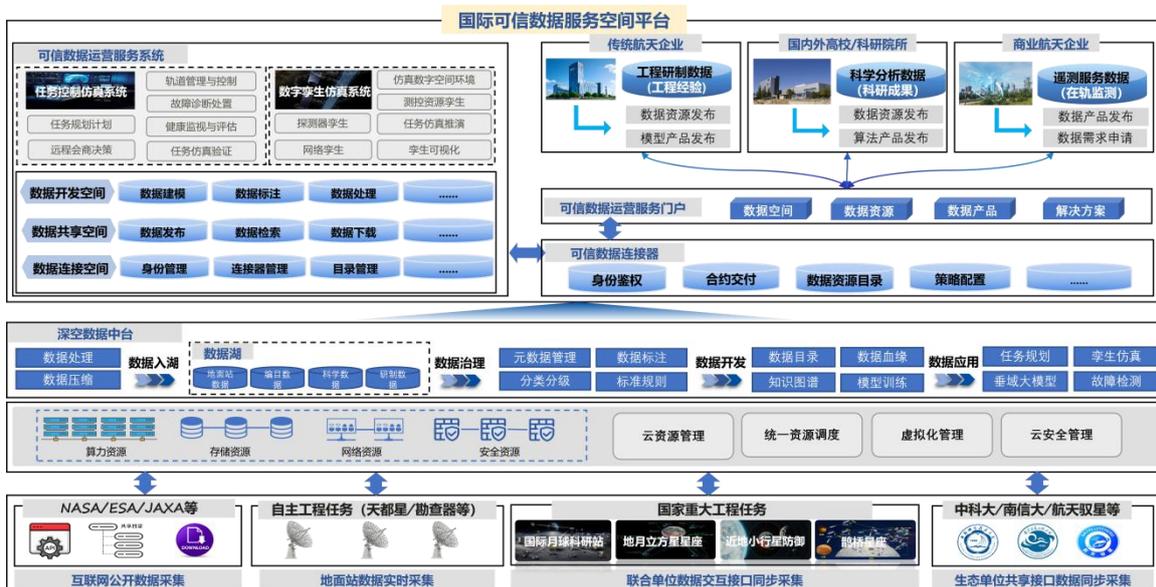


图 1 平台总体能力体系示意图

一是汇聚全球深空探测多类数据，支撑国际合作与科研创新。汇聚全球深空探测任务多类型数据，包括探测器设计、观测、试验等工程数据，月球及深空天体地质、资源等科研数据，覆盖多国多机构协同观测数据，数据总量达 25.8TB、约 129 亿条，为国际合作与科研创新提供支撑。

二是标准统一破解数据异构难题，确权流通构建可信合作体系。

融合国内外数据标准规范，设计“三类五级”数据分级分类方案、统一语义模型与关联算法，解决数据格式和标准异质性问题。设计数据动态确权模型，根据各参与国技术、资金贡献大小计算数据所有权，结合区块链、数字水印、访问控制等技术，制定多边合作数据流通机制。建设含六大模块的可信数据空间，实现数据订阅、检索和协同分析等功能，保障数据安全可信流通。

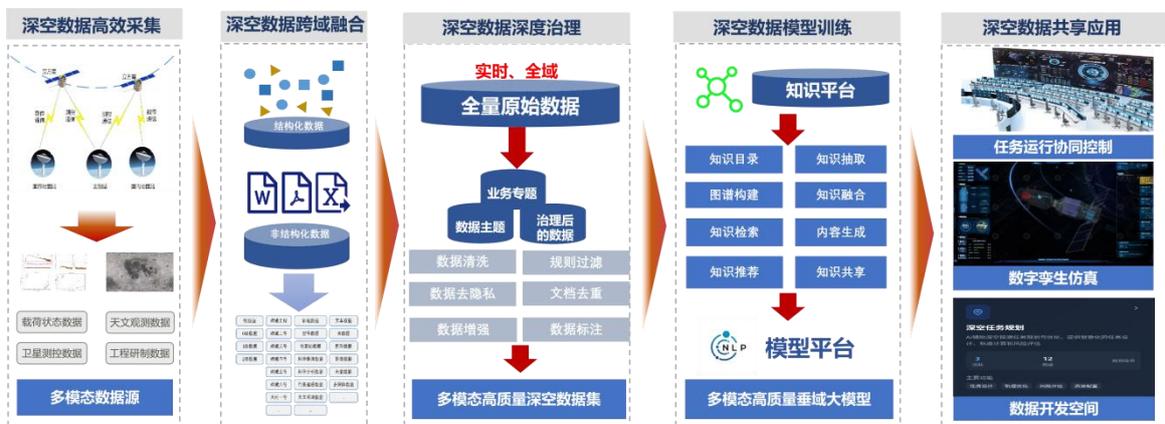


图 2 数据平台架构图

三是多场景赋能深空探测发展，提效协同构建深空经济生态。数据应用于探测器设计仿真、深空资源评估、国际协同观测、商业服务等场景。AI 仿真助手将探测器设计仿真时间缩短 30%以上，提升研发效率；平台支撑多国协同观测和任务联合分析，提升国际合作数据利用率；推动深空数据向工程应用、资源评估、商业服务延伸，构建深空经济合作生态。

四是赋能深空经济带产业发展，协同全球合作提升国际话语权。在经济效益方面，推动深空经济领域合作，带动相关产业发展；在社会价值方面，通过全球首个国际深空探测学会，拉通 17 个国

家与 50 余家机构参与合作，促进全球科技治理；**在产业创新方面**，建成相关平台并授权多项发明专利软著，为国际深空探测合作提供数据支撑与机制保障，助力我国深空探测领域国际话语权提升。

三等奖

面向 AI 原生的高质量数据集创新—构建“Data4AI”的教育高质量数据集实践

我国教育 AI 产业面临方法论缺失、主力军缺位、技术创新滞后、要素割裂等问题，丰富的应用场景数据难以形成协同生态，制约数据价值释放。上海库帕思科技有限公司提出“三个重构”，聚焦教育垂类大模型全流程数据供给需求，按“2+2+1”模式（即教育行业通识数据集，包括文本数据集和音视频数据集；教育行业专业专识数据集，包括思维链强推理数据集、监督微调数据集；教育大模型评测数据集）构建高质量数据集，为教育大模型训练、微调、评测提供支撑。

一是汇聚教育行业多模态多层次数据，构建全阶段数据集。汇聚教育行业多模态、多语言、多层次数据，包括专业书籍、专业教材、在线课程、教学案例、教学过程、教学行为等各类知识与数据。分类构建数据集，覆盖预训练阶段的大规模文本与视频数据，后训练阶段的高质量思维链与精调数据，以及应用阶段的实用型测评数据。

二是全流程保障教育大模型数据供给，以多元运营开放赋能教育生态。聚焦教育大模型“事前、事中、事后”全流程数据供给需求，综合采用公开数据再生产、合作开发、众包采集等方式获取数据。建立分级分类运营可持续发展模式和“3+X”业务矩阵（即语料产品和服务、数据基础设施服务、咨询服务、生态链接服务），

兼顾商业运营和开源开放，开放国内首个数据公共服务平台，开源超 100 万条数据，累计下载次数超 150 万次。

三是多主体赋能全领域数据应用，海量供给支撑大模型研发落地。数据应用覆盖多类主体，基模公司利用多模态数据集进行预训练；高校院所利用微调数据集开展场景适配、特色数据集构建和人才培养；智能体开发者利用开源数据进行二次开发。数据集规模突破 7029TB，涵盖教育、医疗、工业等全领域 1106 个子项数据集，为 100 余家企业提供数据服务，支持不少于 10 个大模型研发。

四是营收增效驱动教育 AI 发展，以体系创新填补行业方法空白。在经济效益方面，产生 1.2 亿元直接营业收入，推动教育 AI 产业降本增效；在社会价值方面，培养 AI 人才，促进教育数字化转型，提升教育服务质量；在产业创新方面，构建“4+1+X”高质量数据体系（即基石数据、SFT 数据、CoT 数据、评测数据、知识图谱、专业数据集），构建“Data4AI”技术体系，为行业提供可复制范式。

三等奖

“从山顶到海洋”的污碳协同遥感数智赋能

“构建从山顶到海洋的保护治理大格局”“减污降碳协同增效”是生态文明建设的重要方向，但传统现场观测难以实现跨空间、多尺度的污碳一体化动态监测。自然资源部第二海洋研究所牵头联合多家单位，依托卫星遥感技术优势，研制水质和碳参数遥感数据产品，为浙江省入海河流、沿海排口、海湾及近海区域提供超 100TB 污碳监测数据服务，提升水体环境监测效率。



图1 项目宣传图

一是汇聚多主体多类型水相关数据，全维度支撑污碳协同监测需求。汇聚多主体、多结构、多类型数据，包括水质自动站、浮标、船测等实测数据，Sentinel-2、Landsat 等系列卫星遥感数据，以及土地利用类型、工业企业分布、农业种植结构等辅助数据。数据覆盖点状站点数据、线状走航数据、面状遥感数据及三维再分析数据，全面支撑污碳协同监测需求。

二是技模融合赋能全流程数据治理，标准引领产出高价值监测产品。创新性将人工智能算法、传统机理模型与海洋数值模拟深度融合，应用于数据治理全过程。攻克“非光学活性物质”遥感反演难题，制定污碳遥感监测行业标准 1 项和团体标准 2 项，建立统一数据处理规范。构建“数据汇聚—清洗—建模—产品化”全流程，形成总氮、总磷、有机碳浓度等高价值遥感数据产品，保障数据可复用、可流通。

三是双核心场景落地数据应用，精准监测赋能生态环境治理。基于“AI+机理+模型”技术架构，融合 XGBoost、FVCOM 等算法与数值模型，数据应用于陆源排污监测和近海碳汇监测两大核心场景。入海排口监测场景中，24 小时内响应 4000 余个排口异常排放，推动异常排口数量三年下降 75%，发现 24 个未申报排污口；海湾与近海监测场景中，实现富营养化监测、碳源汇逐小时监测等，支撑杭州亚运会海洋环境保障。

四是业务化运行创造显著经济效益，创新监管赋能生态治理转型。在经济效益方面，连续四年在浙江七地市业务化运行，累计产

生经济效益近 1200 万元；**在社会价值方面**，推动海洋生态环境监管向高频次、准实时、全覆盖转型，相关成果入选“浙江省生态环境十大科技创新”；**在推广价值方面**，已在粤港澳大湾区复制落地，纳入生态环境科技成果转化典型案例，SatCO2 平台在全球 20 多个国家推广，获批联合国海洋十年项目。

三等奖

“高质量基础要素数据+顺丰物流大模型”，赋能城市数智化转型

城市数智化转型中，时空决策缺乏精准数据支撑，产业侧物流成本高、库存优化难，政府端民生诉求响应与应急调度效率有待提升。昆山市大数据资源管理中心联合丰图科技等单位，以“高质量时空数据+垂类大模型”双轮驱动，整合政企多源时空数据，训练“丰知”大模型，形成近千项智能服务，覆盖 50 余个业务场景。



图 1 高质量时空数据赋能解决方案架构图

一是汇聚多形态多领域城市相关数据，全场景支撑城市多元发展需求。汇聚多形态、多领域数据，包括顺丰物流亿级地址数据、昆山市治理数据，以及新能源、高端制造、大闸蟹等特色产业数据。数据涵盖地址文本、图像视频、信号指纹等多形态信息，全面支撑城市物流、应急、政务、产业等多场景需求。

二是双机制保障数据质效安全，双轨运营赋能政企多元需求。建立多源数据交叉验证+业务实战反馈闭环校验机制，保障数据准确、鲜活、一致。成立跨部门数据治理委员会，遵循内部规范与国家法规，经测绘主管部门审核，确保数据安全合规。采用 TOG 非营利与 TOB 商业双轨运营模式：面向政府提供公益性数据服务，支撑城市治理；面向企业推出按需付费增值服务，满足个性化需求；通过数据交易所、数商等专业机构开展规范化交易，提供数据集、API 等标准化产品。

三是模型驱动落地四大核心场景，数据赋能产业治理双重增效。以“丰知”大模型与时空智能技术为支撑，数据应用于四大核心场景。交通运输场景融合地址与路况数据，优化“畅行昆山”通行效率；公共安全场景整合地址与事件数据，通过时空智能体提升事件调度转派精度；政务服务场景对接民生诉求与时空数据，赋能“马上办”处置效率；应急处置场景融合标注数据，通过模拟推演生成精准预案。应用成效显著，助力特色产业物流成本降低 20%，库存优化 15%，城市治理预警精准度与处置效率大幅提升。

四是降本增效赋能产业提质升级，创新治理引领城市转型发展。在经济效益方面，助力顺丰年节约运营成本超 2 亿元，提升快递时效与可靠性，赋能昆山特色产业降本增效；在社会价值方面，打造 30 余个智慧城市创新应用，优化民生服务与应急指挥，提升城市治理现代化水平；在产业创新方面，创新政企协同数据开发模式，参编多项国际与国家标准，为数据要素赋能城市转型提供新范式。