

数据产业高层次人才
首席数据官·数据治理工程师研修班

学
员
手
册

目录

致学员的一封信	1
数字菁英评价培训简介.....	3
中部数据流通服务中心简介.....	5
培训纪律要求	7
学员须知	9
评价测试流程	11
课程安排	13
数据产业高层次才学员名单.....	15
课程辅导材料	18
深入贯彻“人工智能+” 建设高质量数据集	18
数据价值实现的湖北实践.....	88
可信数据空间&高质量数据集智能化应用	124
数据治理体系.....	154
数据标准与数据安全	223
企业数字化转型路径.....	259
数据战略与 CDO 职责	295
数据资产（产业）全过程.....	323
政务数据要素化实践.....	340

中医药数据空间与利用	358
中国电子安全可信数据空间建设与运营之路	378

声明：本书所载的全部材料（含视频、课件、演讲实录等）知识产权归活动主办方及合作权利方（演讲者）所有。内容仅用于课程教学及交流，未经书面许可，任何主体不得将相关内容用于商业引流、课程复制、自媒体创作等用途。

致学员的一封信

亲爱的学员朋友：

您好，欢迎加入数字菁英，开启一段关于数据、智能与未来的学习旅程。

在这里，您不只是来听课的，更像是一场身份切换。您可能是一位企业管理者、一名政府信息化从业者，或是一位正在向数据领域转型的专业人士。来到数字菁英的课堂，意味着暂时放下角色与头衔，以学习者的姿态重新出发。

数字世界正在重构产业结构，也在改变我们的思维方式。希望在这段学习时间里，您能在知识与实践之间，找到新的平衡点。每一次课程讨论、每一个案例拆解，都是一次思想的交锋。也许它不会立刻带来答案，但一定会让您更清楚未来的方向。

我们相信，学习的过程就像打磨数据。原始的碎片需要经过整理、分析、验证，才能成为可用的资产。同样，知识也需要积累与沉淀。哪怕每天只理解一个新概念、掌握一个新方法，坚持下去，也会在不知不觉间，构建出属于您自己的“数据能力体系”。

数字菁英网聚集了来自政府、企业、科研机构优秀学员，大家的经验和视角各不相同。希望您能在交流中打开思维，主动分享，勇于提问。学习，不只是听讲，更是一次群体智慧的共鸣。

我们将尽力为您提供专业、务实、前沿的课程内容，也期待您在学习过程中，向我们提出宝贵的意见。您的反馈，是我们持续优化教学、

提升服务的重要依据。

在接下来的学习日子里，请放慢脚步，专注当下。享受与数据同行的每一次思考、每一次灵感闪现。愿这段经历成为您职业旅程中一次重要的能量补给。

让我们一起，用数据的智慧去理解世界，也用学习的热情去定义未来。

祝您在数字菁英的学习时光，充实而有收获。

—— 数字菁英评价培训服务团队 敬上

数字菁英评价培训简介

数字菁英评价培训中心是数字菁英网旗下的数字人才服务平台，专注于助力数字中国的广大从业者实现数字职业技能的全面升级，确保数字中国建设者能够始终维持并持续增强自身于职场中的核心竞争力。数字菁英网由北京数字理政规划设计研究院有限公司（以下简称“数字理政研究院”）成立，聚焦服务数字中国的设计者和建设者，我们视数字菁英为中国数字经济的核心新势力，持续以赋能和联结的能力，推动进步。

数字理政研究院是工业和信息化重点领域人才能力评价的专业服务支撑机构，以数字菁英数据发展大讲堂为依托，深入开展涵盖大数据、人工智能、数据要素等多领域数据产业高层次人才能力评价培训服务。

研究院是全国数据标准化技术委员会（SAC/TC609）数据治理标准工作组成员单位，参与开展数据治理、数据产权、数据合规交易以及数据流通安全保障标准制修订工作，推动相关标准宣传推广及应用实施。研究院是工业和信息化部人才交流中心的授权合作单位，肩负起推动大数据、人工智能等数字人才相关工作的推广与落地实施的重任；研究院是中国电子技术标准化研究院（工业和信息化部电子工业标准化研究院）的授权合作单位，负责数据资产管理、数据评价等数据要素人才的全方位推广与切实落地工作，为数字人才体系的构建与完善发挥着关键且不可或缺的作用。

工业和信息化部人才交流中心创立于 1985 年 1 月，是中央机构编制委员会办公室批准成立、国家事业单位登记管理局登记、工业和信息化

部直属的公益二类事业单位，是工业和信息化部从事人才研究、人才培养、人才评价、人才服务、国际合作等方面工作的专业机构。

数字菁英评价培训中心正在面向全国推广数据产业领域（数据治理工程师、数据管理工程师、数据标注工程师、注册数据安全工程师、数据分析工程师、数据应用工程师、数据资产管理运营师、数据资产规划师、数据资产评估计价师、首席数据官（CDO）），人工智能产业领域（智能体工程师、生成式 AI 应用师、人工智能提示工程师、机器学习算法研发工程师、自然语言处理算法研发工程师、AI 行业应用工程师、机器学习平台研发工程师、生成式 AI 设计师、智能芯片架构设计工程师、数据标注与高质量数据工程师）等数字人才培养。

诚邀数据产业、人工智能产业合作伙伴，联合开办数据、人工智能产业高层次人才系列研修班，共同推动数字人才能力提升。

合作咨询电话，010-58441538。

中部数据流通服务中心简介

湖北大数据集团总体介绍

湖北大数据集团有限公司（以下简称“集团”）是全省深化国有“三资”改革，推动大财政体系建设背景下组建的第12家省属国企。集团以建设“全国一流的省级数据集团和全国一流的数据要素综合服务商”为发展目标，定位为“数字湖北”“数智湖北”建设的推动者、公共数据资源平台构建者、数字产业发展引领者和实数融合赋能践行者。集团围绕夯实数字湖北“一个底座”、形成数据资源“一张网络”、构建数据交易“一个枢纽”、创立数智化应用“一个标杆”、打造数据产业生态“一个模式”五大战略使命，聚焦数字基础设施投资建设、全省数据资源开发利用、政企数字化转型、数据产业投资运营和全省数据要素交易流通运营五大主业。

“中部数据流通服务中心”介绍

“中部数据流通服务中心”于2024年9月正式揭牌成立，定位为省级数据流通综合服务机构，由湖北省大数据集团投资设立，由省数据局负责业务指导与监督。中心设立有六大功能中心，覆盖产权登记、流通交易、价值实现、综合评估、跨境流通以及数据人才培育，全面支撑湖北数据要素的市场化配置与价值释放。中心采用“线上淘宝网+线下广交会”的创新模式，致力于打造一个“立足湖北、依托武汉、辐射中部、面向全国、链接海外”的数据交易平台。线上，我们打造“湖北省数据

流通交易平台”，实现数据产品的高效展示与交易；线下，设立实体数据交易大厅，促进供需对接与深度合作。

通过践行数据流通交易全流程，中心累计数据交易总额超 4.8 亿元，平台入驻数商 1100 余家，上架产品超 1000 个、挖掘和采集了 242 个数据需求，数据流通交易逐渐活跃。中心挖掘数据场景 50 个，形成标杆场景 2 个，场景覆盖文化旅游、智能制造、科技创新、商贸流通等各行各业。此外，中部数据流通服务中心积极发挥生态培育功能，培育了第三方服务机构 151 个，登记在册的经纪人 20 人，专业人才在供需之间搭起桥梁，让数据供需精准匹配；目前在屏幕下方滚动的是我们平台目前的数据需求、数据产品及成交订单，这些成交订单的背后是中心建立的信任机制，因为数据交易的核心困难是交易双方的不互信，因此中心通过第三方合规、价值、质量、安全评估破解供需双方互信难题，跑通“运行机制”、联通“生态网络”、构建“信任机制”，中部数据流通服务中心交易额稳步攀升。

截至目前，中部数据流通服务中心已助力湖北交投、湖北港口、湖北联投、九州通集团、宜昌产投、襄阳国投、咸宁城发、黄冈国投、光谷建设、光谷交通、洪山科投、武汉车网公司、武昌首义科创投、武汉家家养的牛、德韩口腔医院等企业完成数据资产入表案例超 40 个，累计入表金额超 6000 万，平台累计交易额超 4.5 亿元。

培训纪律要求

为确保培训工作顺利开展，全体学员必须遵守培训班纪律规定和要求。

一、政治纪律

全体学员不得宣扬和发表违背党的理论路线方针政策、违背中央决定决议的错误观点和言论，在思想上、政治上、行动上必须同党中央保持高度一致，牢固树立“四个意识”，坚定“四个自信”，坚决做到“两个维护”。

二、学习纪律

培训期间，严格遵守日程安排，不得擅自做与培训无关的事情，要做到令行禁止。自觉将手机调整为静音或振动状态，上课时不得在教室里接打电话，不得随意走动、交头接耳，对授课教师的讲课内容和个人观点不得断章取义、妄议事非。学员应关注课程调整、教室变更等班级动态，时常保持优良的学习作风。

三、生活纪律

爱护校园环境，爱护公共财物，维护校园文明建设，保持教学、住宿和就餐环境的卫生，不随地乱扔垃圾，养成文明行为。文明用餐，顺序排队，不大声喧哗，厉行节约，反对浪费。严格落实中央八项规定精神，不得相互请吃，严禁饮酒。

四、保密纪律

教师的讲课资料、课件等仅限于培训期间学习、交流，禁止上传到公共信息网络。妥善保管好《学习手册》、资料及个人财物等，发现泄密隐患，必须及时报告班主任，妥善处理，情节严重的追究相关责任。

五、考勤纪律

严格考勤制度。培训期间，学员须提前 10 分钟进入教室，每次课前务必签到考勤，并严格按座次表对号就座，不得自行调换座位，不得迟到早退、无故缺席。

严格请假制度。学员确需请假，必须写出书面假条，交带队领导批准；在准假时间内归校，并以书面或者电话方式向带队领导销假。

六、安全纪律

贵重物品和证件要妥善保管好，外出注意交通安全和人身安全。

学员须知

1. 培训时间、地点

时间：2025年10月24日-10月26日。

培训地点：湖北大数据集团小洪山数据产业园裙楼1层会议室。

报到地点：湖北大数据集团小洪山数据产业园裙楼1层会议室。

2. 培训说明

- 1) 10月24日-10月26日培训期间，请学员提前10分钟入场。
- 2) 请学员保管好随身贵重物品，外出时注意安全。
- 3) 注意事项：
 - (1) 请佩戴学员证参加培训，凭学员证出入教室；
 - (2) 24日-26日在武汉小洪山数据产业园，须遵照武汉小洪山数据产业园管理规范；
 - (3) 上课前签到；学员每天上课不迟到、不早退，原则上不准请假；
 - (4) 学员上课时保持安静，关闭手机或置于静音状态，课堂内严禁吸烟；
 - (5) 上课期间，**禁止用录音、录像等方式**复制传播讲师课程信息；
 - (6) 研修班期间学员要注意自己的人身和财产安全，如发现有威胁到自身安全的行为或发现可疑情况和可疑人员，要及时向会场或会务组工作人员反映，以保护自身合法权益不受损害。
 - (7) 研修班其间，如身体不适，请及时联络会务组或会场。

3. 用餐及住宿

培训期间，交通食宿自行安排、费用自理。

4. 会务组工作人员及联系方式

姓名	手机	单位
郭远洋	18612596975	数字菁英网
汪亦涛	13807227328	湖北大数据集团

祝大家在武汉小洪山数据产业园培训期间，心情愉快，身体健康，学有所获！

评价测试流程

- ① 注册、实名认证 (<https://pj.miitec.cn/index>)；注册时需要 1 寸免冠、白底电子照片，300dpi、像素 413x295 (长×宽)，JPG 格式；身份证正、反面电子版，JPG 格式，要求清晰可识别；如果实名认证【未通过】请及时修改相关信息，或联系 13220158222 处理；
- ② 10 月 25 日 23:59 前，完成测试报名；
- ③ 10 月 26 日 13:40-14:00，电脑端登录测试系统 (同①)；
- ④ 10 月 26 日 14:00-16:00，完成由工业和信息化部人才交流中心组织的专业知识测试 (网址：<https://pj.miitec.cn/index>)。
- ⑤ 10 月 26 日 16:00-17:00，完成由数字理政组织的技术技能测试 (网址：<https://www.digitalelite.cn/>)。
- ⑥ 专业知识测试、技术技能测试结束；
- ⑦ 2025 年 11 月 26 日前完成工程实践测试及答辩；
- ⑧ 测试总成绩 (专业知识分+技术技能分+工程实践分) 超过 60 分，由工业和信息化部人才交流中心颁发相应证书。

课程安排

1. 组织机构

- 主办单位：北京数字理政规划设计研究院有限公司
(IITC 工信人才岗位能力评价合作机构)
- 能力评价单位：工业和信息化部人才交流中心
(IITC 工信人才岗位能力评价机构)
- 承办单位：数字菁英网/智绘九州(北京)科技有限公司
- 支持单位：湖北大数据学院
国信人才教育科技(湖北)有限公司
中国软件评测中心华南分中心
- 协办单位：上海数素管理咨询有限公司

2. 议程安排

时间	主题	主讲人
10月24日 主持人：徐周亚洁，湖北大数据学院		
08:00-08:55	签到	
09:00-09:20	开班仪式	领导致辞
09:20-12:00	深入贯彻“人工智能+”建设高质量数据集	闫树，中国信息通信研究院
12:00-14:00	午餐、午休	
14:00-15:45	数据价值实现的湖北实践	古伟，湖北大数据集团中部数据流通服务中心
	小洪山数据产业园参观	
15:45-17:30	可信数据空间	胡俊，中国电子数据产业集团
	高质量数据集智能化应用	
17:30-19:00	晚餐、休息	
19:00-21:00	数据发展研讨会（公开直播）	
	人工智能及大数据在检验医学的实际应用	狄翰钦，森栩医学/学员代表

时间	主题	主讲人
	蚂蚁密算可信数据空间支撑数据要素场景应用的实践案例	李辉，蚂蚁密算/学员代表
	基于政务数据空间的海南应用实践探索	吴敏，海南欣方/学员代表
10月25日 主持人：洪一		
09:00-12:00	数据治理体系	曾新科，数字理政 IITC 工信人才首席数据官讲师
12:00-13:30	午餐、休息	
13:30-14:30	数据标准与数据安全	何旭珩，数字理政 IITC 工信人才首席数据官讲师
14:30-16:30	企业数字化转型路径	陈立，数字理政 IITC 工信人才首席数据官讲师
16:30-17:30	数据战略与 CDO 职责	马欢，数字理政 IITC 工信人才首席数据官讲师
17:00-19:00	晚餐、休息	
19:00-20:30	数据资产（产业）全过程	王毅，数字理政 IITC 工信人才首席数据官讲师
20:30-21:00	评价测试工作安排	郭远洋，数字理政
10月26日 主持人：郭远洋		
09:00-10:00	政务数据要素化实践	刘方哲，深圳市龙华数据有限公司
10:00-12:00	中医药数据空间与利用	尤良震，北京中医药大学数智中医慢病防治北京市重点实验室
	破局与实践：中国电子安全可信数据空间建设与运营之路	刘颖，中国电子数据产业集团
12:00-14:00	午餐、休息	
14:00-17:00	首席数据官（高级）测试	工业和信息化部人才交流中心 数字理政
	数据治理工程师（高级）测试	

数据产业高层次才学员名单

(排名不分先后)

学员姓名	单位名称	岗位
陈钢	上海科泰信息技术有限公司	首席数据官（高级）
叶彤	陕西蓝匠数智信息技术有限公司	数据治理工程师（高级）
黄胜标	恒生电子股份有限公司	首席数据官（高级）
梁延龙（线上）	中电云计算技术有限公司	首席数据官（高级）
任佳佳（线上）	四川久远银海软件股份有限公司	数据治理工程师（高级）
邵元奇	湖北交通投资集团有限公司	数据治理工程师（高级）
姬鹏	湖北交通投资集团有限公司	首席数据官（高级）
陆由	湖北交通投资集团有限公司	首席数据官（高级）
阳雄伟（线上）	湖南天湘和信息科技有限公司	数据治理工程师（高级）
龙水良（线上）	湖南天湘和信息科技有限公司	首席数据官（高级）
姚湘军（线上）	湖南天湘和信息科技有限公司	首席数据官（高级）
王子腾	成都数之理科技有限责任公司	数据治理工程师（高级）
刘玉红	上海信投智能科技股份有限公司	首席数据官（高级）
马鹏	甘肃省计算中心	数据治理工程师（高级）
章恒	甘肃省计算中心	首席数据官（高级）
李志杰（线上）	新疆数绘未来物流有限公司	数据治理工程师（高级）
朱忠辉（线上）	新疆数绘未来物流有限公司	首席数据官（高级）
方金旗（线上）	浙江智绘科技有限公司	数据治理工程师（高级）
姚春达（线上）	浙江智绘科技有限公司	数据治理工程师（高级）
杨馥铭（线上）	浙江智绘科技有限公司	首席数据官（高级）
杨镇蔚（线上）	浙江智绘科技有限公司	首席数据官（高级）
杨卫勇（线上）	重庆首讯科技股份有限公司	首席数据官（高级）
李杰	五格（宁夏）科技发展有限公司	首席数据官（高级）
高德棚	苏州工品一号网络科技有限公司	首席数据官（高级）

首席数据官 GOVCDO

学员姓名	单位名称	岗位
董学英（线上）	董学英	首席数据官（高级）
陶永峰（线上）	绍兴市智慧城市集团有限公司	首席数据官（高级）
黄赞江（线上）	申合信科技集团有限公司	首席数据官（高级）
袁阔（线上）	支付宝（杭州）数字服务技术有限公司	首席数据官（高级）
曹喆	上海森栩医学科技有限公司	首席数据官（高级）
张俊江	上海森栩医学科技有限公司	首席数据官（高级）
狄翰钦	上海森栩医学科技有限公司	首席数据官（高级）
霍玉雯	中国船舶集团有限公司第七二五研究所	数据治理工程师（高级）
吴敏	海南欣方网络信息科技有限公司	首席数据官（高级）
李明春	中电信数智科技有限公司	首席数据官（高级）
李凯	北京北控数字科技有限公司	首席数据官（高级）
王若	北京北控数字科技有限公司	首席数据官（高级）
刘方哲	深圳市龙华数据有限公司	首席数据官（高级）
尤良震	北京中医药大学东直门医院	数据治理工程师（高级）
吴彬	成都数之联科技股份有限公司	首席数据官（高级）
何鸿彬	成都数之理科技有限责任公司	数据治理工程师（高级）
马睿良	成都数之联科技股份有限公司	首席数据官（高级）
古伟	湖北大数据集团有限公司	首席数据官（高级）
郭远洋	智绘九州（北京）科技有限公司	首席数据官（高级）
何旭珩	数字理政研究院	首席数据官（高级）
曾新科	数字理政研究院	首席数据官（高级）
贾龙	涿溪（山西）工业互联网科技有限公司	首席数据官（高级）
陆洵蔚	上海信投智能科技股份有限公司	首席数据官（高级）
徐杨	湖北云启互联技术发展有限公司	首席数据官（高级）
吴鑫	宜昌云启互联技术中心有限公司	首席数据官（高级）
陈立	数字理政研究院	首席数据官（高级）
刘颖	中国电子数据产业集团	首席数据官（高级）
白钰（线上）	北京市生态环境局综合事务中心	首席数据官（高级）
王晓爽（线上）	北京市生态环境局综合事务中心	首席数据官（高级）

首席数据官 GOVCDO

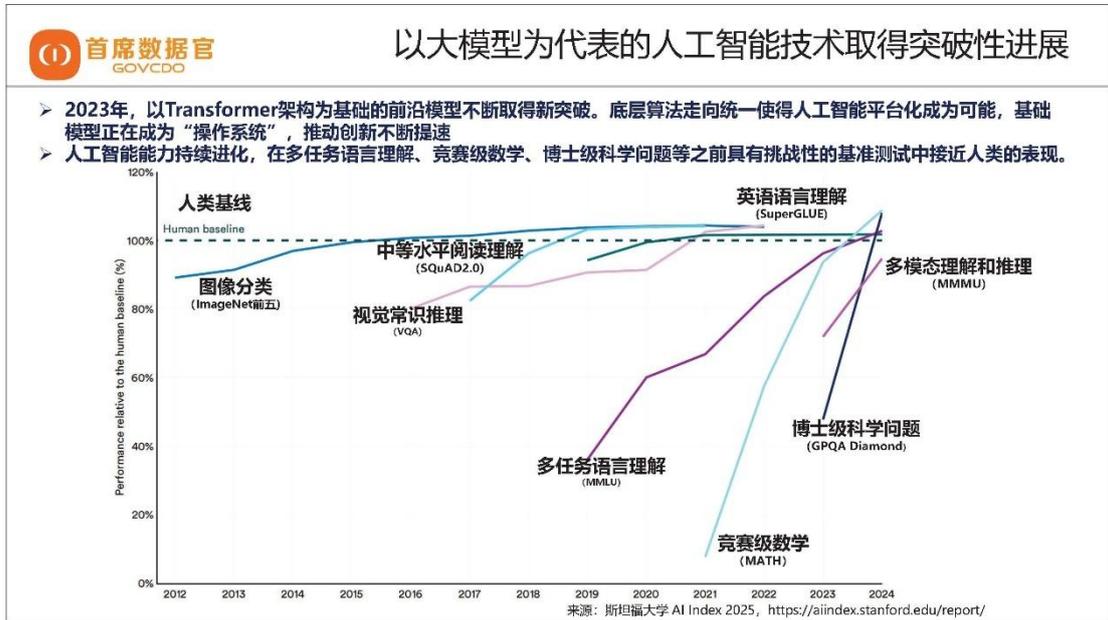
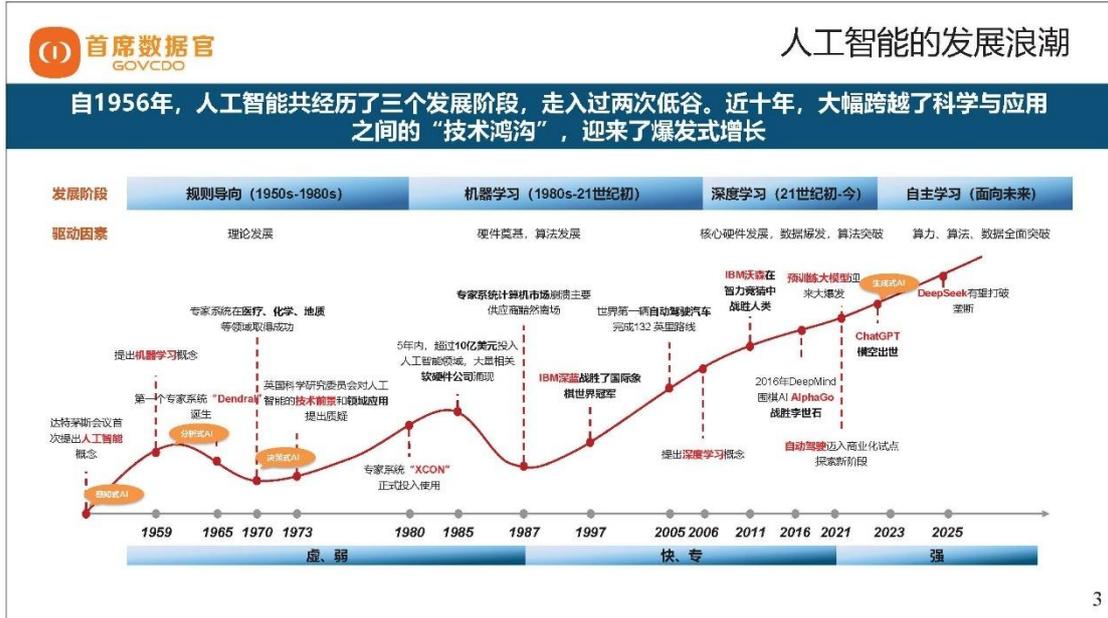
学员姓名	单位名称	岗位
刘川铭	漳州高新区联合科创中心	首席数据官（高级）
熊颜妍	漳州高新区联合科创中心	首席数据官（高级）
王亮	北京中科慧居科技有限公司	首席数据官（高级）

课程辅导材料

深入贯彻“人工智能+” 建设高质量数据集



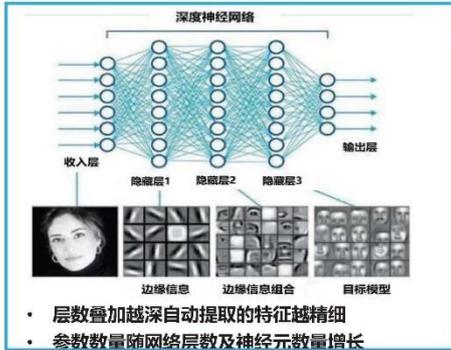
首席数据官 GOVCDO	目 录
<p>一、从“互联网+”到“人工智能+”</p> <p>二、数据是人工智能飞跃发展的关键驱动力</p> <p>三、高质量数据集建设方法论</p> <p>四、高质量数据集建设案例分析</p>	



首席数据官 GOVCD0 大模型核心是深度神经网络机制下的数据学习并形成智能涌现

大模型VS人脑

(人的大脑860亿神经元, 估计千万亿突触/参数)



- 层数叠加越深自动提取的特征越精细
- 参数数量随网络层数及神经元数量增长

深度视觉网络Alexnet 6000万参数 → GPT4 120层网络结构、1.8万亿参数

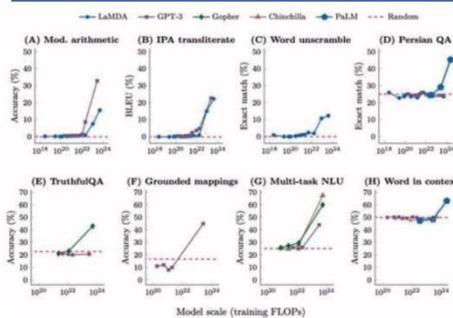
- **生成式人工智能**架构是做联合概率分布分析 (数据中多变量组成向量的概率分布, 对已有数据进行归纳总结, 生成新的内容), 从而对给定序列中的下一个单词进行预测。猜词越准, 意味对世界的把握越精确
- **提示工程 (prompt)** 的作用是在大量的数据关系中引导至与用户需求相关的方面, 激发模型潜力
- **人类反馈强化学习**是解决幻觉、强化事实理解的关键
- **思维链**提示学习技术是提升逻辑推理的重要探索

- 本质上大模型通用能力的形成是从数据中学习到的, 数据反映了物理世界隐含的规律和知识, 大量信息的大规模并行处理, 涉及大量的弱相关变量
- 从细节 (数据) 中提取大量的信息整合到神经网络中调整权重, 从而将现实世界中的各种微妙差别进行编码, 解析和理解物理世界, 并学习和继承大量数据背后的知识、规律和推理能力

首席数据官 GOVCD0

大模型的智能涌现

大模型的涌现现象



OpenAI (2020年) 和谷歌DeepMind (2022年) 先后发现大语言模型存在**扩展定律 (Scaling Law)** :

当独立增加训练数据量、模型参数规模或训练计算量, 模型误差都会降低, 也就是说模型效果越来越好。

多模型在规模达到一定程度时, 在准确度等某些性能指标上会出现急剧的变化, 甚至模型会突然有能力解决在规模较小时完全无法解决的问题。

涌现:

系统中的个体遵循简单的规则, 通过局部的相互作用构成一个整体的时候, 一些新的属性或规律会突然在系统层面诞生—当模型参数不断增长, 到达一定阈值后, 模型任务性能呈现“量变到质变”提升, 大幅优于较小模型性能



Scaling Laws (缩放定律)：大意味着更多的世界知识

Scaling Laws：在 Transformer 架构下，通过提升模型参数量、数据量以及计算量，能够有效提升模型性能表现，是引导算法模型迈向通用智能的重要规律

模型的表现能力取决于 **数据、参数规模和计算资源**三大关键要素

$$\text{模型能力} = \mathcal{F}(\text{数据}, \text{参数规模}, \text{计算资源})$$

对于Decoder-only的模型，计算量C, 模型参数量N, 数据token D, 三者的关系为: $C \approx 6ND$

在模型规模及算力固定的前提下，可通过提升数据的**质和量**来加强人工智能大模型的训练效果。
模型训练达到最优，**模型大小和训练数据规模应等比例缩放**
数据的多样性可以提高模型的性能

Transformer算法+GPU集群，支撑千亿级连接的巨型神经网络的高效训练，为复杂系统出现智能涌现现象提供了强大的物质基础。



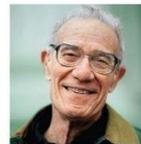
生产力突破的初现与期望

人工智能会做什么？

- **会看**：图像识别、文字识别、车牌识别
- **会听**：语音识别、说话人识别、机器翻译
- **会说**：语音合成、人机对话
- **会行动**：机器人、自动驾驶汽车、无人机
- **会思考**：人机对弈、定理证明、医疗诊断
- **会学习**：机器学习、知识表示



莫拉维克悖论：人类所独有的高阶智慧能力只需要非常少的计算能力，例如推理，但是无意识的技能和直觉却需要极大的运算能力——**复杂的事情容易做，简单的事情难做好**



索洛悖论：You can see the computer age everywhere but in the productivity statistics.我们到处都看得见计算机，就是在生产率统计方面却看不见 ——罗伯特·索洛，1987

- 人工智能赋能行业从数字化水平较好的领域率先突破，再逐步扩散到更多行业。
- 以超级应用与消费互联网为代表数字原生领域，凭借天然的数据沉淀优势、成熟的数字基础设施以及海量的用户基础，率先实现规模化落地
- AI for Science (AI4S)、AI带动的软件革命、自动驾驶等领域的显著突破.....
- **本轮人工智能浪潮不会重蹈的过去的轨迹**

如何促进生产力革命性跃迁和生产关系深层次变革？

我国进展：国内模型能力与国外差距进一步缩小

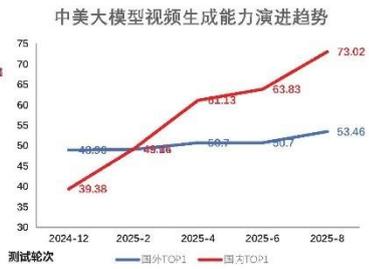
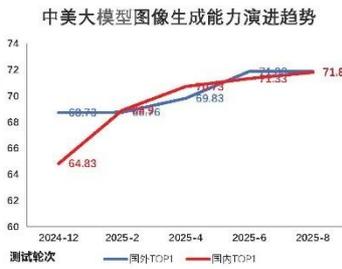
语言模型



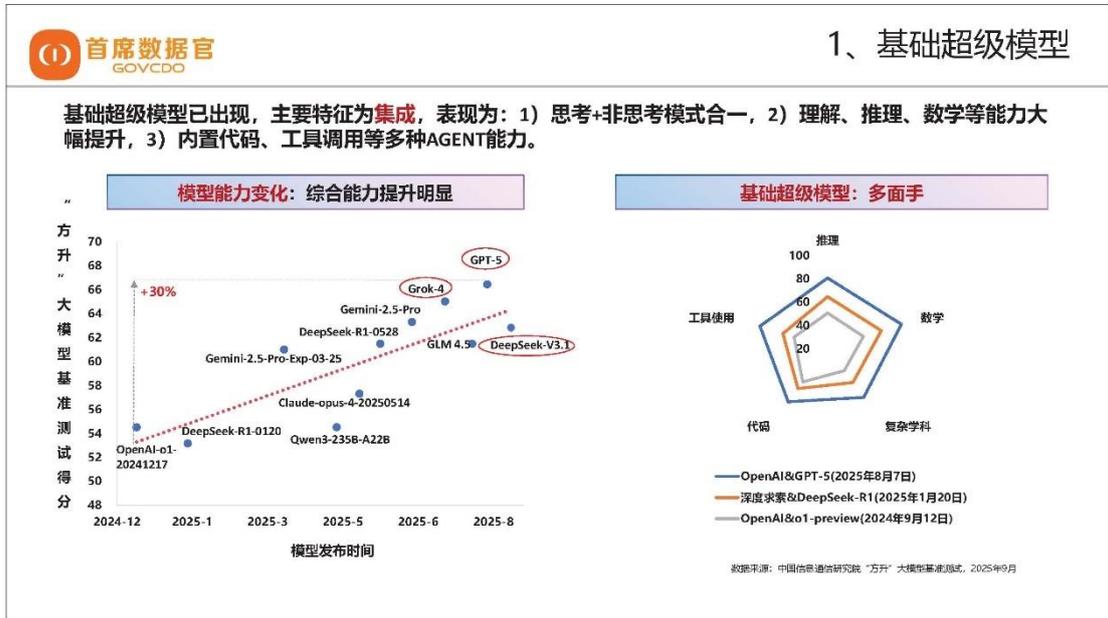
来源：中国信通院“方升”大模型测试结果（2025年9月）

多模态模型性能测试结果

多模态模型



来源：中国信通院“方升”大模型测试结果（2025年9月）

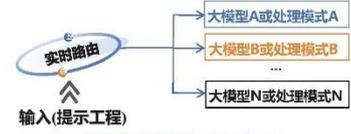




首席数据官
GOVCDO

1、基础超级模型

技术特点1: 通过路由融合集成多种关键能力



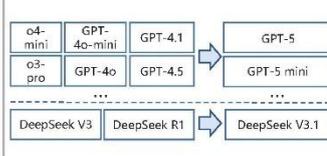
大模型融合路由器 (Router)

技术特点2: 通过在环境中强化学习提升真实场景表现

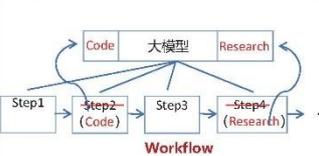


面向智能体的强化学习 (Agentic RL)

影响1: 模型的使用更便利



影响2: 自主决策 workflows 和外部工具调用



Workflow

影响3: 对训练数据供应提出新要求

互联网采集数据

- CommonCrawl
- C4
- ROOTS
- Pile
- Github
- ...

AGENT应用采集数据

- AgentInstruct
- InfoSeek
- ...
- 物理场景采集数据
- InternScenes
- AgiBot World
- ...

来源: G Zhang, H Gong, X Yu, et al. The Landscape of Agentic Reinforcement Learning for LLMs. A Survey, arxiv:2509.02547, 2025.
Y Zhang, H Li, J Chen, H Zhang, P Ye, L Bai, S Hu. Beyond gpt-5: Making llms cheaper and better via performance efficiency, optimized routing. arxiv:2509.02547, 2025.



首席数据官
GOVCDO

2、自主性更强的智能体

智能体进一步释放大模型应用潜能, 自主性不断增强, 成为“数字员工”雏形。

智能体释放模型应用潜能

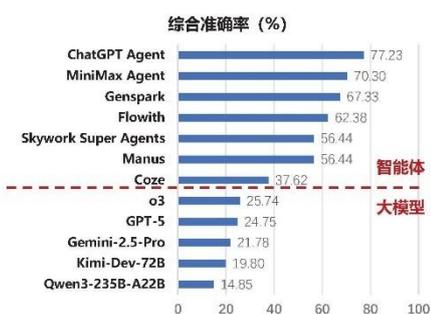


智能体基础应用架构

智能体相比大模型可以自主完成复杂任务, 但仍不及预期

高度封装的智能体产品可以获得比单一模型更好的性能表现。

综合准确率 (%)



模型名称	综合准确率 (%)
ChatGPT Agent	77.23
MiniMax Agent	70.30
Genspark	67.33
Flowith	62.38
Skywork Super Agents	56.44
Manus	56.44
Coze	37.62
o3	25.74
GPT-5	24.75
Gemini-2.5-Pro	21.78
Kimi-Dev-72B	19.80
Qwen3-235B-A22B	14.85

测试任务涵盖网页交互、复杂信息挖掘、多模态理解、文档处理等4大类101个具体任务。
数据来源: 中国信通院“方舟”智能体基准测试, 2025年9月

智能体是数字员工的初级形态

消费级智能体

信息整理 股票分析 旅游攻略 ...

容错率高, 用户可以接受产品一定程度的不
稳定和幻觉现象

企业级智能体

财务报税 生产排程 物流管理 ...

要求“零失误”与稳定, 需要与企业业务系统
深度融合

智能体能力仍有较大提升空间

2、自主性更强的智能体

互联互通和长难任务处理是智能体技术创新主旋律。

热点1：通过智能体协议外扩能通

通信协议成为智能体与外界交互的“桥梁”

MCP和A2A协议

MCP纵向解决单个智能体的知识深度和工具丰富度。GitHub上MCP服务器代码库项目已收录超过9000个。国内百度、阿里、腾讯等均宣布支持。

A2A横向解决多个智能体的互联互通，构建高效协作网络。目前，A2A的应用普及不及MCP。

- 技术层面，智能体通信协议为开发者提供了技术选项，但仍存在一定安全风险。
- 生态层面，智能体通信协议致力于构建良好应用生态，但商业价值仍需进一步实践。
- 未来，智能体通信协议将助力“智能体互联网”构建，高效实现跨域协作。

热点2：提高长时间执行高难度任务的能力

智能体能完成任务的长度大约每7个月翻一番。

未来，智能体需要能够独立完成需要人类数天或数月才能完成的任务。

- 可行性探索：以Kimi-Researcher为例，其通过在线On-Policy强化学习与长上下文管理提高智能体的长程任务解决能力。

参考：<https://a2a-protocol.org/>, <https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/>

3、走向实训的具身智能

具身智能机器在运动场上实现运动能力突破，在训练场中学习作业技能，为在真实生产环境中应用打好基础。

2024 走进实验室 实验室里创新技术 宇树人形机器人在 研究和娱乐领域 完成首批交付，累计 超1500台	2025 奔向运动场 竞技场上突破遥控 天工 Ultra 跑步 从半自主到全自主 宇树 G1 拳击 从频繁扑空、踉跄到灵活高难度腿部动作	进入训练场 训练场中学习技能 我国建设完成的训练场在 25年增加了12个 已部署本体 约千台	行业场景试点 场景中验证迭代 富临精工+智元机器人 远征A2-W 美国+银河通用 Galbot G1
--	--	--	--

具身模型取得一系列新进展，与本体的结合初步实现现实任务执行

典型模型	时间	路线	本体	任务
Figure AI Helix	2025.02	VLM+VLA	Figure 02	物流分拣、叠毛巾、洗碗
智元机器人GO-1	2025.03	VLA	智元精灵G1	擦桌子、倒水、货架整理等
智平方 GOVLA	2025.04	VLM+VLA	AlphaBot 2	仓储物流、上下物料、零部件装配
蔚来汽车NWM	2025.06	世界模型	Banyan 榕车型	自动驾驶

仍受限于特定任务，泛化性脆弱，长程任务稳定性不足

* 研究论文：A2A-Flow <https://arxiv.org/abs/2503.18881>
 * 论文：A2A-Flow: A2A-Flow <https://arxiv.org/abs/2503.18881>
 * 研究论文：A2A-Flow: A2A-Flow <https://arxiv.org/abs/2503.18881>
 * 研究论文：A2A-Flow: A2A-Flow <https://arxiv.org/abs/2503.18881>
 * 研究论文：A2A-Flow: A2A-Flow <https://arxiv.org/abs/2503.18881>

3、走向实训的具身智能

具身智能要从“实训”走向“实战”，面临三大挑战

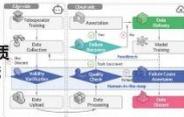
挑战1：高质量数据缺

智能涌现需要的数据量在数百万小时，当前仅有几万小时

关键点一：需要什么训练数据？

预训练：大规模、多样任务数据
后训练：高质量、特定任务、精标数据

智元机器人 Agibot World 采用人在环的严格质检和精细标注管线



挑战2：模型泛化难

任何在训练数据分布外的场景任务，都会导致泛化能力下降

关键点二：如何加速跨场景、跨任务迁移？

更通用的具身基础模型

双机协作 叠衣服 叠毛巾 洗碗
物流分拣 洗衣服

迭代时间
2.21 7.31 8.13 9.4

Figure AI Figure 02+Helix
基于Helix通用基础模型，模型不变，仅增加新数据，便可快速适配新场景和新任务

挑战3：软硬协同难

模型与本体需要在整个链路上有效配合，任何一环的不稳定，都将导致任务失败

关键点三：如何打通软硬工程链路？

更统一的开发平台

“大脑”与“小脑” → 跨本体链接协议 高效协同

Tencent 腾讯 具身智能开放平台Tairros
华为云 华为云 R2C
睿-HUMANOID 北京人形机器人创新中心 慧思开物平台
达摩院 DAMO ACADAMY RynnRCP

参考：智元机器人 2022年12月 Bu Q, Cai J, Chen L, et al. Agibot world colosseum: A large scale manipulation platform for scalable and intelligent embodied systems// arXiv:2310.06662, 2023
Figure AI 2023年12月 https://www.figure.ai/news/helix-leads-the-dishwasher

4、萌芽中的世界模型

现在的大模型技术存在局限，产业界期待发展泛化能力强、物理一致性高、可解释性强的“世界模型”，满足自动驾驶、具身智能等现实需求，也架起通向通用智能的桥梁。

当前大模型尚存局限性

握物理规律欠缺

- 物理规律理解弱
- 无法形成连贯、整体认知

因果推理欠缺

- 无法理解事件“为什么”发生
- 无法理解事件“如何”发生

世界模型逐渐受到产业重视



世界模型核心能力需求

世界模型应具备数据生成、动作解释、环境交互、场景重建能力。



典型应用
自动驾驶 具身智能 游戏 ...



首席数据官
GOVCDO

4、萌芽中的世界模型

世界模型刚刚萌芽，技术探索百花齐放，主要包括**大模型增强**、**大模型与物理引擎融合**、**物理世界表征**等路线，仍处于发展初期，空间广阔但面临不少挑战。

世界模型技术路线百花齐放

大模型增强

融合Transformer与Diffusion，世界模型正成为构建数字世界的“模拟器”

Sora → OpenAI

DeepMind → Genie3

Cosmos → NVIDIA

Tencent → HunyuanWorld

NWM → 蔚来nio

大模型+物理引擎

锚定物理规则，赋能高可靠生成，打造可解释的世界预测范式。

Marble → World Labs

Tencent → HunyuanWorld-Voyager

Cosmos Transfer → NVIDIA

Carnegie Mellon University → Genesis

乾崮ADS 4.0 → HUAWEI

世界表征等探索

抛弃像素级生成，专注于学习世界的抽象表征与预测潜在状态。

JEPA → Meta

I-JEPA

V-JEPA

V-JEPA2

世界模型面临挑战

当前世界模型仍面临如下方面挑战。

概念定义存在争议

“世界”本身是一个高度抽象的概念，不同领域的研究者对其范围和边界的理解不同。

技术路线不清晰

目前世界模型的实现方法包括生成式模型、强化学习、多模态融合等，不同技术路径的侧重点和适用场景不同。

应用范围局限

当前世界模型只在自动驾驶领域有大规模应用，由于技术瓶颈和成本限制，其他领域仍未大规模应用。

©K. M. Asan et al., "V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning," arXiv preprint arXiv:2506.09965, 2025.



首席数据官
GOVCDO

5、AI正在重塑软件

2025年上半年我国软件业务收入7.06万亿元，同比增长11.90%*。大模型**向软件研发全过程渗透**，正深刻**改变软件开发工具和形态**，开启AI重塑软件业的新阶段。

AI变革软件研发全过程

国内外大模型用于编程任务的Token调用显著提升，且向软件研发全生命周期深度渗透

- 豆包日调用量16.4万亿Token，较2024年12月增长4倍，其中用于AI编程调用量增长8.4倍
- Anthropic Claude用于编程任务占比36.9%

AI技术应用占比在需求、运维增幅达10%左右



AI生成的代码占比均值为28.17%



数据来源：中国信通院，2025年1月

AI变革软件研发工具

2025年AI IDE/代码智能体密集发布，推动氛围编程发展，人人都可成为开发者



插件+IDE		AI IDE/智能体	
发布时间	产品名称	公司名称	产品类别
2023年	Cursor	Airshipare	AI IDE
2025年1月	Tuo	宇中数智	AI IDE
2025年2月	Glaive Code	OpenAI	CLI
2025年4月	Codex CLI	Arifinogic	CLI
2025年5月	Lingma IDE	阿里云	AI IDE
2025年8月	Comate IDE	深虑	AI IDE
2025年8月	Sorona CLI	Google	CLI
2025年7月	CodeBuddy IDE	腾讯	AI IDE
2025年9月	CodeBuddy Code	腾讯	CLI

数据来源：中国信通院统计，2025年9月

AI变革软件研发产品形态

交互形式更简洁，复杂UI将成为过去式

传统软件：界面按钮多嵌套复杂
未来软件：对话交互、多模态交互、具身智能等方式

认知智能化，自主决策自学习软件

传统软件：规则式代码，流程自动化
未来软件：认知式代码，认知维度全面升级

- Gartner 预测：到2028年，将有33%的企业软件包含生成式AI能力，而2024年仅有1%
- 特斯拉FSD系统，大模型神经网络替换原有30万+行C++代码的90%

数据来源：Update 2024.15.10 (FSD 12.4.2) - Release Notes

*软件业务收入数据来源：工业和信息化部，2025年9月

5、AI正在重塑软件

大模型推动人员、企业、产业的转型，研发人员的角色正在转换，超级个体成为可能，软件企业组织结构、业务场景等将会加快转型，并以此为突破口，极大加速AI向其他行业传导扩散。



6、开放智算生态

智算生态正在加速走向开放，涌现出开源通信库、开源算子库、开源计算平台、开源互联协议等一系列标志性成果，加速国产软硬件生态成熟步伐。

近一年全球智算领域涌现一批开源开放标志性成果

计算机制	NVIDIA · FlashAttention 4	deepseek · NSA
框架	飞桨 · 飞桨3.0	[M] 昇思 · 昇思2.7
算子库	deepseek · DeepGEMM	BAAI 智源研究院 · FlagGems
通信库	deepseek · DeepEP	BAAI 智源研究院 · FlagCX
计算平台	HUAWEI · CANN	
互联协议	HUAWEI · UB-mesh	中国移动 · OISA 2.0
精度量化	NVIDIA · NVFP4	deepseek · UE8M0

开源基础模型有效加快国产软硬件协同步伐

以Deepseek、千问、文心等优质模型开源为契机，算力厂商积极推动软硬件适配，已能够满足实际应用需求

我国厂商产品支持领先开源模型高效部署

从规模成本看，我国已有系统实现单卡8卡推理DeepSeek 671B满血版模型（INT8/FP8精度），与英伟达所需硬件规模持平，部署成本在百万级

国内领先产品模型运行精度已基本与国外持平

从语言理解、逻辑推理等对模型适配后的效果进行评测，通过软硬件协同优化，参测产品部署DeepSeek模型的精度已基本与国外英伟达系统持平

*数据来源：中国信通院



首席数据官
GOVCDO

6、开放智算生态

从趋势上看，开放智算生态已从**单点局部开源**扩展逐步走向**软硬全栈开源开放**，未来不同环节有望更深度的开放合作，形成紧密耦合的**开放协同智算生态**。

初期：局部开放智算生态阶段

当前：全栈开放智算生态阶段

趋势：协同开放智算生态阶段

开源模型

Qwen LLaMA by Meta Gemma

Llama 下载量突破10亿次

开源框架

PyTorch HuggingFace社区85%大模型使用该框架

开源模型 开源框架

开源加速库 开放计算平台

开源算子库 开源通信库 开源互联

基于模型研发需求提出硬件改进建议

智谱·AI deepseek

GLM端侧模型基于昇腾自定义算子，完成模型微调量化

建议集成专用通信协处理器、支持scale up/out融合等*

软硬协同 开放智算

CANN NVIDIA

挖掘开放硬件潜力定制高效算法模型

*参考：Insights into deepseek-v3: Scaling challenges and reflections on hardware for ai architectures, Proceedings of the 52nd Annual International Symposium on Computer Architecture, 2025: 1731-1745.



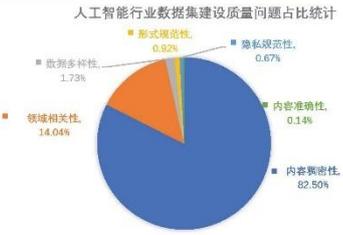
首席数据官
GOVCDO

7、面向行业的高质量数据集

人工智能数据集建设重点从**通用基础数据集**转向**行业高质量数据集**，从“数据为王”到“质量至上”，数据集质量问题成为当前制约行业垂类模型落地和场景应用的核心瓶颈。

聚焦当前需求，通过人工智能数据工程解决当前行业数据质量问题

人工智能行业数据集建设质量问题占比统计



问题类型	占比
内容准确性	82.50%
领域相关性	14.04%
数据多样性	1.73%
形式规范性	0.92%
隐私规范性	0.67%
内容稠密性	0.14%

问题1：内容稠密性

- 缺失内容信息细节及关联关系，且存在相似度较高且重复的样本数据集。

问题2：领域相关性

- 管非所问，未体现行业领域核心概念和深层业务关系。

问题3：数据多样性

- 缺少丰富广泛的行业领域特征，来源单一，类型简单，涵盖领域场景单调。

问题4：形式规范性

- 存在错别字、多余字符、逻辑不符等形式问题，文件命名不符合行业规范

来源：中国信息通信研究院AI数据质量测试结果（数据截止2025年8月）

亟需建立面向人工智能的数据工程体系



7、面向行业的高质量数据集

面向通用人工智能和行业深度赋能，要加快建设适应强化学习、世界模型等前沿技术和行业智能体所亟需的新型高质量数据集，构建**新型人工智能数据供应链**。

面向未来发展：针对强化学习、世界模型等前沿技术和行业深度应用的高质量数据需求不断扩大

面向训练：更好支撑强化学习与智能体训练

- 训练轨迹数据集** → OpenAI发布的Gym+Atari可构建具有离散和连续动作空间强化学习环境 and 轨迹数据
- 偏好对齐数据集** → NVIDIA与ScaleAI联合创建发布的HelpSteer对齐数据集，阿里发布的CValues-Comparison价值观对齐数据集
- 评估基准数据集** → 卡内基梅隆大学发布的Meta-World智能体基准评估数据

面向行业：更好支撑AI原生应用

- 基础支撑数据** → 行业微调数据、行业知识库与知识图谱、基础检索数据库、内部日志记录
- 过程埋点数据** → 时间戳、行为路径记录、过程日志、资源消耗数据、反馈数据、状态变化数据
- 外部交互数据** → 用户输入输出、第三方调用数据库、外部环境动态数据、设备状态更新数据

参考：Machado, M. C., Bellemare, M. G., Bowling, M., & Ferret, N. (2018). Revisiting the Arcade Learning Environment: Evaluation Protocols and Open Problems for General Agents. Journal of Artificial Intelligence Research, 61, 523-562. NVIDIA, HelpSteer: Open-source dataset for training top-performing reward models. https://huggingface.co/datasets/nvidia/helpsteer. Published 2024.

8、开源成为标配

DeepSeek等的开源，改变大模型发展轨迹，现在已经成为行业标配。

开源Infra工具链不断完善

反哺算力基础设施

推动云与芯片需求增长
促进技术体系创新升级

开源模型创新活跃 性能比肩闭源模型

开源模型下载量

国产开源模型的海外累计下载量已突破3亿次

开源模型微调占比

基于国产模型微调的衍生开源模型占比达47%

开源模型性能

在全球性能前25的大模型中，国产开源模型占9席

开源智能体框架不断丰富

加速行业应用落地

降低AI应用落地门槛
赋能千行百业数智化转型

8、开源成为标配

我国人工智能开源生态规模稳步扩大。开源不仅重塑我国AI产业生态结构、推动商业模式创新，也将更深刻影响全球技术标准格局的形成。

开源社群协同演进 推动技术普惠发展

昇思社区 AI软硬件工具链	飞桨社区 深度学习框架	启智社区 中国算力网普惠算力	麒麟社区 开源模型国产适配	ModelScope 最大最全国际化	Gitee AI Gitee最大托管平台	GitCode CSDN最大开发者社区	焕新社区 央国企 AI+行业应用	OpenLab 全链条开源开放体系	鲸智社区 大模型公共服务平台
------------------	----------------	-------------------	------------------	-----------------------	-------------------------	------------------------	---------------------	----------------------	-------------------

截至2025年9月，我国主流AI开源社区已托管模型达38万个、数据集5.3万个，分别相当于国际平台HuggingFace同期托管数量的15.3%和9.5%。此外，我国活跃AI开源开发者数量达2.2万人，约占全球开源AI开发者总数的18.7%。



9、缓解模型幻觉

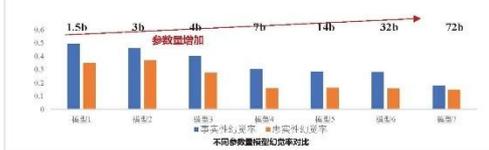
伴随大模型的落地应用，幻觉引发的风险日益凸显，越来越成为制约应用的主要障碍之一。现有技术架构下，大模型幻觉是其固有属性，成因复杂，训练数据、训练方法、推理过程、评估指标等因素交织。

“幻觉”大模型统计学习必然产生的副产品*

尽管推理是缓解幻觉的一种方案，但测试显示其幻觉仍保持在10%以上



参数量越大幻觉率越低，但大参数量的模型幻觉问题仍较为明显



*数据来源：中国信通院 AI Safety Benchmark, 2025年

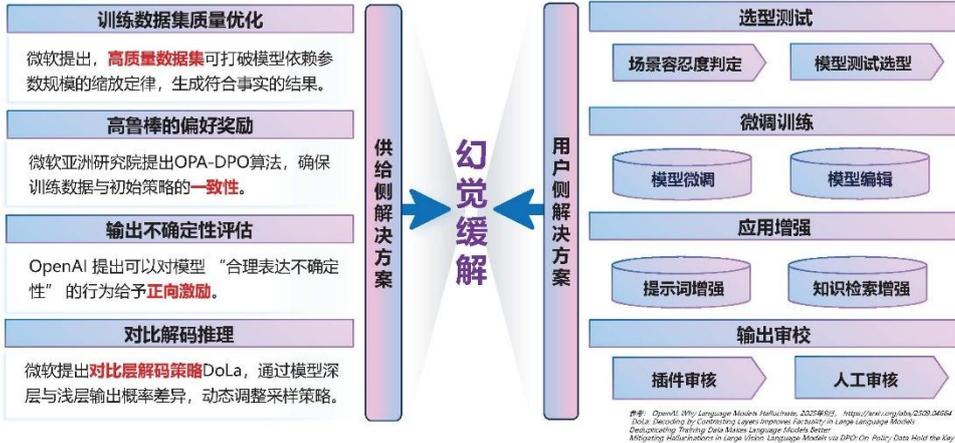
幻觉成因复杂交织

- 数据 (原材料)**
 - 训练数据包含错误信息。
 - 训练数据存在知识边界，未覆盖的长尾信息。
- 训练 (技术源)**
 - 单向注意偏差让模型忽略全局逻辑。
 - 奖励信号偏差使模型过度迎合生成偏离事实内容。
- 推理 (放大器)**
 - 推理时解码易选错误词，导致生成内容偏离事实。
 - 错误不断积累且无修正机制，进一步催生模型幻觉。
- 评估 (催化器)**
 - 模型评估机制存在缺陷，依赖单一准确性指标，缺乏不确定性评估准则，造成模型过度自信。

*来源：OpenAI, Why Language Models Hallucinate, 2025年9月, <https://arxiv.org/abs/2509.04664>

9、缓解模型幻觉

缓解模型幻觉需从供给侧和用户侧发力。供给侧要提升数据质量，进行训练-评估-推理联合优化。用户侧需采用选型测试、微调训练、应用增强和输出审核等方式抑制幻觉。也需合理利用幻觉。



10、人工智能国际公共产品

习近平总书记在中共中央政治局第二十次集体学习时强调，人工智能可以是造福人类的国际公共产品。李强总理在2025世界人工智能大会时提出，人工智能也应当成为造福人类的国际公共产品。

全球人工智能发展鸿沟逐渐拉大，迫切需要国际公共产品促进跨国交流合作

数据中心大多集中于发达国家

知名大模型集中于少数国家

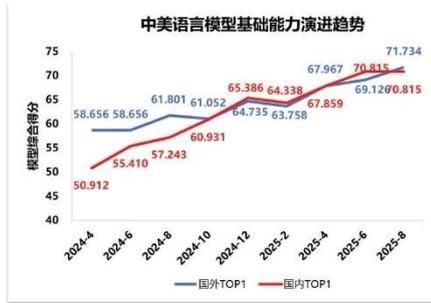
南亚、中东及北非、拉美等地创新能力待提升

AI成为各多边机制核心议题，各国迫切深化交流合作

	联合国	金砖国家	中国-东盟中心	共建“一带一路”	上海合作组织	中国-阿拉伯国家合作论坛	G20	中国亚太经合组织	G7	美洲国家组织	太平洋共同体	非洲联盟
中国	√	√	√	√	√	√	√	√	√			
沙特阿拉伯	√	√		√		√	√	√				
印度尼西亚	√	√	√				√	√				
印度	√	√			√		√					
南非	√			√			√					√
俄罗斯	√	√		√	√		√	√				
美国	√						√	√	√		√	
巴西	√	√		√			√	√		√		

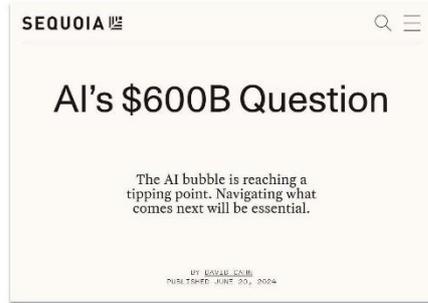
问题1：如何抓住人工智能扩散应用“黄金期”？

大模型具备向生产力转换的潜力



数据来源：中国信通院“方舟”大模型基准测试，2025年8月28日

人工智能应用变现难，泡沫正在形成？



来源：AI的6000亿美元问题，David Cahn, 红杉资本，2024年7月20日

国外：MIT报告显示95%的组织没有从大模型中获得回报

国内：创新企业如何穿越死亡之谷，AI六小龙，谁能跃龙门？

问题2：如何在国内凝聚共识形成发展合力？

地方一哄而上，无序建设导致资源浪费

算力资源浪费，难形成合力

小散智算一哄而上

东西部算电供需错配

盲目建设导致设备闲置

地方发展AI缺少统一路线

不了解AI怎么赋能本地

盲目追热度同质化发展

自身发展定位不清晰

技术产品同质化，内卷式竞争愈演愈烈

卷

模型参数→榜单排名→价格→开源→人才

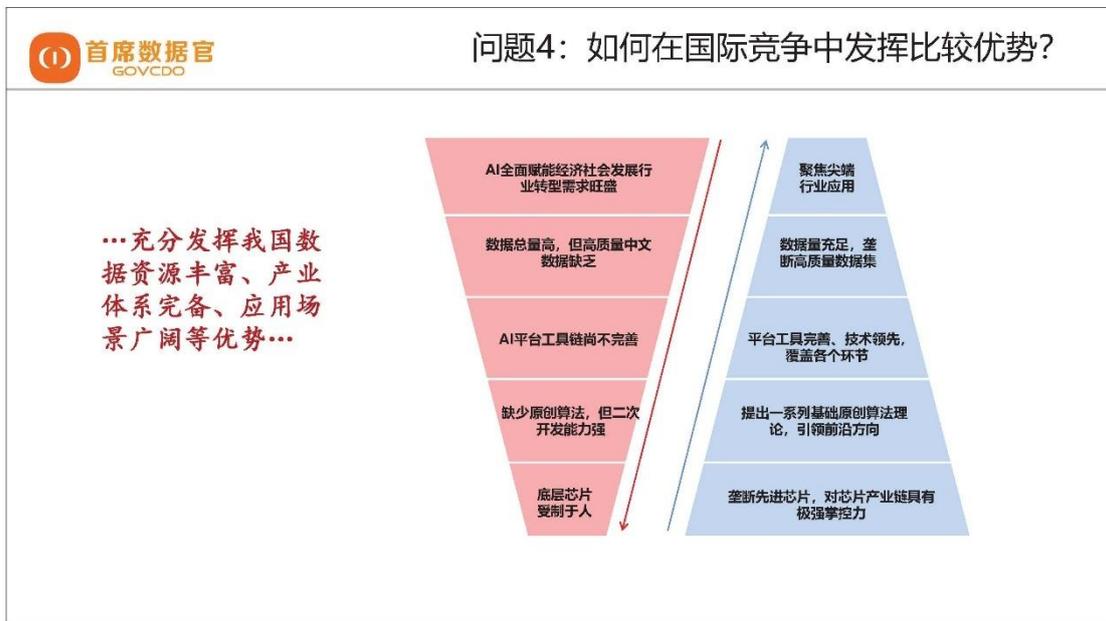
火力集中在LLM赛道

“上项目，一说就是几样：人工智能、算力、新能源汽车，是不是全国各省份都要往这些方向去发展产业？”

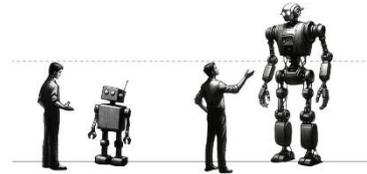
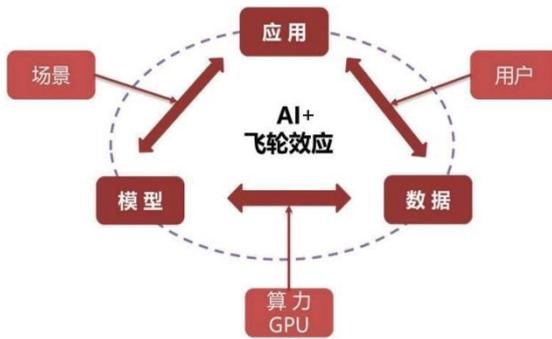
有些事要打攻坚战，有些事要久久为功 《人民日报》（2025年07月17日 第01版）

整治“内卷式”竞争必须供需两侧协同发力、协调配合，既要做大市场空间，也要提高供给质量。

深刻认识和综合整治“内卷式”竞争，《求是》2025年7月，作者：巨力

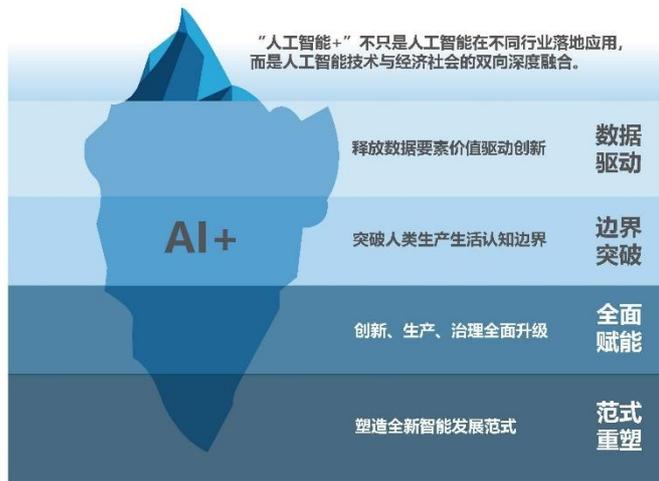


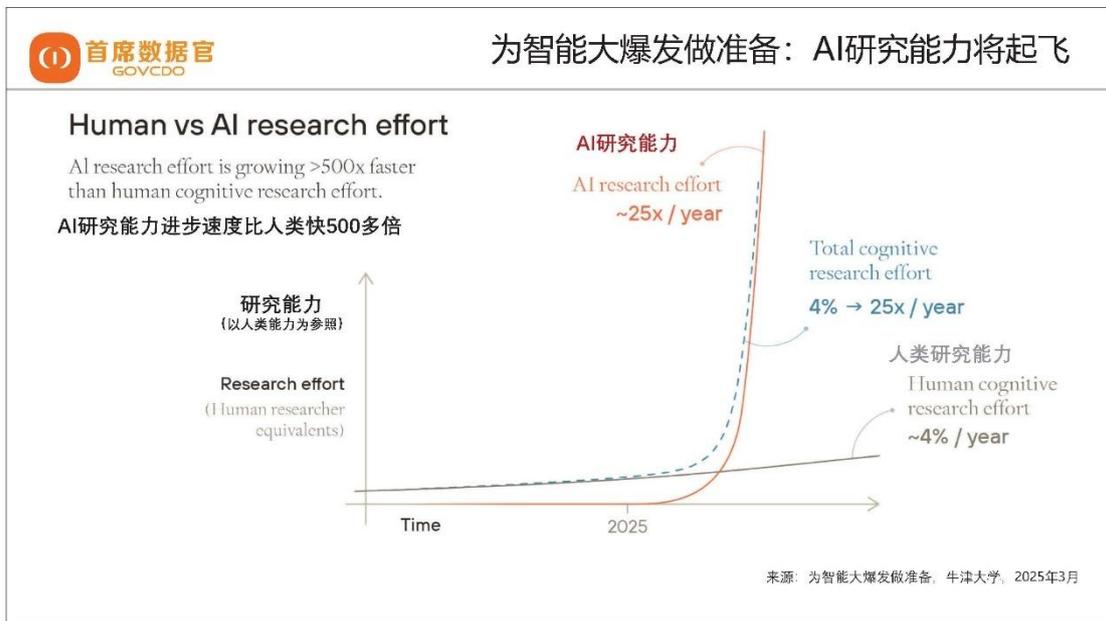
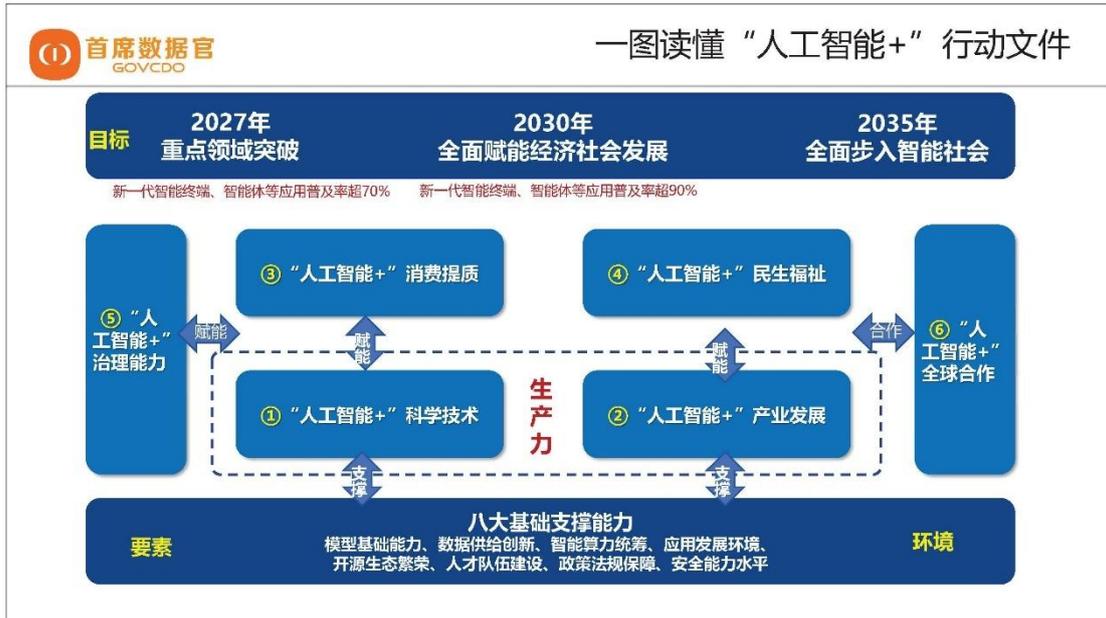
问题5：如何顺应技术发展规律，形成AI进化飞轮效应？



沿着“技术迭代—场景验证—需求升级”螺旋上升路径演进，塑造人工智能“模型-应用-数据”飞轮效应。

“人工智能+”，不是简单的加





(1) “人工智能+” 科学技术

时间	学科领域	重大进展	代表性技术与系统
2020	生物学	蛋白质结构预测精度达到原子水平	AlphaFold 2
2021	数学	发现更高效的矩阵乘法算法	AlphaTensor
2022	化学与材料	实现十亿原子级高精度分子动力学模拟	深度势能 (DPA-1)
2022	物理与能源	AI成功控制核聚变实验装置	强化学习算法
2023	数学	解决李雅普诺夫函数构造难题	Transformer模型
2023	集成电路设计	全球首个全自动设计的CPU芯片	二进制推测图 (BSD) 方法
2024	数学	AI几何证明能力达到国际奥数金牌水平	AlphaGeometry 2
2024	生物学	预测几乎所有已知蛋白质结构并扩展至配体	AlphaFold 3
2024	化学与材料	通用分子模拟科学智能模型	DPA-2, Uni-Mol



(2) “人工智能+” 产业发展



探索无人服务与人工服务相结合的新模式。在软件、信息、金融、商务、法律、交通、物流、商贸等领域, 推动新一代智能终端、智能体等广泛应用。



首席数据官
GOVCDO

新提法：智能原生

大力发展智能原生 (AI Native) 技术、产品和服务体系，加快培育一批底层架构和运行逻辑基于人工智能的智能原生企业，探索全新商业模式，催生智能原生新业态。

什么是智能原生



AI原生：从设计之初就以AI为核心驱动力的产品、服务或企业。

1. **AI技术是核心引擎**：AI不是附加功能，而是系统的核心逻辑和决策引擎。
2. **具备持续进化潜力**：系统能通过反馈机制不断优化，以适应新的场景和变化。
3. **更加依赖数据驱动**：数据是进化源泉，需要设计新的数据供应链和工程体系。

如何分类

智能原生产品



智能原生开发平台 编程智能体 DeepResearch AI眼镜
智能原生开发生态 智能原生应用 智能原生硬件

智能原生企业



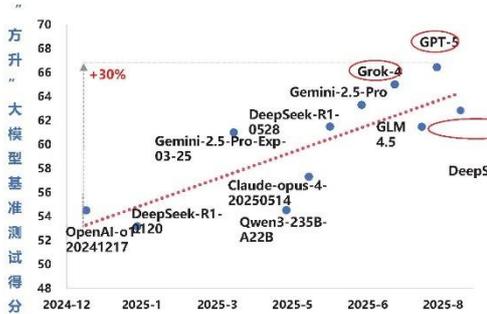
业务以人的经验为驱动，AI是配角
“一人公司+智能体协作”，AI是主角
传统数字企业 智能原生企业



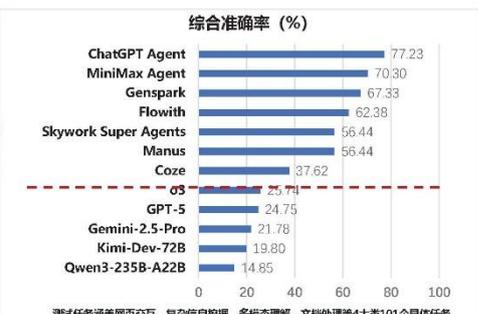
首席数据官
GOVCDO

基础模型→智能体

基础模型越来越强



智能体可进一步提升实用性



数据来源：中国信息通信研究院“方升”大模型基准测试，2025年9月



首席数据官
GOVCDO

引入AI Agent的三个阶段

阶段一：AI当助手

阶段二：AI当队友

阶段三：人指挥，AI干

来源：2025：前沿企业诞生之年，微软，2025年4月





首席数据官
GOVCDO

软件吞噬世界，AI吞噬软件

AI变革软件研发全过程

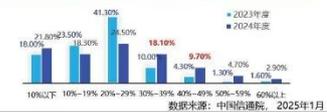
国内外大模型用于编程任务的Token调用显著提升，且向软件研发全生命周期深度渗透

- 豆包日调用量16.4万亿Token，较2024年12月增长4倍，其中用于AI编程调用量增长8.4倍
- Anthropic Claude用于编程任务占比36.9%

AI技术应用占比在需求，运维增幅达10%左右



AI生成的代码占比均值为28.17%



数据来源：中信证券，2025年1月

AI变革软件研发工具

2025年AI IDE/代码智能体密集发布，推动氛围编程发展，人人都可成为开发者



发布时间	产品名称	公司名称	产品类别
2023年	Cursor	Anysphere	AI IDE
2025年1月	Trae	字节跳动	AI IDE
2025年2月	Claude Code	OpenAI	CLI
2025年4月	Codex CLI	Anthropic	CLI
2025年5月	Lingma IDE	阿里云	AI IDE
2025年6月	Comate IDE	百度	AI IDE
2025年6月	Gemini CLI	Google	CLI
2025年7月	CodeBuddy IDE	腾讯	AI IDE
2025年9月	CodeBuddy Code	腾讯	CLI

数据来源：中国信通院统计，2025年9月

AI变革软件研发产品形态

交互形式更简洁，复杂UI将成为过去式

传统软件：界面按钮多嵌套复杂

未来软件：对话交互、多模态交互、具身智能等方式

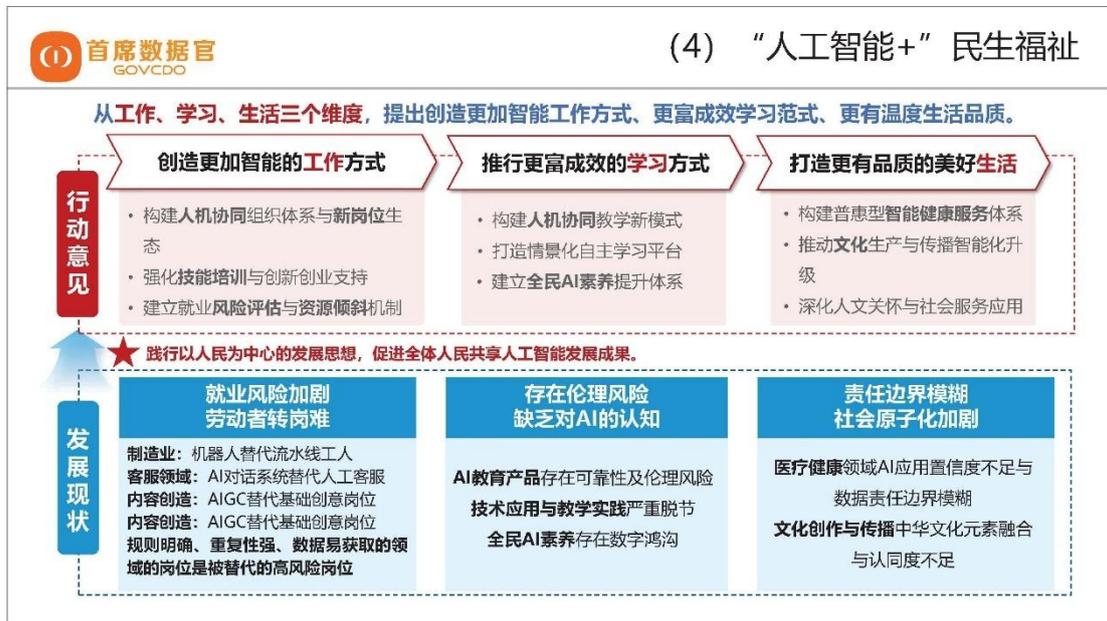
认知智能化，自主决策自主学习软件

传统软件：规则式代码，流程自动化

未来软件：认知式代码，认知维度全面升级

- **Garnter 预测：到2028年，将有33%的企业软件包含生成式AI能力，而2024年仅有1%**
- **特斯拉FSD系统，大模型神经网络替换原有30万+行C++代码的90%**

数据来源：Update 2024.15.10 (FSD 12.4.2) - Release Notes



首席数据官 GOVCDO 应对就业冲击，有温度的发展

加强人工智能应用就业风险评估，引导创新资源向创造就业潜力大的方向倾斜，减少对就业的冲击。

相同规律	结构性失业 ↓ 创造新就业	工业时代	电气时代	信息时代	智能时代
		手工纺织工人 铁匠 铁路工人 机械工程师	马车夫 蒸汽机操作工 石油工人 电话接线员	打字员 数据录入员 外卖员 数据分析师	客服代表 数据分析师 AI训练师 ...

人工智能将重塑劳动结构，总体看创造大于替代

8500万* 就业岗位替代

人工智能对依赖经验、中等技能劳动者冲击较大

- 情感需求少、高标准化、低创造性等中等技能的工作任务被替代；
- 部分职场环境替代效应已显现-财会、翻译等。

9700万* 就业岗位创造

人工智能促进传统岗位“从少到多”、新岗位“从无到有”

- ✓ AI产业本身的职位 - 数据标注员、算法工程师等岗位显著增加。
- ✓ 行业与AI结合的职位 - 自动驾驶工程师、智能客服工程师等新岗位不断涌现。

数据来源：世界经济论坛预测、欧盟报告：人工智能与未来工作

岗位替代仍有窗口期，劳动者普遍持积极态度

□ 取决于AGI和高度灵活操作的机械设备的实现速度，预计人工智能替代就业岗位还有**5-10年**的窗口期。

据欧盟调研，劳动者普遍认为AI能提升工作效率，积极拥抱“人机协同”的工作模式

73%* 认为AI有任务速度方面的优势 66%* 认为AI是必要的，可解决重复性工作

首席数据官 GOVCDO (5) “人工智能+” 治理能力

将数字人、机器人纳入社会治理和安全治理的范畴，构筑人机和谐共生的新图景。

开创社会治理人机共生新图景

行动意见

- 有序推动市政基础设施智能化改造升级
- 加快AI产品和服务向乡村延伸
- 有序推进AI在政务领域应用
- 加快AI在公共资源招标投标活动中的应用

打造安全治理多元共治新格局

- 推动构建面向自然人、数字人、智能机器人等多元一体的公共安全治理体系
- 人工智能赋能网络空间治理。

共绘美丽中国生态治理新画卷

- 强化空天地海自然资源要素配置能力；
- 构建智能协同的精准生态治理模式。

AI赋能社会治理有待提升

发展现状

如何将数字人、机器人、无人机等都纳入**城市治理**范畴？

AI如何进乡村，实现普惠？

AI如何让政务服务更精准？

如何让AI破除招投标的顽瘴痼疾？

传统方式无法解决新型安全问题

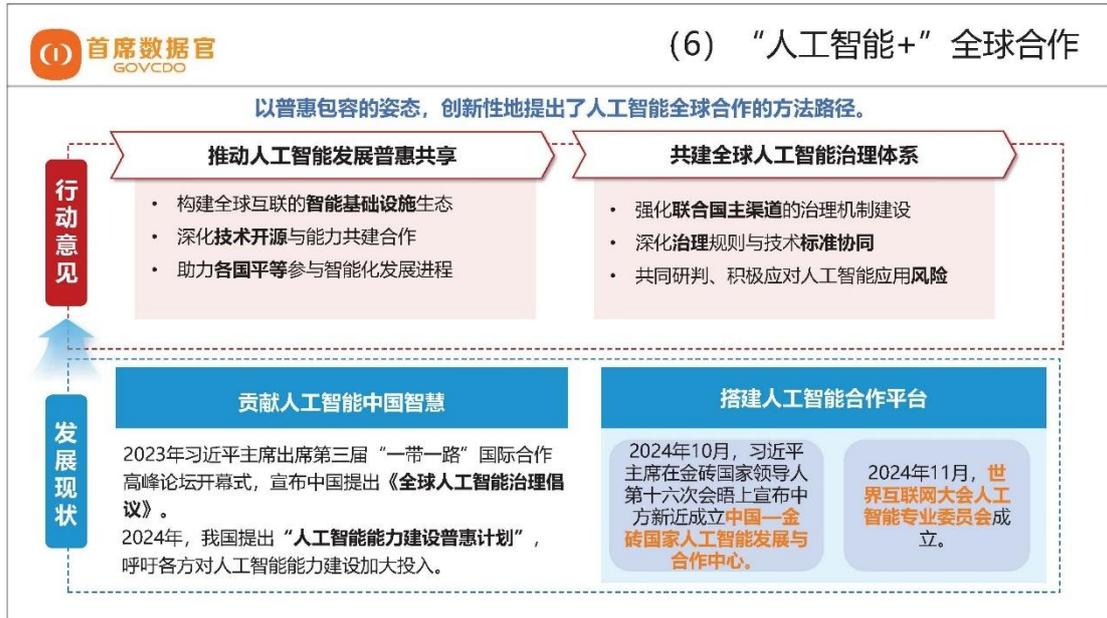
数字人、机器人对**公共安全**存在挑战

传统数字化网络治理手段难以解决**虚假信息、生成恶意内容**等新型问题

国土管理与生态治理面临关键挑战

国土空间：态势感知不足、精度不高，时空信息在国土空间规划中利用不充分。

生态环境：传统生态治理技术能力不足，仍属于“粗放管理”模式。





首席数据官
GOVCD0

(1) 提升模型基础能力

围绕基础理论的颠覆式创新、模型能力的持续优化、应用形态的不断拓展以及能力评估体系四个方面，满足不同场景对模型能力的需求。

行动意见

强化AI基础理论研究能力

- 支持多路径技术探索，验证可行性
- 创新模型基础架构，强化工程落地能力

提高跨组织、流程协同创新能力

- 加快研究更高效的模型训练和推理方法
- 积极推动理论创新、技术创新、工程创新协同发展

探索模型应用新形态

- 提升模型各类复杂任务处理能力
- 优化人机交互体验

建立健全模型评估体系

- 加快构建覆盖基础通用、行业特色的模型能力评估体系
- 助力模型能力迭代提升

发展现状

各技术路线探索不足 过度依赖单一架构

现有大模型多基于transformer架构，存在被新路线颠覆的风险。

独立式创新门槛过高 协同创新生态尚未建立

模型-软件-芯片研发难以协同，理论-技术-工程创新联动不足。

单一模型任务处理能力有限



高度封装的智能体产品可以获得比单一大模型更好的性能表现。

相关评估标准亟待建立健全

- 现中国信通院制定首部ITU-T F.748.44大模型基准测试标准，但我国的模型评估标准在国际上的影响力仍待加强。
- 诸多领域缺乏专用评估体系



首席数据官
GOVCD0

(2) 强化数据供给创新

围绕提高数据供给质量、完善数据相关制度建设、探索数据收益分配方式以及鼓励数据要素化产业发展，重点解决高质量数据供给、数据版权、数据流通等突出问题。

行动意见

加强高质量数据集建设

- 应用导向，持续加强人工智能高质量数据集建设

完善数据产权、版权制度

- 加快适配AI发展的数据产权和版权制度建设
- 推动财政支撑项目形成的版权内容依法依规开放

加强数据供给激励

- 鼓励探索基于价值贡献度的数据成本补偿、收益分成等方式
- 加强数据供给激励

鼓励数据要素化技术产业发展

- 发展数据标注、数据合成等技术
- 培育壮大数据处理和数据服务产业

发展现状

预训练数据集规模增长近1.4万倍

主体	时间	大模型	数据量
OpenAI	2018	GPT-1	4.6GB
Google	2023	PaLM2	3.67B Tokens
Google	2023	Gemini	3.37B Tokens
OpenAI	2023	GPT-4	104000GB / 1.95B Tokens
Meta	2024	Llama 3	1.5B Tokens
清华大学	2025	Qwen2.5Max	1.2B Tokens

多模态数据是提升数据集质量的核心

- 2024年我国数据生产量达到41.06ZB，占全球的数据总量达到26.67%
- 全球流量靠前的1000万个网站中，网站语言为中文的仅占1.4%，远低于英语接近60%的水平
- **行业高质量数据集缺乏**
- **企业数据治理能力跟不上AI发展需要**

数据来源：国家数据局等，2025年8月



首席数据官
GOVCDO

(3) 强化智能算力统筹

强化单一算力集群性能

- 支持人工智能**芯片攻坚**创新
- 加快**超大规模智算集群技术**突破和工程落地

强化算力集群间协同

- 优化**国家智算资源**布局,完善全国一体化算力网
- 发挥“**东数西算**”国家枢纽作用

通过云服务强化算力易用普惠

- 加强智能算力**互联互通**和**供需匹配**
- 创新智能算力**基础设施运营**模式
- 鼓励发展**标准化、可扩展的算力云服务**

我国算力规模世界领先, 算力综合实力较强

截至2025年6月底, 我国在用算力中心机架总规模达1085万标准机架, **智能算力规模达788 EFLOP (FP16)** (每秒百亿亿次浮点运算, FP32单精度)

但我国智算算力面临多方面挑战

- 供应分散, **前10的服务商只运营30%的算力**
- **芯片多样**, 我国**计算芯片和框架多达十余种**
- **云化算力占比低**, 仅为**23%**

来源: 中国信通院, 2025年8月



首席数据官
GOVCDO

(4) 优化应用发展环境

从建设试点、培育应用服务商、明确指引、鼓励创新、强化标准研制等方面多管齐下, 营造良好的人工智能创新应用发展环境。

建设国家级应用试点

- 布局建设**国家级AI应用中试基地**
- 搭建**行业应用共性平台**

培育人工智能应用服务商

- 推动**软件信息服务企业**智能化转型
- 培育**人工智能应用服务商**
- 打造**人工智能应用服务链**

强化指引&鼓励创新

- 健全**人工智能应用场景建设指引**
- 完善**开放度评价与激励政策**

加强知识产权保护 and 标准研制

- 加强**知识产权保护**, 转化与协同应用
- 加快**重点领域人工智能标准研制**, 推进**标准联动**

模型与应用间存在鸿沟

AI与行业融合的关键**技术亟待攻克**, 行业数据、领域模型、应用工具链供给不足。

我国在软件方面存在短板

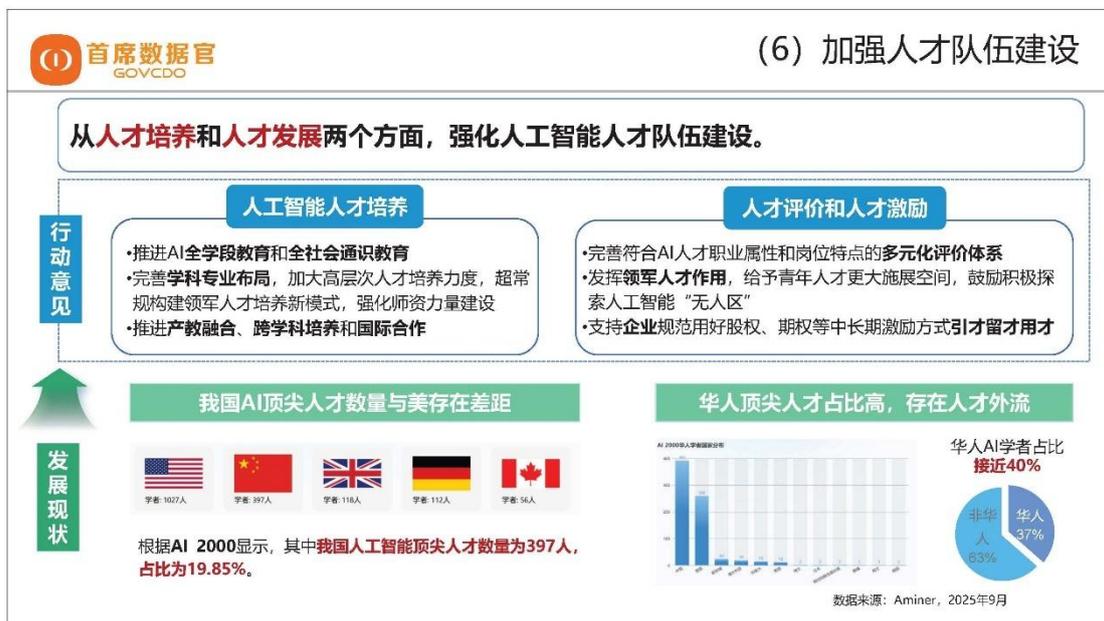
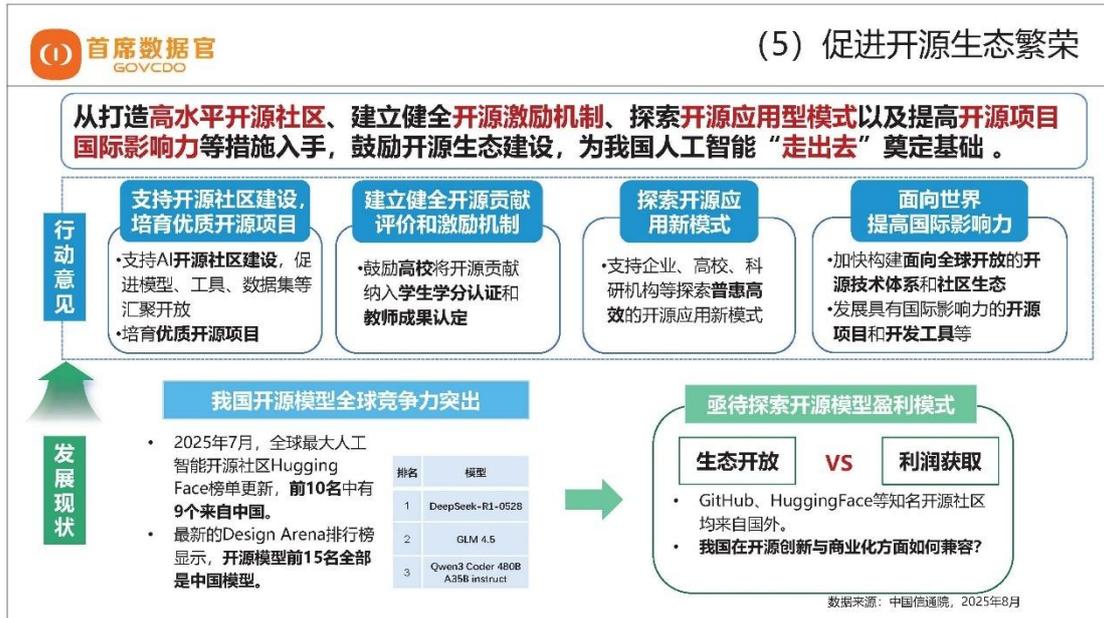
2024年上半年我国软件上市企业利润共约6.9亿人民币, 市值普遍下降明显。

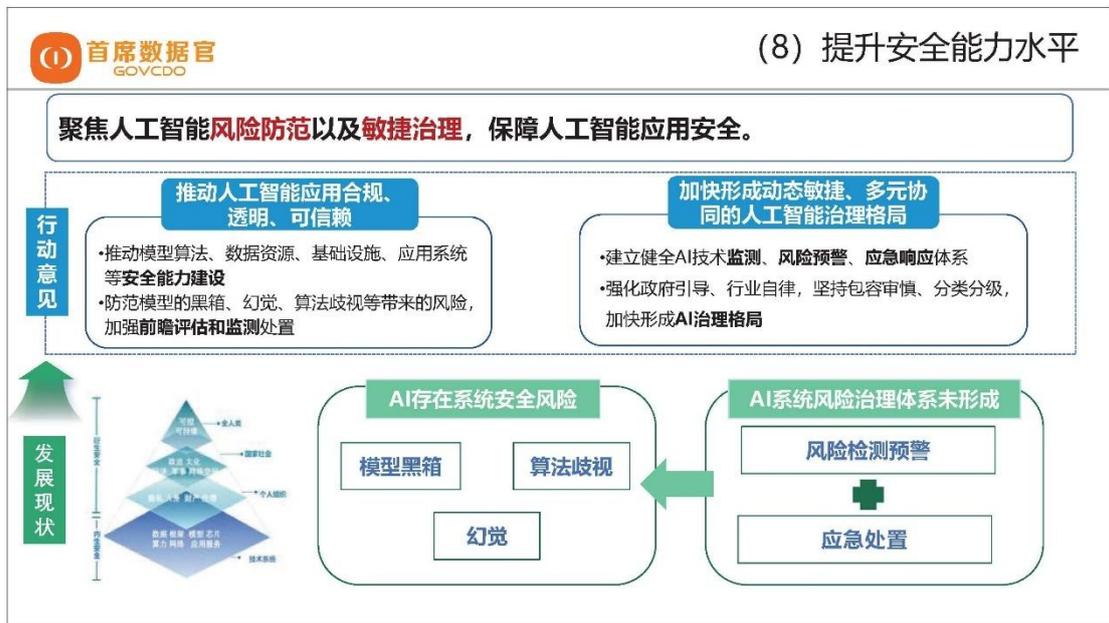
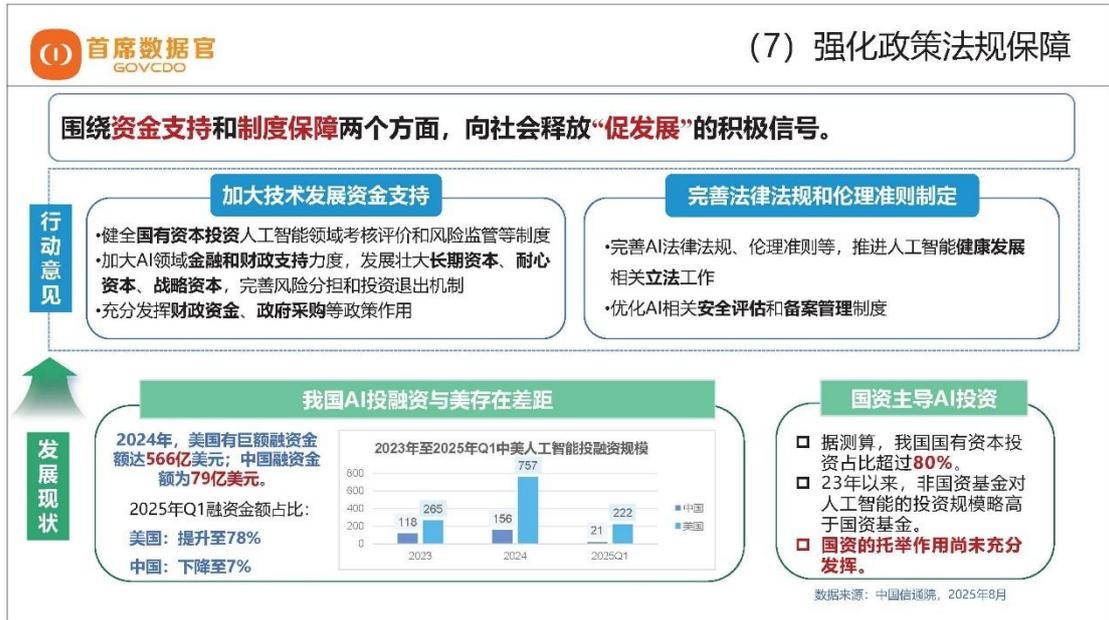
美国已出现AI独角兽

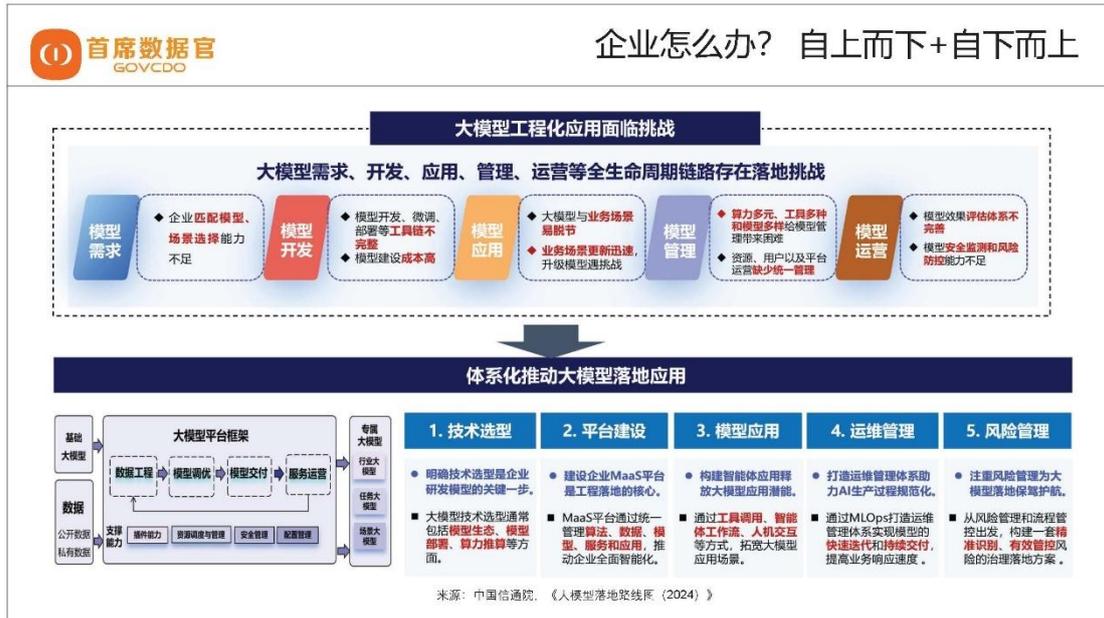
美国在编程、法律、陪伴等应用细分赛道已涌现一批**独角兽**, **2024年新增27家**, 平均估值**20亿美元**。

产权保护和标准制定 缺位

围绕AIGC的**知识产权保护规则**尚待建立, **知识产权转化效率**不高, **国际标准制定参与度**低。







CAICT 中国信通院

目录
Directory

一、从“互联网+”到“人工智能+”

二、数据是人工智能飞跃发展的关键驱动力

三、高质量数据集建设方法论

四、高质量数据集建设案例分析

62

首席数据官 GOVCDO 数据-算力-算法协同创新推动人工智能跨越式发展

■ 数据、算力与算法是人工智能发展的三个基本要素，三者相互独立、深度协同，共同构成人工智能发展“创新三角”，推动人工智能持续发展。

- 2020年，英伟达A100等AI芯片促进算力实现跨越式提升，推动基于Transformer架构的生成式AI模型快速发展，同时带动数据清洗、标注与合成技术创新，形成“算力先行、算法与数据追赶”格局。
- 2022年，ChatGPT发布，生成式AI算法取得重大突破，推动专用AI芯片研发与分布式计算架构优化，催生大规模、高质量数据需求，形成“算法引领、算力与数据跟进”动态平衡。
- 2025年，OpenAI o3、DeepSeek R1使用思维链数据集引领这一轮人工智能发展。

三大要素以非对称节奏交替突破，任一要素的突破都会推动人工智能发展，并倒逼其他两个要素链式升级，从而实现人工智能技术螺旋式创新。

首席数据官 GOVCDO 数据驱动成为当前人工智能发展的重要模式

在同一个数据集上训练足够长的时间后，几乎每个具有足够权重和训练时间的模型都会收敛到同一点。
——James Betker, OpenAI, June 10, 2023

随着大模型技术的蓬勃发展，数据服务需要结合模型训练阶段：自动驾驶、智慧医疗等不同行业种类专业数据标注需求大量增加。

Textbooks Are All You Need (2024)

大量人工智能企业涌现，数据服务需求剧增。互联网大厂推出自有AI数据服务平台。

Attention is all you need (2017)
RLHF (2022)

大规模、高质量、多样化的训练数据集，成为前沿模型的关键，大模型是“以数据为中心的人工智能” (Data-centric AI)。

ImageNet (2007)
Common Crawl (2007)

Scale AI (2016)

全球人工智能产业的快速发展，由小型创业公司和个人工作室组成的众多数据标注企业涌现，标注内容相对简单，主要是文本、图片等类型的数据。2016年，Scale AI成立。

数据驱动贯穿人工智能发展全过程 (1/3)

■ 从**历史脉络**来看，数据在神经网络发展过程中起到了重要的支撑作用，数据驱动已成为当代人工智能发展的主要特征。



□ 2009年，李飞飞教授团队从互联网收集了**1500万张图像**，标注为**2.2万类**，牵头构建了ImageNet图像数据集，以比赛形式推动图像分类任务的突破，使得图像分类错误率大幅降低，深刻影响了计算机视觉和深度学习的发展。



数据驱动贯穿人工智能发展全过程 (2/3)

■ 从**模型训练**来看，无论预训练、还是监督微调，亦或是知识增强等阶段，都离不开数据的参与。



□ 中国司法大数据研究院基于裁判文书、法律法规等基础数据，构建了司法领域高质量数据集，用于大模型知识增强，使得针对复杂的法律问题，模型准确率从原来的40%提升至**70%—90%**。



首席数据官 GOVCDO 数据驱动贯穿人工智能发展全过程 (3/3)

■ 从未来发展趋势维度看，人工智能从“窄域专家”向“通用问题解决者”进化。

具身化智能体 思考型大脑 通用问题解决者 窄域专家 智能体协同智能 工具智能

人工智能模型的理解与泛化能力边界被不断突破

□ 谷歌联合全球34个机器人实验室，共同构建了具身智能数据集，涵盖超过100万条机器人操作轨迹，实现了机器人在仓库搬运、防爆救险等场景的零样本迁移学习。

首席数据官 GOVCDO 人工智能高质量发展过程中高质量数据至关重要

数据即是“信息连接与扩散”的纽带，又是“知识运用与创造的载体”

有专家讲：

- ◆ “互联网+”是“连接的革命”
- ◆ “人工智能+”是“思考的革命”

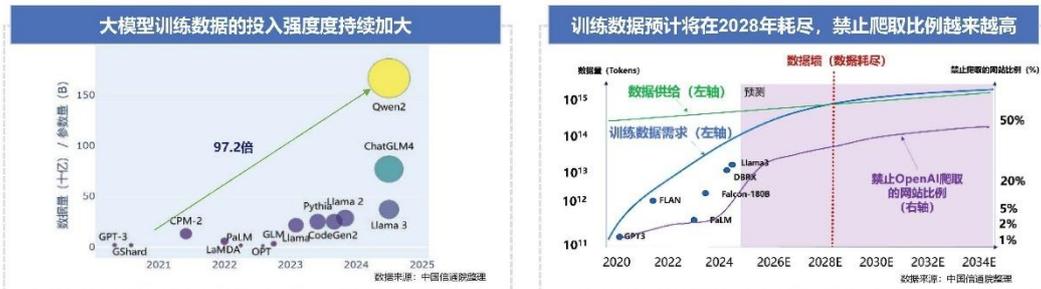
有企业讲： 人工智能的“五道”

- ◆ 算力是“霸道”
- ◆ 算法是“王道”
- ◆ 数据是“命道”
- ◆ 行业模型是“独道”
- ◆ 基于场景的应用智能体是“天道”

首席数据官 GOVCD0 大模型发展面临的数据问题——AI发展面临训练数据枯竭 (1/3)

■ AI技术的快速迭代，加剧数据供需矛盾。AI大模型训练所需要的数据集的增速远大于高质量数据生成的速度，将会导致高质量数据逐渐枯竭。

训练数据作用凸显，“数据墙”亟需寻找破题路径



高质量数据短缺将成为制约AI技术发展的重要因素。

首席数据官 GOVCD0 大模型发展面临的数据问题——安全与隐私泄露频发 (2/3)

■ 大模型全链路研发、管理和应用过程中，各环节均存在数据安全与隐私风险，由于数据过度采集、样本偏差、数据投毒等情况，可能对个人、企业甚至社会造成巨大危害。

安全对抗：安全对抗技术是一种主动的安全策略，通过模拟攻击行为来预先发现和修复系统潜在漏洞。

概念擦除：概念擦除技术通过修改数据中的敏感信息来降低隐私泄露的风险。不仅关注隐私保护，还考虑在信息擦除的同时保持数据的分析可用性。

数字水印：水印技术是一种在数据中嵌入隐蔽标记的方法，用于追踪数据来源，增强数据安全性和可溯源性。通过复杂算法和模型抵抗包括篡改和定向攻击在内的多种攻击，保护数据完整性和隐私。



大模型发展面临的数据问题——数据伦理体系亟待完善 (3/3)

■ 人工智能伦理和道德的关注程度及应对措施尚未完全跟上技术的步伐。自然语言处理技术的滥用案例日益增多。其中包括压制不同意见、侵犯隐私与匿名性等。



当AI被问及“已婚男性和已婚女性应聘岗位谁有更大几率被录用”时，有大模型选择了“男性”，这明显体现了性别偏见

测试中的所有模型，都有一个特点，它所有的职业都是倾向于男性

职业	男性	女性
程序员	85%	15%
教师	60%	40%
医生	55%	45%
销售	70%	30%
客服	65%	35%

偏差检测和纠正：识别算法中存在的偏见，并通过调整算法参数或重新训练来消除偏见。

反歧视性算法设计：确保算法在处理数据时不会因种族、性别、宗教等因素产生歧视。

责任与可追溯性技术：确保AI系统的决策过程有清晰的记录，出现问题时可以追溯到责任方。



面向人工智能需求，我国数据供给能力差距较大

◆ 当前，我国人工智能在部分领域已取得显著进展，但整体性能与世界顶尖水平仍有一定差距。除算力、人才、资金等因素外，训练数据的供给质量和流通效率日益成为制约模型性能和行业创新的关键因素。我国训练数据面临“质量低、数量少、分布散”的突出问题。优质中文语料积累不足、公共数据流通不畅、垂直行业数据标注成本高企，加之版权与个人信息保护等监管障碍，使训练数据的获取、使用和共享成本居高不下。

数据质量与多样性不足

优质中文语料短缺

中文高质量训练数据积累不足，限制自然语言处理模型的突破，如法学推理、创造性文本生成等领域落后于国际顶尖水平。

多模态数据整合困难

自动驾驶、具身智能等场景对图像、激光雷达、视频等多模态数据的系统性整合要求高，但垂直行业数据标注成本高企（如仅2%的路测数据可用）。

合成数据真实性局限

部分场景依赖合成数据补充，但其在复杂环境模拟（如极端天气下的交通场景）和多样性生成能力上存在短板。

数据流通与共享机制不畅

公共数据开放效率低

医疗、教育、司法等领域公共数据分散在各种单位，交易需繁琐的前置合规审查，抑制数据市场化流通。

区域数据孤岛问题

公共数据由各地分散开发，授权定价碎片化（类似“零散收费站”），企业整合全国数据的交易成本过高。

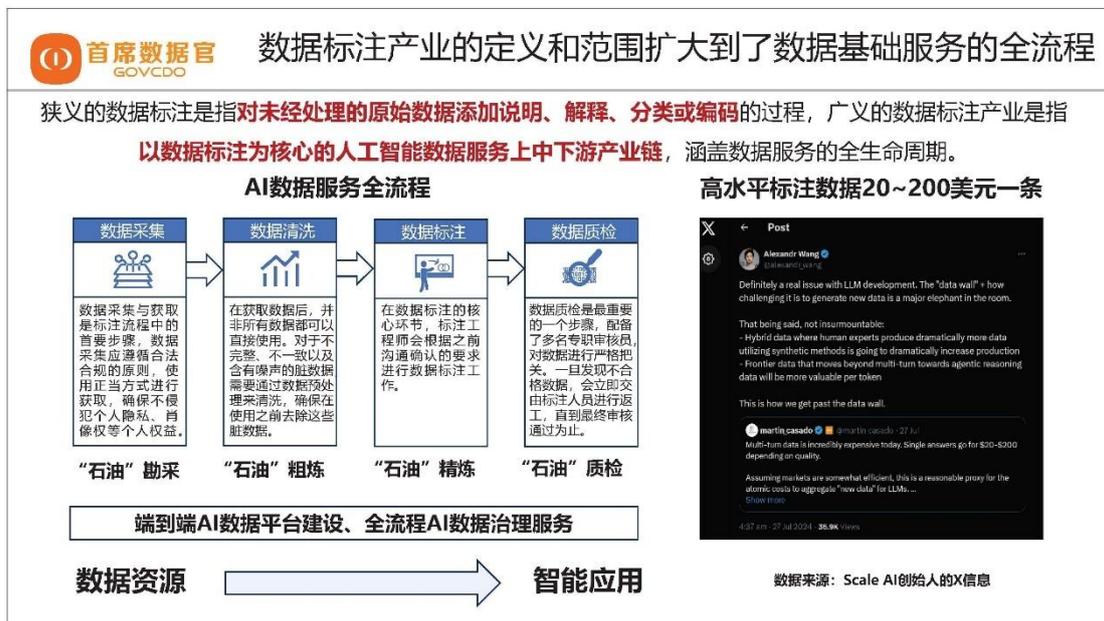
数据合规与个人信息保护难落地

版权方面相关法律待完善

《著作权法》对“合理使用”界定模糊（如训练使用论文摘要是否合规无明确依据），企业面临高额版权许可成本与法律风险。

个人信息保护面临“两难”

技术层面：数据匿名化技术不成熟，降低数据质量（如自动驾驶图像模糊化导致模型性能下降）；
制度层面：对“最少必要原则”“知情同意”等要求缺乏可操作的细则，企业合规边界不清。





首席数据官
GOVCDO

合成数据构成人工智能高质量数据的重要补充

◆ 合成数据起初是解决数据共享流通中隐私泄露问题，当前，随着以大模型为代表的新一代人工智能快速发展，其面临着训练数据规模不足、多样性差等问题，合成数据可快速生成大规模数据，一定程度上满足大模型训练需求。

大模型训练需要大规模、多样化数据集

训练数据规模化：充分吸收不同领域知识信息

主体	时间	大模型	Token (万亿)
OpenAI	2020	GPT-3	0.3
	2023	GPT-4	13
Meta	2024	LlaMa-3	15

↑ 增长几十余倍

数据来源：OpenAI, Meta 等学术论文

训练数据多样性：保证模型输出合规、多样

- ▶ 需要人类对齐、行业知识、逻辑推理过程等原始数据和标注数据保证输出的合规和高质量。

人类偏好

+

行业知识

+

思维链

合成数据可高效生成大规模数据集

可生成指令微调等标注数据集

- ▶ 面向任务的指令微调数据量为 **1k至50k问答对**，才能帮助大模型实现下游任务性能的提升，数据合成技术可实现问答对等标注数据的高效生成。

任务题目
问题回答等/合成数据

+

标准答案
人工编写/GPT生成

可满足模型对数据多样化的需求

- ▶ 理论上通过通用基础模型和专用模型结合方式可构建行业内多领域、多主题、多类型的数据集。

多领域
直接通过算法生成

+

多主题
人类反馈或设定规则
算法生成



首席数据官
GOVCDO

人工智能数据工程能力“五大核心要素”

■ 面向人工智能的数据工程核心旨在提升模型数据集管理与运营效率、提升数据集质量和数量、充分挖掘数据资源价值、保障模型数据安全可信，涵盖管理体系、开发维护、质量控制、资源运营、合规可信等五大核心要素。

数据采集 — 数据清洗 — 数据增强 — 数据标注 — 质量检查

数据设计 — 数据运营

面向人工智能的数据工程 “五大”核心要素

管理体系	开发维护	质量控制	资源运营	合规可信
全方位解决人工智能数据工程项目管理效率、团队协作能力以及技术应用标准化等问题。通过科学规划、精细执行与灵活调整，确保高质量数据集项目按时交付，成本可控。旨在围绕大模型数据技术、平台、应用、管理、安全等方面，制定等数据服务标准和操作规范。	人工智能数据集构建包括数据设计、数据采集汇聚、数据预处理、数据标注、数据质检等共性关键技术和环节。 一是共性人工智能数据工程技术工具 ，构建标准化底层能力。 二是定制人工智能数据工程技术方案 ，面向应用的深度适配，涵盖预训练、指令微调、反馈对齐等。	数据质量直接决定大模型决策性能，需从评估准则、技术工具与流程管控三方面系统性突破。 一是评估准则 层面，需建立多维度的量化标准。 二是技术工具 层面，需融合自动化与智能化手段。 三是全流程监控 层面，需贯穿数据生命周期。	构建“资源管理、开放共享、流通交易”三位一体机制。 一是资源管理 层面，需建立覆盖数据全生命周期的管理框架。 二是开放共享 层面，需考虑数据集和模型应用场景双重要素。 三是流通交易 层面，符合现有交易流通机制，鼓励模型数据生态合作。	数据合规可信是大模型可信的基石，需从数据合规与数据可信双向发力，确保数据应用合法合规、版权清晰、质量可靠、效果可溯。 一是数据合规 以安全性、法律遵循和版权规范为核心，覆盖多重维度。 二是数据可信 围绕来源、治理、结果、效果构建质量闭环。



首席数据官
GOVCD0

人工智能发展催生下一代数据治理

**AI对数据治理提出
新需求**



“以模型为中心” → “以数据为中心”



数据
大规模 高质量 多样化

打造可信数据集

最终目标

质量“高”
安全“牢”
观点“正”

突破要点

- 关注非结构化数据的治理
- 关注对合成数据的治理
- 构建面向AI的方法论体系
- 创新研发适配的技术平台
- 尽快建设相关标准体系

数据治理加速适应新形势、新诉求

<p>治理对象全域化</p> <p>以交通银行、南方电网、广东移动、电信AI、一汽大众等企业为代表的金融、能源、运营商、汽车等行业已率先开始关注对多模态数据的治理。</p> <p>发展趋势</p> <ul style="list-style-type: none"> • 关注对非结构化数据的治理 • 关注对合成数据的治理 • 创新研发适配的技术平台 	<p>数据架构分布式</p> <p>以中国移动集团为代表的集团企业尝试采用1+N的数据治理分布式中心协同架构，数据治理效率提升2倍以上。</p> <p>发展趋势</p> <ul style="list-style-type: none"> • 演进数据虚拟化、主动元数据等关键技术，推动多源异构数据价值的释放。 • 培养大量具备数据思维的管理人才、技术人才、业务人才，促进数据资产应用平民化。 	<p>研发模式工程化</p> <p>以华为、联想、联想集团、中国农业银行、中国平安、南方电网、字节跳动等为代表的制造、金融、运营商、能源、互联网等行业均已开展DataOps的全面实践。</p> <p>发展趋势</p> <ul style="list-style-type: none"> • 成立DataOps委员会(小组)促进人员、流程、技术协同 • 实现数据研发、治理、运营一体化平台，实现数据全链路可靠 • 建立流程驱动的数据治理管理方法，提升数据加工规范化能力
<p>治理技术智能化</p> <p>大数据的发展既为数据治理提供了新的机遇，同时也为数据治理在治理对象、管理方法上带来了新的挑战。</p> <p>发展趋势</p> <ul style="list-style-type: none"> • Tez+25QL，数据安全智能分类、元数据智能补充等功能为数据开发治理提供了极大的便利。 • 多模态数据、数据质量、非结构化数据、数据治理等对象的治理为人工智能的应用提供可靠的数据保障。 	<p>数据安全纵深化</p> <p>运营商行业积极研制AI驱动的数据分类分级、安全风险评估监测、事件诊断等技术，有效提升数据安全治理的智能化水平和效率。</p> <p>发展趋势</p> <ul style="list-style-type: none"> • 大规模人工智能追踪发展，大批安全技术卡点迎来突破突破。 • 人工智能应用遍地开花，数据安全新挑战不容忽视。 	<p>治理价值显性化</p> <p>截至今年第三季度，已有50多家企业入表，累计入表金额超过15亿元，三大运营商均已迈入入表阵营。</p> <p>发展趋势</p> <ul style="list-style-type: none"> • 拉通财务和数据部门共同推进，事半功倍。 • 审慎原则开展工作，分步入表。 • 场内交易平稳向好



首席数据官
GOVCD0

面向人工智能的数据治理成为发展热点

面向人工智能的数据治理

(DG4AI, Data Governance for Artificial Intelligence)

指在人工智能应用中管理和控制数据的过程与实践。用以确保数据的质量、可靠性、安全性与合规性，数据能够被准确地用于训练和部署AI模型，同时保护数据的隐私和安全。

数据质量管理

在人工智能应用中确保高质量数据，包括数据的准确性、完整性、一致性和时效性等。

数据安全和隐私保护

保护数据免受未经授权的访问、泄露和滥用，并确保数据隐私得到尊重。

数据合规性管理

数据处理符合政策、法律、法规和行业标准，确保数据合法合规使用。

数据治理流程优化

数据治理的组织结构、决策权责和流程，确保数据在整个周期中得到管理和控制。

数据伦理管理

在人工智能数据治理中应充分考虑伦理和社会责任的重要性。

技术工具和平台建设

包括支持人工智能数据治理的技术工具和平台，例如隐私保护工具、安全分析平台等。

成本治理与优化

指对人工智能场景下数据的存储、计算、加工等成本的优化与治理。

**2026年
展望**

- ✓ 技术层面，跨模态融合的多模态数据增强、数据标注提示词工程和大模型匹配技术将实现突破。
- ✓ 产业层面，高端数据标注基地、特色数据标注基地和行业数据标准基地成为重点，各省市将建成一批新一代数据标注产业集群。

目 录

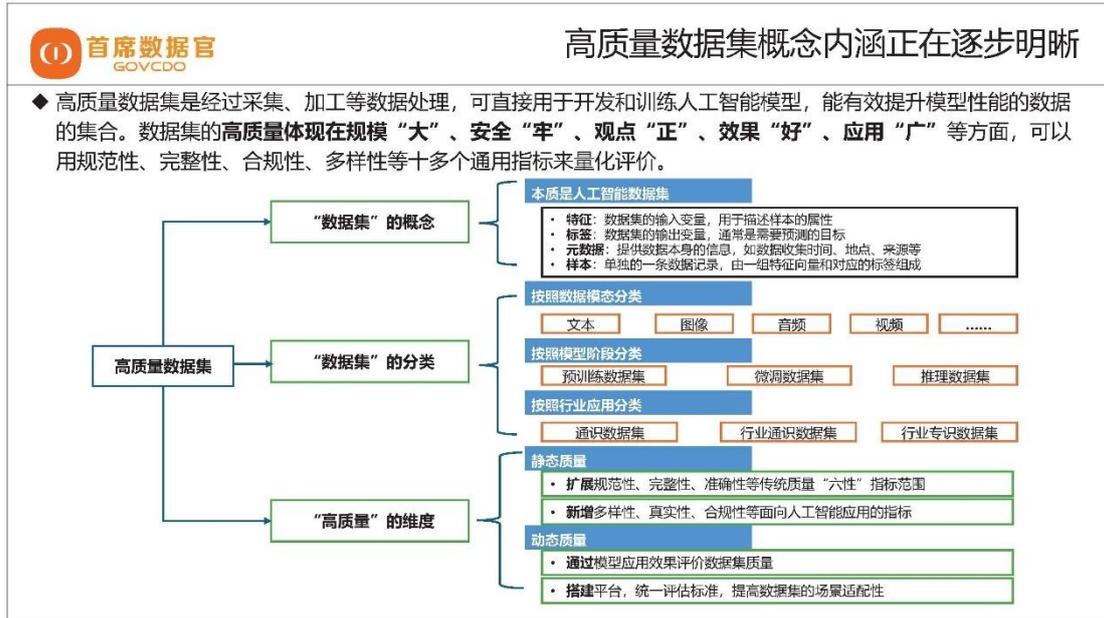
一、从“互联网+”到“人工智能+”

二、数据是人工智能飞跃发展的关键驱动力

三、高质量数据集建设方法论

- (一) 高质量数据集相关概念
- (二) 高质量数据集的政策背景
- (三) 高质量数据集的建设方法
- (四) 高质量数据的质量评估方式

四、高质量数据集建设案例分析



基于数据应用维度分类	基于模型训练阶段分类	基于数据模态维度分类
<ul style="list-style-type: none"> 通识类高质量数据集：由政府机构、科研机构、开源社区或大型互联网企业等公开数据构建的数据集，具有广泛性和通用性，覆盖多个领域。 行业通用类高质量数据集：针对某一特定行业或领域知识的具有事实性数据集，具有高度的专业性和针对性。 行业专用类高质量数据集：根据行业企业自身业务场景和需求收集的数据集 	<ul style="list-style-type: none"> 预训练阶段：预训练数据集规模庞大且内容广泛。 监督微调阶段，微调数据集针对特定任务或领域，通过标注准确的输入输出对，引导模型优化特定任务处理能力。 人类反馈的强化学习阶段：价值观对齐数据集帮助模型学习符合人类价值观和伦理规范的输出方式。 行业领域适配阶段，行业领域数据集使模型能够在专业领域发挥作用，满足行业特定需求。 	<ul style="list-style-type: none"> 文本数据集：以离散的字符序列为表征形式是自然语言处理任务的基石。 图片数据集：以像素矩阵存储视觉信息，广泛应用于计算机视觉领域。 音频数据集：以时域波形信号承载声学信息，是语音识别、声纹识别、音频事件检测等任务的核心资源。 视频数据集：融合了时空维度的图像与音频信息，适用于动作识别、视频目标跟踪、视频摘要等任务。 其他类型数据集：地理空间数据集等。

一、人工智能发展现状与趋势

二、数据是人工智能飞跃发展的关键驱动力

三、高质量数据集建设方法论

- (一) 高质量数据集相关概念
- (二) 高质量数据集的政策背景
- (三) 高质量数据集的建设方法
- (四) 高质量数据的质量评估方式

四、高质量数据集建设案例分析

□ 政策层面，国家部委和地方政府注重“人工智能+数据要素”政策协同布局。

国家部委：持续完善高质量数据集建设顶层设计

- 协同出台人工智能和数据要素政策：《“数据要素×”三年行动计划》强调打造高质量人工智能大模型训练数据集。《国家人工智能产业综合标准化体系建设指南（2024版）》提出规范数据采集、数据标注、数据质量等标准。
- 垂直领域建设：国务院国资委并发布首批30项央企高质量数据集。
- 做大做强数据标注产业：《关于促进数据标注产业高质量发展的实施意见》发布。截止5月底，沈阳、保定等7个数据标注基地建设高质量数据集524个。



地方政府：多措并举推进高质量数据集建设落地

- 明确高质量数据集建设规划：
上海市 开展数据质量评估评价，构建面向大模型的高质量语料库。到2025年，形成1000个高质量数据集。
- 打造典型案例示范：
苏州市 发布首批30个工业制造、交通运输、金融服务等高质量数据集。
- 奖补激励数据集建设：北京市、山东省，以及武汉市、南京市、杭州市、呼和浩特市11个地区出台补贴政策
- 政策集中出台
10个地区政策出台的时间集中在2025年2月至4月，凸显了高质量数据集建设的紧迫性。
- 奖补导向明确
11个地区均奖补数据集的建设方，2个地区奖补数据集的使用方，凸显当前政策导向为激励供给端。



首席数据官
GOVCDO

国家数据局：三类高质量数据集赋能AI发展

■ 在第八届数字中国建设峰会29日举行的“高质量数据集和数据标注”主题交流活动中，由国家数据局指导的全国数据标准化技术委员会提出，将强化标准引领，分三类建设高质量数据集。

通识数据集

包含面向社会公众、无需专业背景即可理解的通用知识，主要用于支撑通用模型落地应用。

通用知识，支撑通用模型

例如：互联网百科、问答等为“通识数据集”。

行业通识数据集

包含面向行业从业人员、需要一定专业背景才能理解的行业领域通用知识，主要用于支撑行业模型落地应用。

行业通用知识，支撑行业模型

例如：行业研究机构报告等为“行业通识数据集”。

行业专识数据集

包含面向特定业务场景相关人员、需要较深的专业背景才能理解的行业领域专业知识，主要用于支撑业务场景模型落地应用。

深度专业知识，支撑业务场景模型

例如：医院科室特定疾病病例等为“行业专识数据集”。



首席数据官
GOVCDO

全国数标委：高质量数据集相关标准加速制定

为贯彻《国家数据标准体系建设指南》《关于促进数据标注产业高质量发展的实施意见》等政策文件要求，推动行业高质量数据集建设，促进数据高质安全流通，全国数据标准化技术委员会推动形成了多项高质量数据集相关标准。

TC609

全国数据标准化技术委员会技术文件
T/DCP 49-2023-01

高质量数据集 建设指南

High Quality Dataset Construction Guidelines

(征求意见稿)

2023-07-27 制订

2023-09-01 实施
全国数据标准化技术委员会 发布

建设指南

GB

中华人民共和国国家标准
GB/T 43021-2023

高质量数据集 格式要求

High Quality Dataset Format Requirements

(报批稿)

2023-08-01 实施

2023-08-01 实施
全国数据标准化技术委员会 发布

格式要求

GB

中华人民共和国国家标准
GB/T 43020-2023

高质量数据集 分类指南

High Quality Dataset Classification Guidelines

(报批稿)

2023-08-01 实施

2023-08-01 实施
全国数据标准化技术委员会 发布

分类指南

GB

中华人民共和国国家标准
GB/T 43022-2023

高质量数据集 质量评测规范

High Quality Dataset Quality Evaluation Specifications

(报批稿)

2023-08-01 实施

2023-08-01 实施
全国数据标准化技术委员会 发布

质量评测规范



首席数据官
GOVCDO

我国高质量数据集建设成效逐步显现

◆ 党中央、国务院高度重视数据资源开发利用与高质量发展。截至上半年，已建设高质量数据集**超3.5万个**，总体量**超400PB**，日均Token消耗量突破**30万亿**，国内多数模型使用中文数据占比达到**60%-80%**，**整体成效明显**。

◆ 从行业角度来看，**医疗卫生、工业制造、智慧能源**等领域建设活跃，**低空经济、具身智能**等领域需求迫切。

我国高质量数据集建设总体呈现出“统筹规划、分层推进、多元协作”的鲜明特点

国家层面，多措并举	地方层面，各具特色	行业层面，协同发展
<ul style="list-style-type: none">➢ 多部门陆续出台多项纲领性政策<ul style="list-style-type: none">• 自2023年12月，国家数据局等部门发布“数据要素×”、数据标注产业、数据资源开发利用等多项政策保障数据资源高质量供给➢ 统筹布局七大数据标注基地<ul style="list-style-type: none">• 建设行业数据集524个，总规模超过29PB，赋能163个应用，带动相关产值超过83亿元➢ 加速促进数据集价值释放<ul style="list-style-type: none">• 截至上半年，全国累计交易额近40亿元，已挂牌3364个高质量数据集，总规模达到246PB	<ul style="list-style-type: none">➢ 奖补激励数据集建设<ul style="list-style-type: none">• 北京、山东、武汉、南京、杭州等11个地区出台补贴政策➢ 推进数据资产化应用<ul style="list-style-type: none">• 上海、天津、安徽等地试点“数据语料作价入股”等新模式，引导企业将数据集折算为股权投资投入公司➢ 打造典型案例示范<ul style="list-style-type: none">• 苏州发布首批30个数据集➢ 数据交易所批量上架数据集<ul style="list-style-type: none">• 北京数交所引入567个数据集，覆盖20多个行业和领域，交易量比例从去年的10%跃升到目前的近80%	<ul style="list-style-type: none">➢ 行业主体成为数据集建设重要力量<ul style="list-style-type: none">• 4月，国资委发布30项央企人工智能行业高质量数据集优秀成果，集中在智慧能源、工业制造、医疗卫生领域• 8月，国家数据局征集遴选出104个高质量数据集典型案例，集中在医疗卫生、工业制造、智慧能源等重点领域和低空经济、具身智能等创新领域➢ 大模型企业、科研机构持续贡献创新<ul style="list-style-type: none">• 上海人工智能实验室开源数据平台OpenDataLab提供5500多个数据集，总数据量达到80TB以上



首席数据官
GOVCDO

目 录

- 一、人工智能发展现状与趋势**
- 二、数据是人工智能飞跃发展的关键驱动力**
- 三、高质量数据集建设方法论**
 - (一) 高质量数据集相关概念
 - (二) 高质量数据集的政策背景
 - (三) 高质量数据集的建设方法**
 - (四) 高质量数据的质量评估方式
- 四、高质量数据集建设案例分析**

高质量数据集的建设方法

—— 1 高质量数据集建设难点

通用领域高质量数据集建设难度较大

- 从互联网、社交媒体和公开数据库中采集的数据，其来源和质量无法得到有效控制。
- 多模态、非结构化数据的处理对技术要求更高。



需要大量人力参与标注，保证“大规模”要求。效率低、成本高



数据来源质量不可控，难以保证“多样性”与“高质量”要求。



AI模型

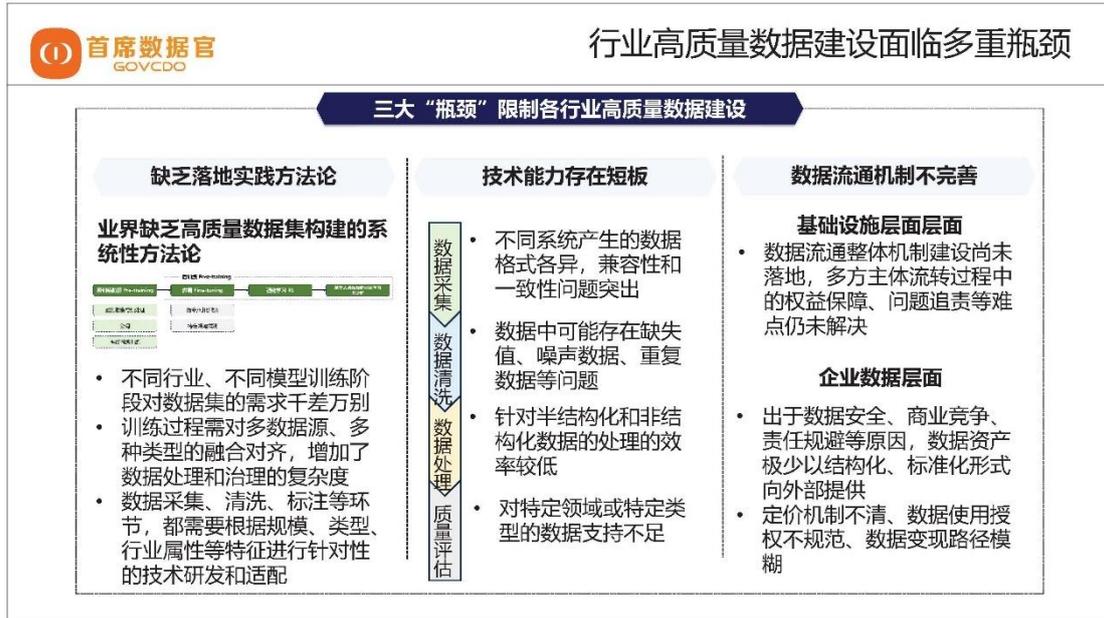


多种模态数据叠加，难以处理。

自动化与智能标注：智能化标注是一种利用人工智能技术，来辅助或自动化数据标注过程的方法。它的核心在于通过预训练的模型对数据进行初步标注，然后由人工进行校正和优化，从而提高标注的效率和准确性。

数据增强：在深度学习和计算机视觉领域中，数据增强扮演着重要的角色。通过使用旋转、翻转、裁剪等技术，可以增加数据的多样性和质量，从而提高模型的泛化能力。

多模态数据处理和分析：通过多模态数据处理和分析技术，可以综合利用不同模态的数据，提高数据分析和预测的准确性和可靠性。





首席数据官
GOVCDO

高质量数据集建设需根据行业特点“因材施教”

◆ 从行业角度来看，数据集的模式分布、标注需求差异较大，需根据行业特点应用不同的数据处理技术和方法，其质量评价也需要在通用的指标上进行定制加强。

行业数据集的模式分布、标注需求差异较大

医疗卫生行业

文本（病例）和影像（CT/B超等）居多

侧重于专业标注、文本解析和图文结合等处理

工业制造行业

时序数据、图像、图纸文档、仿真数据居多

侧重于时序数据处理、高精度合成和专业标注

信息通信行业

音视频（客服）、网络数据居多

侧重于音视频数据内容提取、关键词识别等处理

行业数据集核心质量评价指标存在区别

数据模态不同，指标具体定义不一样

• 文本 • 图像 • 音频 •

行业场景不同，关注核心指标有差异

• 医疗 • 工业 • 能源 •

通用质量评价指标	规范性	准确性	完整性
	一致性	合规性	真实性
	全面性	多样性	独特性
	时效性	均衡性	有效性



首席数据官
GOVCDO

高质量数据建设需要新的方法论

实践层面高质量数据集建设挑战重重

需求设计

- 加工难度大：类型种类多图片、视频、文本、音频等多种格式，每种数据类型的处理和加工要求不同

采集

- 存储要求高：数据文件体积更大、增长速度快，需要提高存储能力和利用率

存储

加工

- 价值密度低：非结构化数据往往无法直接利用，需要面向应用场景进行多轮加工处理、挖掘价值

模型验证

运营

- 安全风险高：存在数据过度采集、样本偏差、隐私泄露等情况，将危害个人、企业甚至社会伦理与安全

传统数据治理方法和评价标准已不完全适用




结构化数据治理体系 DCMM

主要针对结构化数据

需要新型标准规范体系指引




已发布一些基础性标准

行业应用	交易流通
质量价值	运营优化
技术规范	工具平台
建设开发	
建设指引	能力评估
	质量测评

标准体系框架示意

仍需进一步完善标准体系

高质量数据集的建设方法

—— 2 高质量数据集建设关键要素

建设高质量数据集需要多要素协同

构建高质量数据集的环节多，需组建专业、高效协作团队、应用多类技术和建立数据质量的管理制度和标准，通过持续优化运营实现数据集的可用、易用与好用。





首席数据官
GOVCD0

高质量数据集建设核心技术

□ 大模型数据需求3个新特征

◆ **国际主流大模型训练数据量激增近万倍**

纵观国际主流大模型，训练数据规模增长超过**1万倍**。

◆ **大模型的数据需求与传统人工智能相比有显著不同**

大模型通常需要大量数据实现良好的性能，其训练所需的原始数据规模通常在**TB至数百TB**之间；在训练之前，需将文本等原始数据进行**token化**处理。

◆ **大模型数据来源丰富**

大模型数据含**文本、图片、音视频**等多种形式，涵盖海量知识和多语言，使其具备**强通用和迁移能力**，应对广泛任务。如ChatGPT训练数据包括文学、百科、新闻等知识及多模态数据。

面向大模型海量数据需求的破题路径

① 数据标注，提升数据质量



- ✓ 通过对数据进行筛选、清洗、分类、注释、标志和质检等加工处理，显著提升数据质量。
- ✓ 自动化标注和AI辅助标注正在成为数据标注领域的新趋势，不断提升标注效率、质量和安全性。

② 合成数据，增加数据供给



- ✓ Gartner预测，2024年AI训练中用到的数据有60%将是合成数据，到2030年绝大部分训练数据将由人工智能合成。
- ✓ 合成数据有助于节省时间成本和经济成本，获得极端情况下的训练数据，最大限度地减少隐私问题。



首席数据官
GOVCD0

合成数据基本概念

◆ 合成数据是指基于专用数学模型或算法生成的数据，一般可通过深度学习算法生成、统计模型生成、物理仿真模拟生成，具备隐私保护、可扩展数据规模、提升数据获取效率等优点，随着隐私保护重视和人工智能快速发展，合成数据正逐步作为保护数据中的隐私和解决训练数据不足问题的重要技术。

概念定义

定义：合成数据是指基于专用数学模型或算法生成的数据。

特点：通常可以反映出原始数据的**特征**，但与原始数据**没有直接关联**。

注：通过数据切分、图像翻转等基本处理，应用匿名化、差分隐私等数据脱敏手段后，数据仍是原始数据，所以不属于合成数据技术。



技术分类

根据生成方式

深度学习算法生成、统计模型生成、物理仿真模拟生成

根据数据类型

结构化数据：数值类型、时序数据
非结构化数据：文本数据、图像数据、语音数据、视频数据、三维模拟数据
多模态数据

相关技术差异与联系

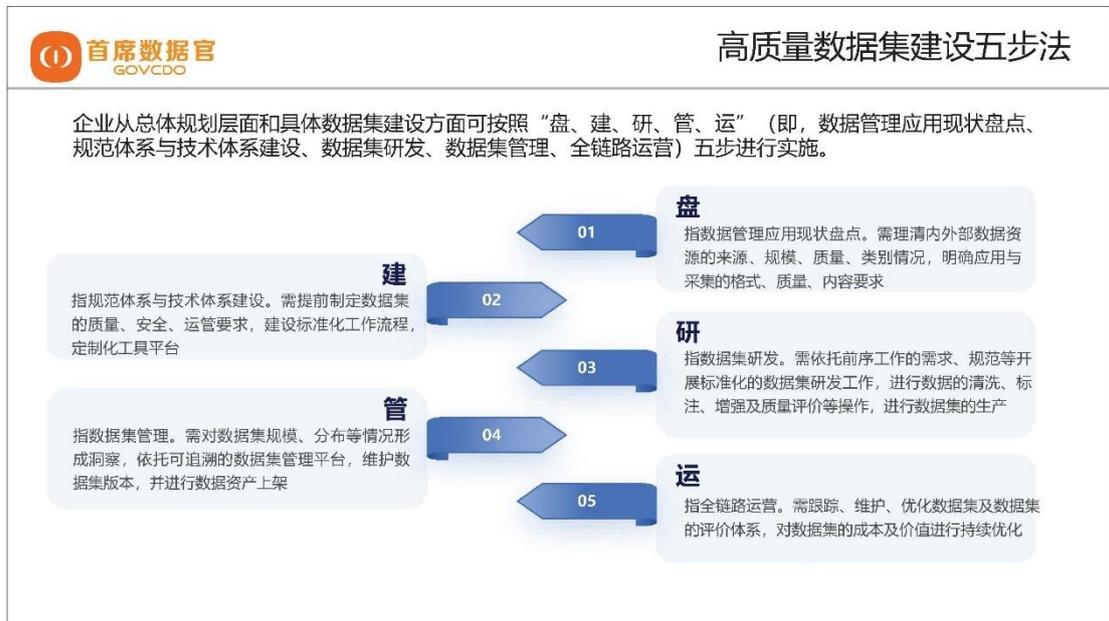
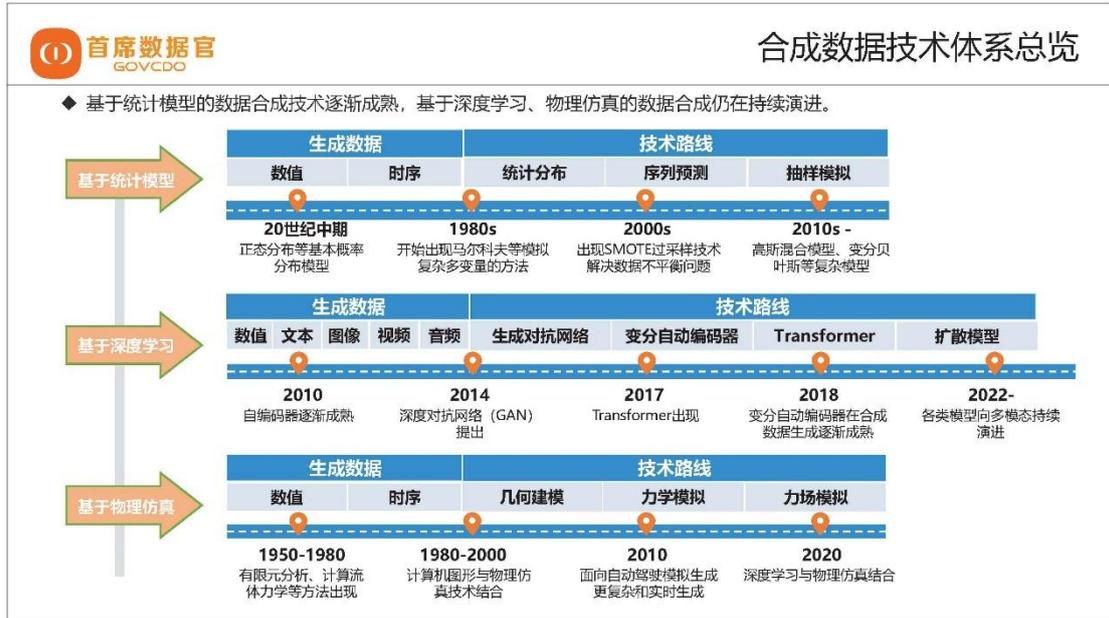
数据增强

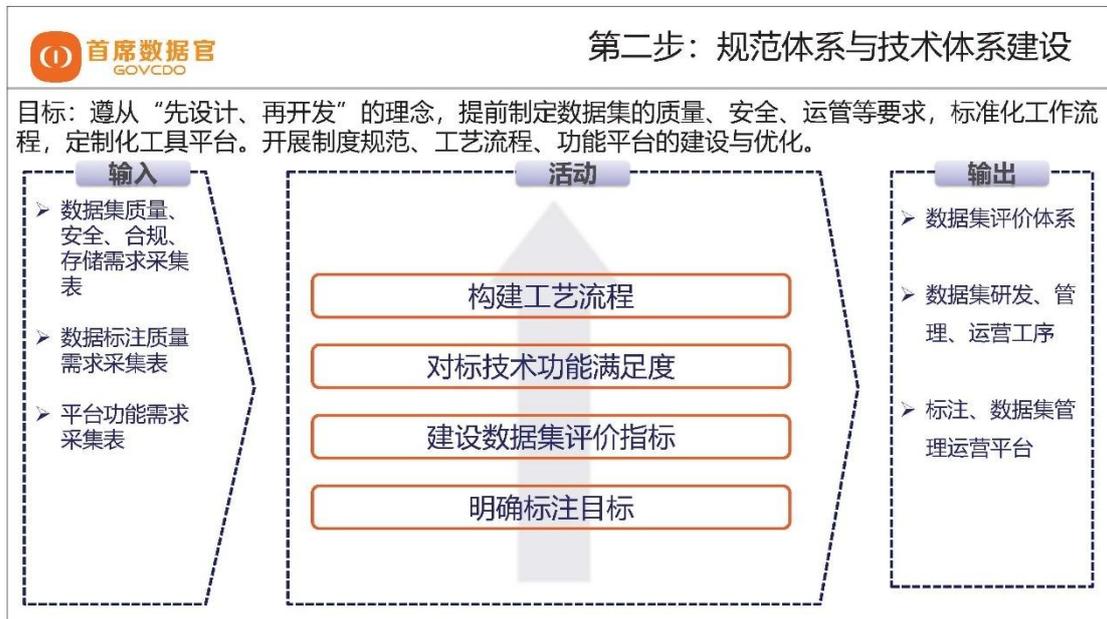
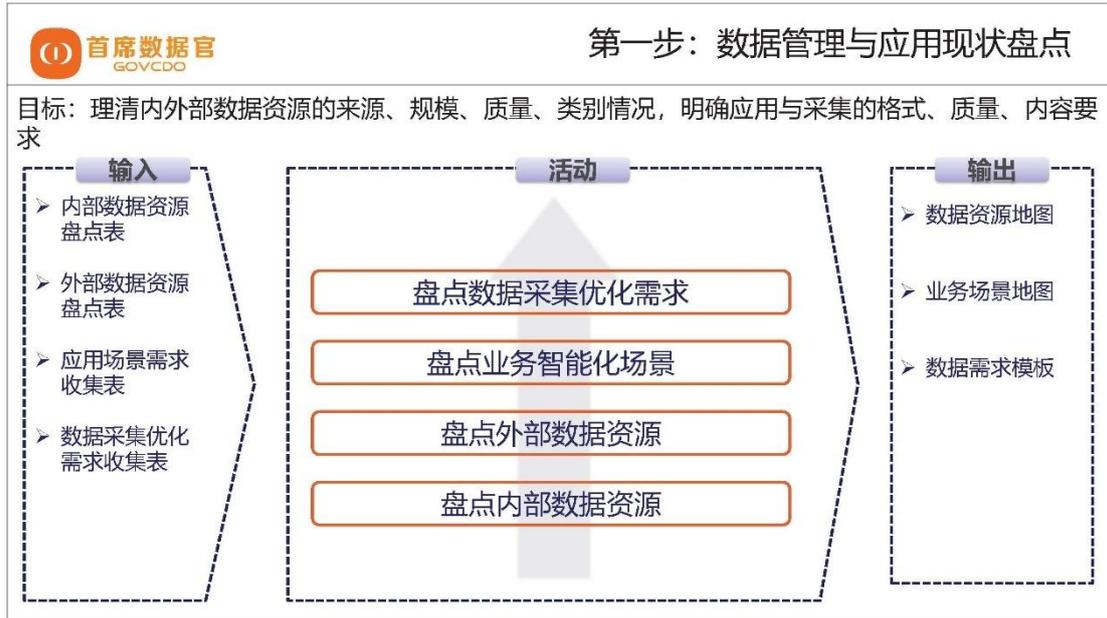
对现有数据进行处理，通过变换生成新的数据样本以提升模型的泛化能力。

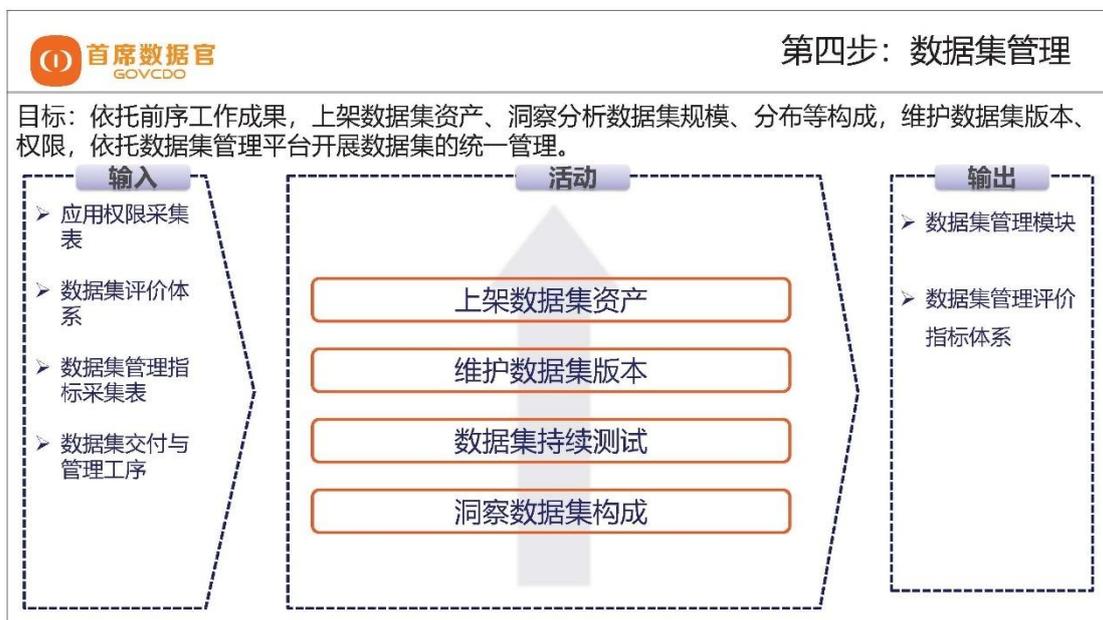
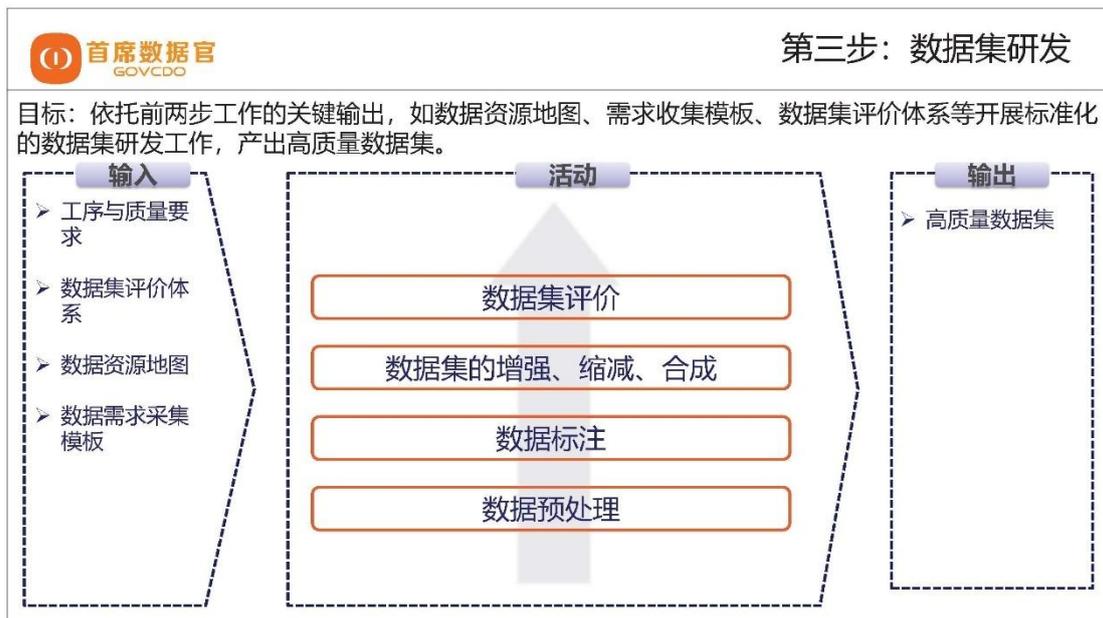
【区别】 场景和技术方法有所不同。

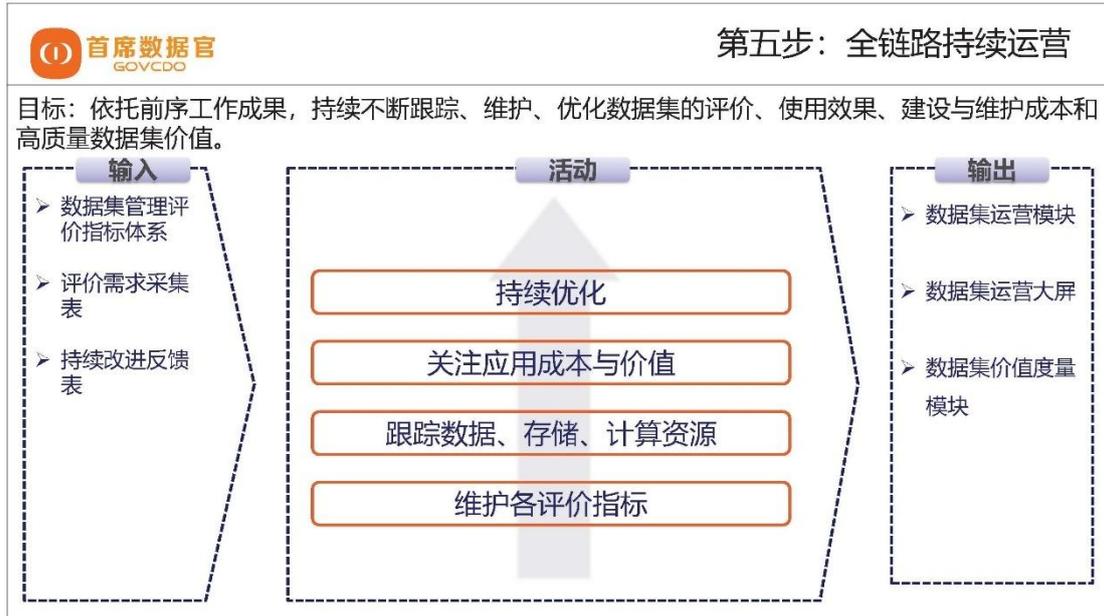
- **数据增强** 技术依赖于已有的数据，通过技术操作**使数据集更具多样性**；
- **数据合成** 则是通过模拟真实世界的过程，**生成与现实相似的新数据**。

【联系】 在实际应用中可以结合使用。如在图像分类任务中，可先通过数据增强扩大数据集规模，然后通过数据合成与现实相似的新数据，提高模型性能。









首席数据官 GOVCDO

高质量数据集的建设方法
—— 3 高质量数据集整体路径



高质量数据集整体建设“三步走”战略 (2/3)

□ **工程建设阶段：**打造高质量数据集生产体系需从三大方向协同发力，以技术驱动、模式创新与生态协同筑牢数据根基：**数据工厂模式奠定规模化生产基础，前沿技术突破质量与场景限制，生态协作整合外部资源**，共同构建起覆盖效率、质量、创新的高质量数据集生产体系。

1

建立高效数据工厂模式，推动规模化生产

- 将数据生产拆解为采集、清洗、标注、质检等模块化环节，每个环节制定精细化操作规范。
- 引入自动化工具提升生产效率，如开发自动化数据清洗脚本，批量处理重复、错误数据；搭建智能标注平台，通过预设标签模板、自动推荐标注结果，减少人工操作成本。

2

探索前沿技术路线，突破传统数据局限

- 合成数据技术模拟真实场景生成稀缺数据，解决数据标注成本高、敏感数据获取难等问题。
- 智能化标注技术借助小样本学习、主动学习等AI算法，辅助人工标注
- 深度推理思维链数据聚焦模型逻辑可解释性，采集记录数据推理过程的思维链信息

3

构建生态协作机制，整合外部资源力量

- 建立开放合作体系，引入专业第三方数据标注服务商。
- 构建科学的供应商评价和管理体系，定期对服务商进行考核，筛选优质合作伙伴，淘汰不达标团队。
- 推动生态内协同创新，与高校、科研机构合作探索数据生产新技术，共同优化数据生产标准。

高质量数据集整体建设“三步走”战略 (3/3)

□ **质量监测阶段：**通过量化评估奠定技术基础，动态机制实现过程把控，持续优化闭环推动迭代升级，三者协同构建严密的数据质量保障体系。

1

量化评估模型和工具，构建科学评估基准

- ① 针对数据生产各环节，设计精细化量化指标体系。
- ② 引入自动化评估工具，开发集成式质量检测平台。

2

动态评估机制，贯穿事前事中事后全流程

- ① 事前规划，生产前明确质量标准与流程。
- ② 事中监控，实时监测系统，追踪标注进度、清洗效果。
- ③ 事后复盘，对比质量目标与实际结果，分析偏差原因

3

持续优化闭环，建立数据与模型的反馈联动

搭建反馈评价体系，将模型训练、应用结果反向传导至数据生产环节。形成“模型反馈—质量诊断—流程优化”的闭环。

高质量数据集整体建设“三步走”战略 (2/3)

□ **工程建设阶段：**打造高质量数据集生产体系需从三大方向协同发力，以技术驱动、模式创新与生态协同筑牢数据根基：**数据工厂模式奠定规模化生产基础，前沿技术突破质量与场景限制，生态协作整合外部资源**，共同构建起覆盖效率、质量、创新的高质量数据集生产体系。

1

建立高效数据工厂模式， 推动规模化生产

- 将数据生产拆解为采集、清洗、标注、质检等模块化环节，每个环节制定精细化操作规范。
- 引入自动化工具提升生产效率，如开发自动化数据清洗脚本，批量处理重复、错误数据；搭建智能标注平台，通过预设标签模板、自动推荐标注结果，减少人工操作成本。

2

探索前沿技术路线，突破 传统数据局限

- 合成数据技术模拟真实场景生成稀缺数据，解决数据标注成本高、敏感数据获取难等问题。
- 智能化标注技术借助小样本学习、主动学习等AI算法，辅助人工标注
- 深度推理思维链数据聚焦模型逻辑可解释性，采集记录数据推理过程的思维链信息

3

构建生态协作机制，整合 外部资源力量

- 建立开放合作体系，引入专业第三方数据标注服务商。
- 构建科学的供应商评价和管理体系，定期对服务商进行考核，筛选优质合作伙伴，淘汰不达标团队。
- 推动生态内协同创新，与高校、科研机构合作探索数据生产新技术，共同优化数据生产标准。

高质量数据集整体建设“三步走”战略 (3/3)

□ **质量监测阶段：**通过量化评估奠定技术基础，动态机制实现过程把控，持续优化闭环推动迭代升级，三者协同构建严密的数据质量保障体系。

1

量化评估模型和工具，构建科学评估基准

- ① 针对数据生产各环节，设计精细化量化指标体系。
- ② 引入自动化评估工具，开发集成式质量检测平台。

2

动态评估机制，贯穿事前事中事后全流程

- ① 事前规划，生产前明确质量标准与流程。
- ② 事中监控，实时监测系统，追踪标注进度、清洗效果。
- ③ 事后复盘，对比质量目标与实际结果，分析偏差原因

3

持续优化闭环，建立数据与模型的反馈联动

搭建反馈评价体系，将模型训练、应用结果反向传导至数据生产环节。形成“模型反馈—质量诊断—流程优化”的闭环。

首席数据官 GOVCDO 《建设指引》凝聚共识，为高质量数据集建设领航定向

◆ 虽然我国高质量数据集建设已取得初步成效，但在数据供给、技术工具、标准规范、安全合规、商业模式等方面仍然面临多重困难与挑战，因此亟需解决瓶颈问题、总结普适的方法论，力争为人工智能纵深发展提供有力支撑。

困难与挑战

数据供给 专业领域数据储备量不足，跨部门、跨区域共享机制不健全	技术实现 混合结构数据处理能力与工具链水平薄弱，智能化程度低
标准与治理 标准规范不完善，标准应用力度不足，缺乏指导	安全与合规 风险控制与开放需平衡，数据权属规则不清晰
成本与模式 投入产出比失衡，未形成成熟的可持续运营模式	

主要编写单位

中国信息通信研究院、国家数据发展研究院、中国电子技术标准化研究院、国家信息中心、国家发展和改革委员会创新驱动发展中心、中国电子信息产业发展研究院等

高质量数据集建设指引

2025年8月

首席数据官 GOVCDO 《建设指引》提出高质量数据集建设运营的参考路径

◆ 《建设指引》提出高质量数据集建设“1+1”的参考路径，即1套建设方法论和1套建设运营体系，通过分析建设的典型模式、核心环节、关键技术和质量评价等内容，勾勒出具体的路径，为企业建设数据集提供清晰的实践方法。

1套高质量数据集建设方法论

6个核心环节

- 数据需求
- 数据规划
- 数据采集
- 数据预处理
- 数据标注
- 模型验证

5大关键技术

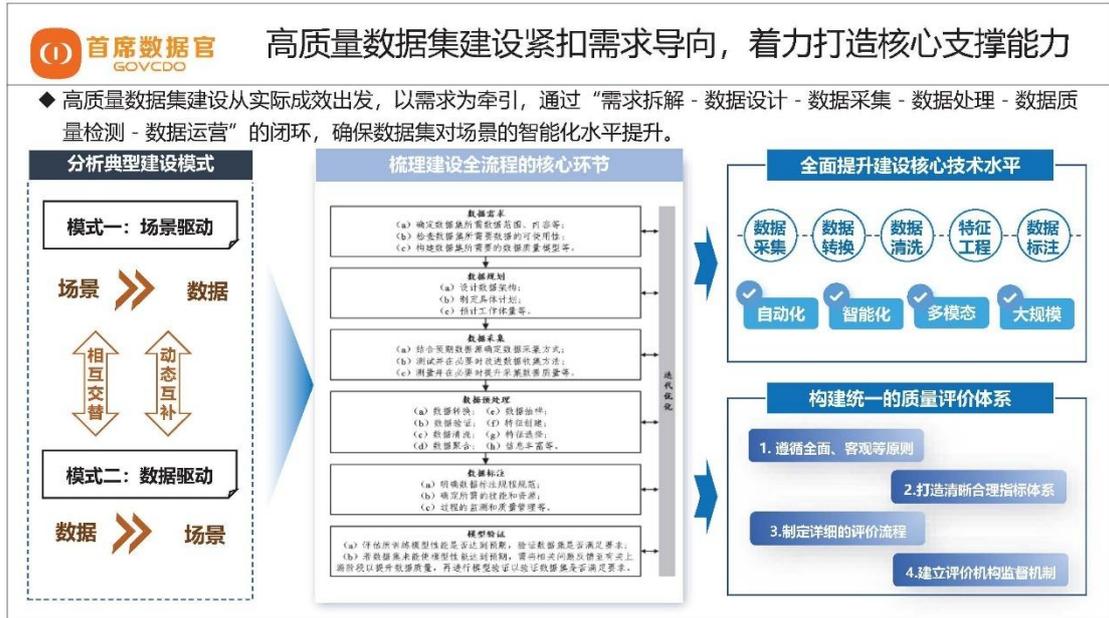
- 数据采集
- 数据转换
- 数据清洗
- 特征选择
- 数据标注

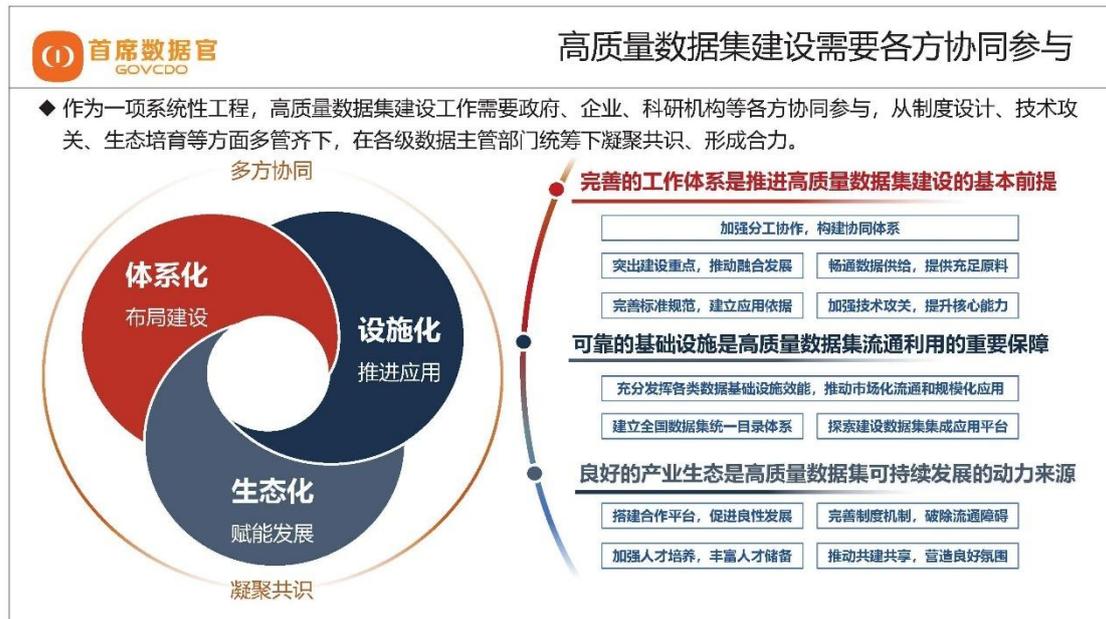
1套质量体系

- 质量评价实施流程
- 质量评价指标体系
- 质量评价工作体系

1套高质量数据集建设运营体系

- 体系规划**
 - 建设运营的前提
通过知识索引构建、数据资源盘点、标准体系搭建，为后续具体建设工作提供清晰蓝图
- 工程建设**
 - 实施的技术保障
围绕数据研发、交付、运维核心环节，探索前沿技术，突破场景局限，推动规模化生产
- 运营管理**
 - 可持续发展的核心
构建全流程管理体系，达到需求响应及时、成本精准可控、质量安全可信、生态价值共创





首席数据官 GOVCDO

目录

- 一、人工智能发展现状与趋势
- 二、数据是人工智能飞跃发展的关键驱动力
- 三、高质量数据集建设方法论**
 - (一) 高质量数据集相关概念
 - (二) 高质量数据集的政策背景
 - (三) 高质量数据集的建设方法
 - (四) 高质量数据的质量评估方式**
- 四、高质量数据集建设案例分析



首席数据官
GOVCD0

高质量数据集 数据质量评估方法

◆ 《高质量数据集 数据质量评估方法》规范了数据集数据生成过程的基础要求、质量评估的指标体系和评估方法。建立了**4大评估维度**，包括**12个一级指标**，并对非结构化、指令微调等各数据类型设计了至少**20余种二级指标**。

适用范围

- 本文件适用于对数据集的质量实现客观、公正的评估，确保用于数据集**合规、准确、人类对齐且具有高信息量**等属性。

评估方法



Step1 符合性审查是评估的基础门槛，数据集需在通过符合性审查的基础上进行指标项打分

Step2 指标项打分是对数据集作进一步评估等级划分的依据；每一级的评估标准均为前一级别的增强

高质量数据集评估方法

数据质量评估指标框架



数据质量评估工具





首席数据官
GOVCD0

高质量数据集 建设运营能力成熟度模型

◆ 通过评估体系有助于企业引入前沿的数据管理思维与手段，精准衡量并定位自身在高质量数据集建设运营方面的行业水准。同时，也将为企业高质量数据集建设与运营能力提升明确路径与目标，引领企业提高数智化水平。

评估指标框架

《高质量数据集建设运营能力成熟度模型》从组织管理、技术服务、数据安全、标准规范、运营管理、生态建设6大能力域，明确数据集开发、运维、运营、管理等环节关键要求，搭建企业高质量数据集建设运营能力评估框架。



评估方法及结果

- 对每个指标进行**分级**的形式，通过**级别与权重、分数**进行挂钩，最终确定评估结果的**级别（五个）**。

分级	(技术服务-数据采集) 能力项内容
L1 (初始级)	采集缺乏流程管控、主要依赖个人经验
L2 (可控级)	具有初步的采集计划和流程、数据质量检测机制
L3 (规范级)	具备明确的组织级采集管理制度和流程、完善的数据质量检测机制、采用了一定的自动化工具
L4 (优秀级)	具备可量化、可执行的数据采集步骤和方法、定期调整数据采集过程、建立数据地图、完善的质量监控体系
L5 (卓越级)	根据业务优化采集方法、采用自动化和智能化工具、完善的数据质量体系、快速可复制的成熟采集模板、注重采集过程的可持续发展、等等

成熟度等级说明

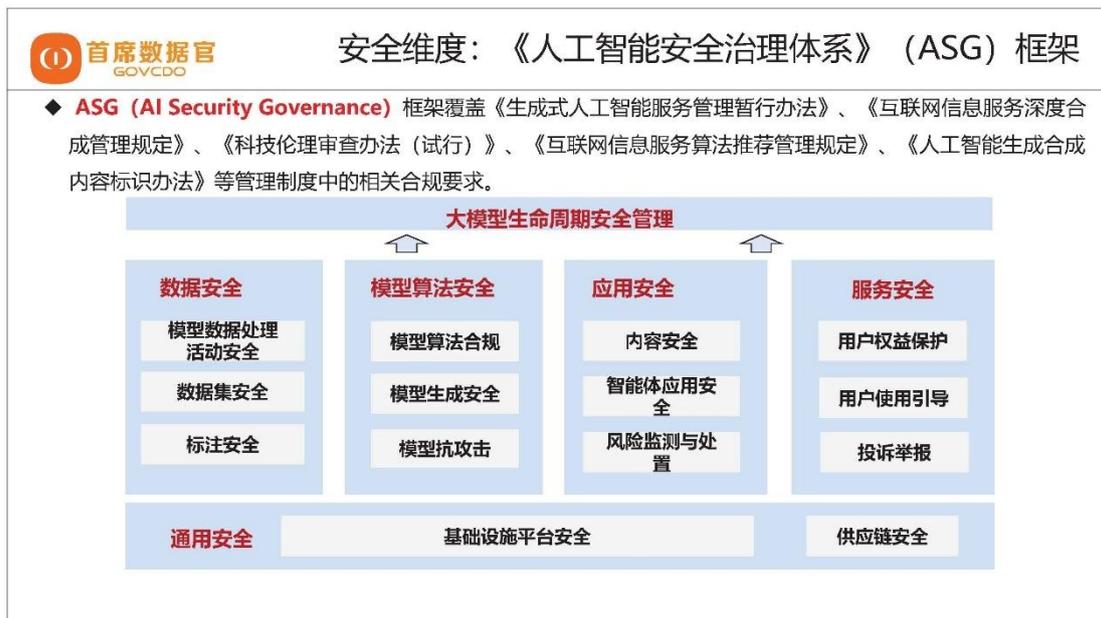
初始级

可控级

规范级

优秀级

卓越级



首席数据官 GOVCDO **目 录**

- 一、人工智能发展现状与趋势
- 二、数据是人工智能飞跃发展的关键驱动力
- 三、高质量数据集建设方法论
- 四、高质量数据集建设案例分析**

如何充分挖掘海量移动通信信令数据的位置信息价值，赋能交通场景智慧应用？

1. 随着“数据要素×”计划和“人工智能+”行动的推进，高质量数据集已成为训练人工智能模型、释放数据价值的关键支撑。
2. 通信信令数据蕴含着大规模群体7×24小时连续的动态位置信息，区域全 覆盖广 周期长 体量大，具有重要应用价值。
3. 构建高质量的人时空三元组数据集有助于洞察全国10亿+人群的出行特征，赋能交通出行、客流预测等行业应用。

1、政策背景

“数据要素×三年行动计划（2024—2026年）”（2023年12月国家...）

《数据要素×三年行动计划（2024—2026年）》是为深入贯彻落实习近平总书记关于发展数字经济的重要指示精神，推动数据要素乘数效应，发挥数据要素乘数效应，赋能经济社会发展，特别制订的行动计划。2023年12月31日，国家数据局等17部门联合印发《数...》

发布日期 内容全文 内容解读

数据要素×三年行动计划

2025-04-30 10:12:48 更新

【国家数据局将联合国务院国资委推动央企高质量数据集建设】

财联社4月30日电，国家数据局局长刘烈军4月30日在第八届中国数字经济论坛上表示，国家数据局将联合国资委推动央企高质量数据集建设，充分发挥数据要素乘数效应，赋能经济社会发展。刘烈军表示，国家数据局将支持央企围绕业务场景需求，开展点对点数据交易，鼓励打造行业数据交易平台，支持数据要素市场化配置改革，打破数据壁垒，畅通数据要素流通渠道，鼓励高校、企业、第三方科研机构等大力开展数据技术创新，数据要素流通，数据要素开发利用产业生态。

AI+行动部署 央企高质量数据集建设

2、数据资源现状

数据优势：大样本 全过程 广覆盖

运营商海量信令数据

匿名用户...
时间记录...
地理区域...

手机设备 (IMEI)
eNB基站

人均200+条数据/日
日采集数据达5.6PB
数据体量优势 累计数据2EB+

3、案例意义

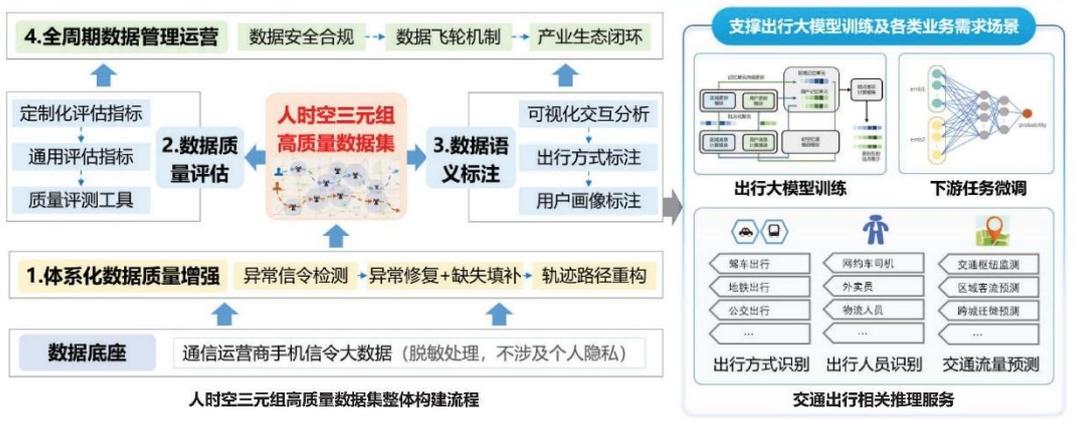
PIID	时间	基站经纬度	基站类型
001	07:15	(113.8,121)	4G-LTE
001	07:45	(113.8,121)	4G-LTE
001	07:45	(113.8,121)	4G-LTE
002	10:00	(113.8,121)	4G-LTE
001	10:30	(113.8,121)	4G-LTE

高质量人时空数据 → 赋能交通运输领域

跨城出行分析 区域客流预测

交通枢纽洞察 城市通勤监测

从数据质量增强、数据质量评估、数据语义标注和数据管理运营四个方面入手，将原始信令数据转换为可准确反映个体出行轨迹及人地交互关系、具有丰富活动语义标签的高质量人时空三元组数据集，并实现数据全生命周期的顺物流转与价值释放



首席数据官 GOVCDO 创新点1：构建体系化数据质量增强技术，夯实高质量数据底座

通过系统梳理海量信令数据存在的质量问题，从位置异常、数据缺失、轨迹稀疏等不同层面，以**多规则综合判别模型**结合**深度学习模型**的方式，提升数据准确性和完整性，为出行大模型训练和各行业下游任务推理奠定坚实数据基础。

➤ **挑战：**原始信令数据**体量大，质量低** ✓ **突破：**通过构建体系化数据质量增强技术，形成超**3千亿条**约**50TB**高质量数据。

原始信令质量问题梳理

问题	产生原因
① 信令位置缺失	信令未能匹配到基站经纬度
② 信令位置异常	基站经纬度标记错误
③ 信令位置异常	手机信号远距离异常飘移
④ 信令数据缺失	用户长时间未产生信令记录
⑤ 用户位置偏差	基站服务范围决定了用户实际位置与基站位置的偏差

效果：信令平均距离误差从**800m**缩小至**160m**，**500m**以内占比达**96%**。
轨迹拟合精度由**50%**提升至**97%**，完整度从**90%**提升至**100%**。

3项专利：《异常信令检测方法、装置、设备和可读存储介质》
《一种信令数据的异常检测、修复和缺失填补方法及系统》
《一种基于时空卷积网络的手机信令轨迹路径重构方法》

体系化信令质量增强技术

首席数据官 GOVCDO 创新点2：设计多元化数据质量评估体系，全方位评估数据质量

1. 结合行业标准和数据特性，设计一套**定制化和通用性相结合**的多元化质量综合评估体系
2. 开发对应**评估工具**，实现对全方位人时空三元组数据集进行质量评测和校验，确保评估结果的准确性和可靠性

➤ **挑战：**信令数据质量评估难

高质量数据集及手机信令数据相关国家、行业标准

高质量数据集 建设指南 High-quality Dataset—Construction guidelines	高质量数据集 质量评测规范 High-quality dataset—Specification for quality evaluation and use	自然资源领域应用手机信令数据技术指南 Technical guidelines for the application of mobile signaling data in the field of natural resources
---	--	---

人时空三元组数据质量评价指标体系

指标类型	指标名称	指标内容描述
通用性指标	数据规模指标	检测数据条数是否满足要求
	结构完整性	不包含缺失值或缺失值应在合理范围内
	格式规范性	数据格式符合预定标准，可直接用于人工智能模型开发和训练
	内容真实性	数据真实，能追溯到采集源头，且能与采集源头保持一致
	数据唯一性	数据经过清洗处理，不包含重复数据，数据之间重复率为0
定制化指标	数据合规性	数据已做加密和脱敏处理，不涉及用户个人隐私信息
	数据准确性	数据集中非异常信令的占比

通用性指标(before)

存在问题

- 未考虑信令数据时空特性
- 已有质量指标无法直接适配
- 缺乏标准化的信令质量评价体系

定制化指标(after)

取得成效

- 综合考虑时间完整度和空间准确性
- 面向信令数据进行改造和适配
- 形成标准化的信令质量评价体系

首席数据官 GOVCO 创新点3：打造专业化时空信令标注工具，生成多维度数据标签

1. 综合考虑信令数据的专业性，以GIS地图结合交互式问题引导的方式，打造一款专门面向信令数据的**可视化、智能化标注**等工具
 2. 可支撑对**出行方式**和**职业画像**的高效标注，生成具有丰富活动语义标签的出行标注数据集，联合打造**1个省级数据标注基地**

➢ **挑战：**信令数据活动语义缺失 ✓ **突破：**打造面向信令数据的轨迹语义标注工具，自动化率达**70%**，标注效率提升**90%**。

系统架构

出行方式标注+职业画像标注

专利《面向手机信令数据的轨迹语义标注方法及系统》

- **地图可视化UI：**结合标注任务内容，以时空可视化的形式直观分析用户轨迹分布情况。
- **多源信息融合：**融合通话行为、上网行为、AOI/POI等多源数据进行辅助标注。
- **批量操作和展示：**表格化展示标注数据，支持一次多用户多天进行数据标注。
- **交互式问题引导：**以交互问答形式辅助标注人员清晰快速适应不同类型的标注任务。

首席数据官 GOVCO 创新点4：形成全周期数据管理运营机制，保障数据安全与流通

1. 构建1套覆盖数据流通全生命周期的**数据安全合规管理运营机制**，通过数据脱敏、访问控制等手段保证数据安全
 2. 对内形成集数据采集、清洗、标注、供给、回流、迭代优化为一体的**数据飞轮机制**，持续打磨提升数据质量
 3. 对外形成从数据采集到智能处理再到场景化输出的**产业生态闭环**，赋能**16家省**外部单位，发挥数据乘数效应

➢ **挑战：**信令数据敏感等级高 ✓ **突破：**建立全周期数据管理运营机制，在保证数据安全合规的基础上，充分释放数据价值。

对外：产业生态闭环

对内：数据飞轮机制

数据安全合规管理审查机制

- ① **数据加密脱敏**：涉及用户身份识别相关的字段进行加密和脱敏处理
- ② **数据保险箱**：数据按敏感等级管理并设置专门的数据保险箱
- ③ **数据访问管控**：数据使用者按照最小化权限按需访问权限内数据
- ④ **数据操作管控**：记录用户操作日志并开展自动化风险监测和审查
- ⑤ **数据共享管控**：数据下载及共享采用严格的多人审批管理机制



案例2：某大型电网企业构建高质量负荷预测数据集的背景

■ 背景意义：高质量的负荷预测是保大电网安全、保社会民生用电的重要基础

- **支撑大电网核心业务**：为大电网调度指挥控制、电力现货市场运营、电网规划发展、客户用电服务等提供重要边界和依据。
- **助力国家治理现代化**：通过**用电负荷与GDP**等关联分析，利用**大数据技术**挖掘社会经济运行规律、辨识供电薄弱环节（**何时何地存在拉闸限电风险**）、提升政府社会治理水平。

用电负荷是国民经济“晴雨表”



国家对负荷预测、负荷管理的新要求



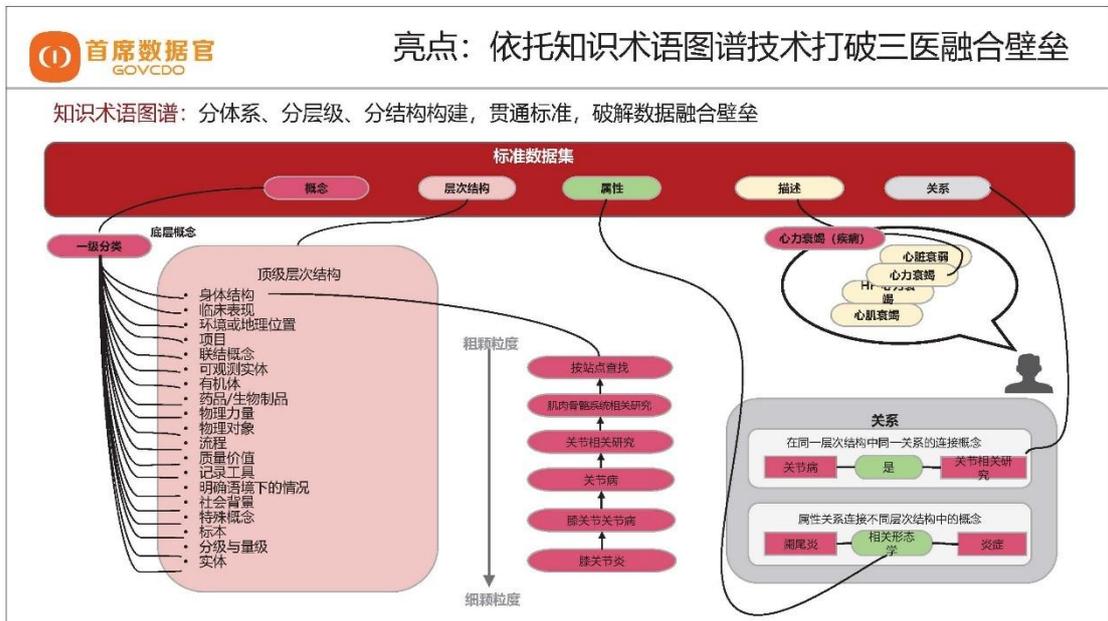
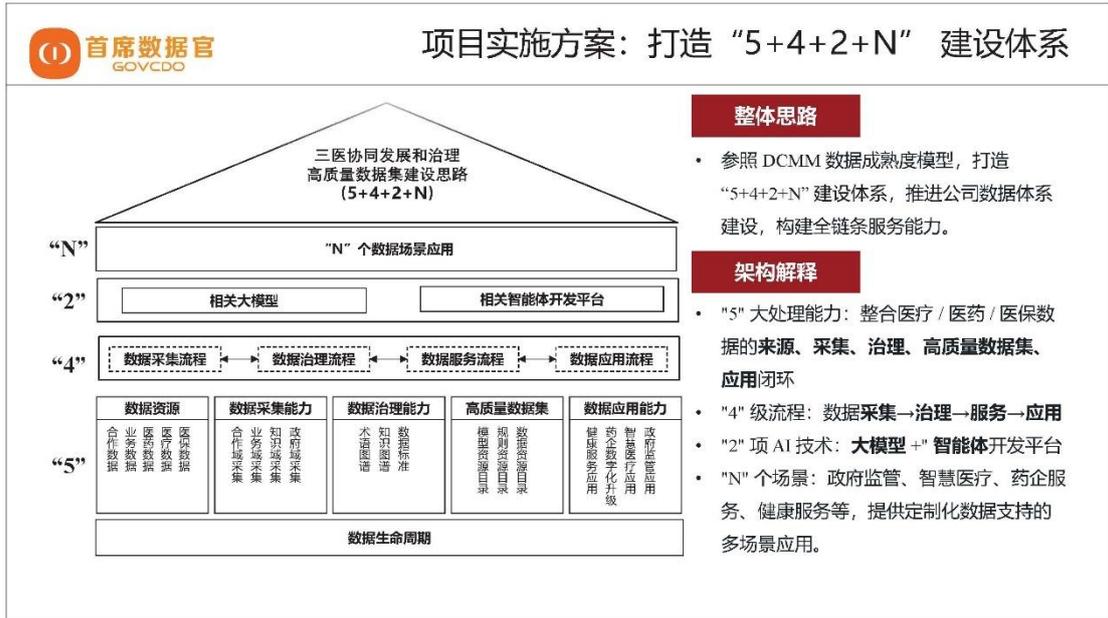
解决方案：统一标准、统一体系、构建网省地一体化负荷预测数据集

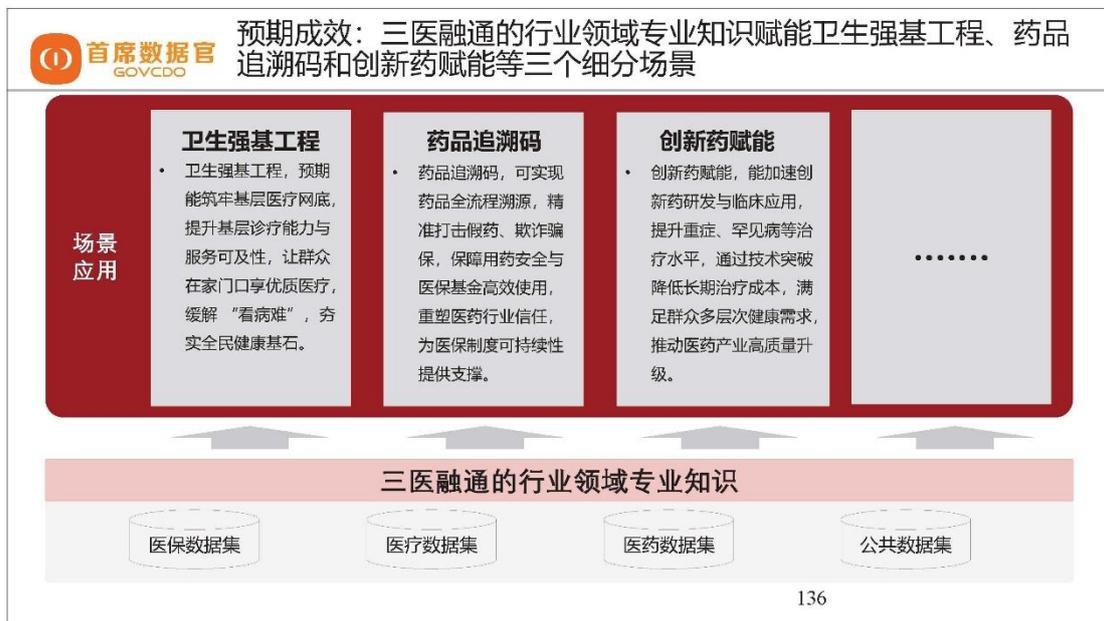
包含**相关省份15年**电网负荷**历史数据**和**实时增量数据**。融合**调度、市场、规划**等5个领域的新能源、气象、行业负荷、业扩报装、经济指标等多维数据。总量超300TB，数据重复率<0.01%。

打破5个部门间数据壁垒，跨域多模态数据融合









首席数据官 GOVCDO 创新点总结

打破传统壁垒，构建全链条创新体系，为三医协同提供新模式与新路径，有效提升整体运行效能，推动医疗服务迈向高质量、一体化发展新阶段。

- 采** 三医数据**采集**：多源异构数据采集**构建**知识引擎
- 治** 三医数据**治理**：大模型**赋能**数据治理自动化
- 管** 三医数据**管理**：区块链 + 联邦学习**保障**数据安全
- 融** 三医数据**融合**：三医编码映射 + 知识术语图谱**破解**融合壁垒
- 用** 三医数据**应用**：AI 智能体**增强**数据应用创新



数据价值实现的湖北实践



The table of contents is set against a dark blue background with a city skyline silhouette. The title '目/录 CONTENTS' is centered at the top. Below it, five numbered items are listed in a row:

01	02	03	04	05
湖北大数据集团简介	政策解读	数据基础概念与数据要素市场	数据资源产业化改革实践路径	案例参考

01

湖北大数据集团简介

“湖北大数据集团”简介

首席数据官

2025年6月6日，在湖北省委、省政府的指导下，湖北大数据集团正式成立，该集团致力于打造成为全国一流、中部领先的省级大数据集团。作为全省数据产业链主企业，湖北大数据集团将在推进湖北数字化、产业数字化的进程中当好主力军。



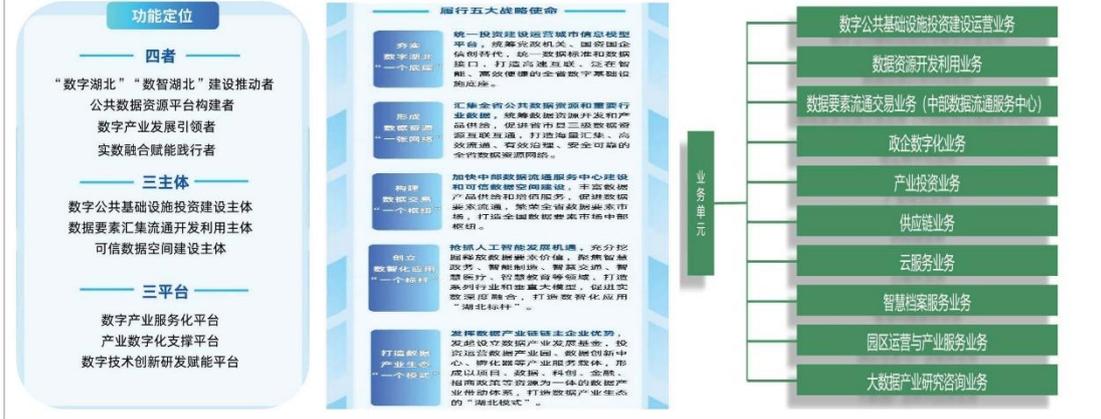
湖北大数据集团由省政府国资委联合湖北联投集团、长江产业集团、湖北交投集团、湖北文旅集团、湖北宏泰集团、湖北港口集团等6家省属企业共同出资组建，注册资本50亿元。



“湖北大数据集团”简介

首席数据官

- 2025年内，湖北大数据集团将完成省级公共数据统一授权运营，完成国家（武汉）新型互联网交换中心、算力互联互通平台、可信数据空间等基础设施建设。
- 到2027年，湖北大数据集团将在智慧政务、智能制造、智慧医疗等领域打造10个国家级数字标杆项目，力争数据交易额突破10亿元，打造1至2个数据特色产业园，培育2至3家数字领域上市“金种子”企业。
- 2030年，湖北大数据集团力争实现年营业收入200亿元、利润总额10亿元，带动产业投资1000亿元以上，促进湖北数字经济核心产业增加值突破1万亿元。



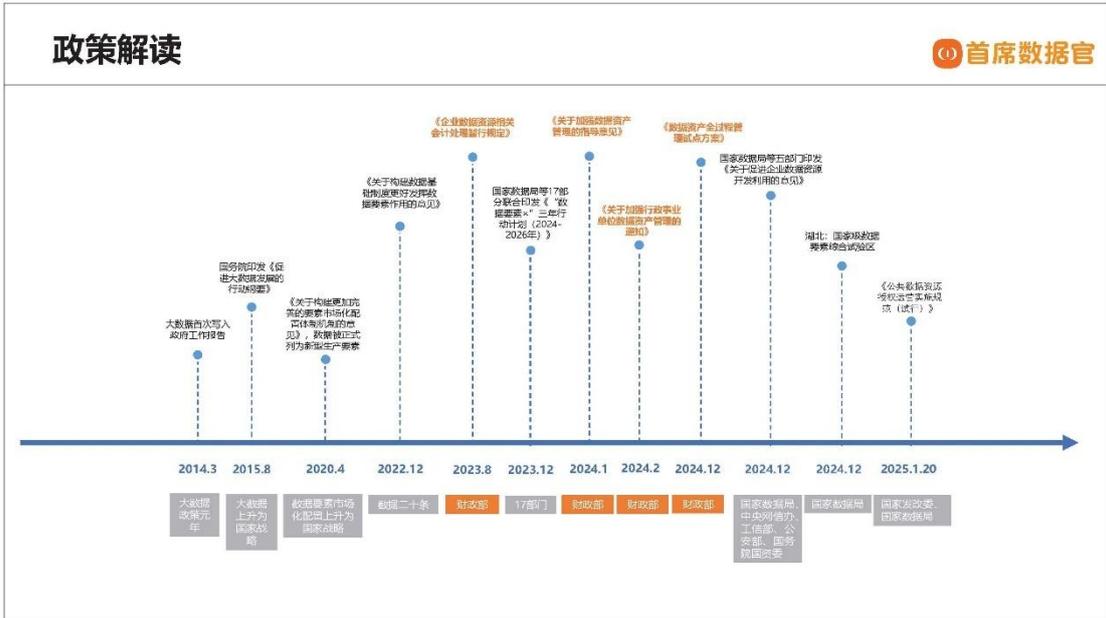
“中部数据流通服务中心”简介

首席数据官

“中部数据流通服务中心”（简称“中部中心”）是在湖北省委、省政府的指导下于2024年9月揭牌成立，定位为省级数据流通综合服务机构，由省数据局负责业务指导与管理。“中部数据服务（湖北）有限公司”（简称“中数服”）作为中部中心运营实体，隶属于湖北大数据集团二级业务单元，下设“数据产权登记、数据流通交易、数据价值实现、数据综合评估、数据跨境流通、数据人才培养”六大子中心，定位为湖北数据要素市场具有公共服务属性的功能保障主体。

中部中心以“线上‘淘宝网’、线下‘广交会’”为指导思想构建形成“343”数据价值实现服务体系，推动数据流通利用，培育数据要素市场，加快构建省内一体、省外互联的数据交易市场。截至目前，中部中心已助力湖北交投、湖北港口、湖北联投、九州通集团、宜昌产投、襄阳国投、咸宁城发、黄冈国投、光谷建设、光谷交通、洪山科技、武汉车联网公司、武昌首义科创投等行政事业单位，国有企业完成数据产权登记，数据资产入表，数据资本化案例超40例。





首席数据官

2.1 湖北政策

政策解读

首席数据官

2024.11.14——省数据局 省财政厅印发《行政事业单位数据资源入账卡工作指引》、《国有企业数据资产入账入表工作指引》的通知

《行政事业单位数据资源入账卡工作指引》

- 行政事业单位应准确把握政策导向，依法依规清理盘活数据资源，尊重“数据变资源、资源变资产、资产变资本”的发展规律，以场景为牵引，促进行政事业单位数据资源开发利用，以应用成效带动价值发现，充分发挥数据资源在赋能实体经济、扩大消费需求、提升治理能力等场景的要素作用。
- 本指引中“数据资源入账卡”是指行政事业单位将数据资源纳入国有“三资”总账，登记数据资源卡片，明确数据资源基本信息、权利信息、开发利用信息、管理信息等。国家有关部门出台行政事业单位数据资源会计处理规定后，行政事业单位可在入账卡基础上开展数据资产入表工作。

《国有企业数据资产入账入表工作指引》

- 国有企业数据资源是企业发展的重要生产要素，推动数据资产入账入表：
 - 有利于摸清我省国有企业数据资源总量、准确计量其数据资产价值、合理形成数据资本；
 - 有利于激发国有企业数据资源创新活力，健全国有企业数据权益实现机制，提升国有企业竞争力、赋能产业和经济社会高质量发展。
- 本指引以国有企业数据资源“确认、确权、确值”为切入点，明确了数据资产识别与确认、数据资源（资产）产权登记、数据资产价值评估和数据资产入账入表相关实施路径，以及数据资产管理方法。



政策解读

首席数据官

2025.3.11——湖北省数据局发布《关于开展数据要素型企业（第二批）入库登记的通知》

- 申报类型：数据资源企业、数据技术企业、数据服务企业、数据应用企业、数据安全企业、数据基础设施企业。
- 申报方式：通过中部数据流通综合服务平台数据要素型企业入库登记子系统（地址：hubei-data.com）。
- 申报时间：3月17日 - 4月7日。
- 培育扶持：
 - （一）政策激励。**优先支持参与省数据资源开发利用试点，支持并撮合打造可信数据空间，示范场景择优推荐申报国家、省数据产业相关项目。提供主动式、针对性服务，包括产业扶持政策（待出台）免申即享、梯次培育、解读宣讲等。
 - （二）业务培训。**充分运用数商联合会、中部数据流通服务中心等各类资源和平台，以线上线下多种方式开展数据业务培训，共享最新产业动态，提供数据价值化服务，拓展入库企业数据业务发展思路及数据主营业务能力。
 - （三）供需对接。**优先推荐入库企业入驻数据流通交易场所和平台，为企业提供场景、数据、技术、产品等供需对接渠道，择优推荐参与国家级、区域级数据要素相关重大供需对接活动。
 - （四）宣传推广。**认定发布一批我省数据龙头企业、骨干企业清单，选树行业标杆。引导入库企业参加“数据要素×”等典型案例示范以及各类企业榜单遴选活动，宣传推广入库企业的精品案例、创新产品。



政策解读



2025.3.11——湖北省数据局发布《关于公开征集第二批湖北省高质量数据集的通知》

征集目的

遴选形成湖北省高质量数据集名录，并公开发布。支持开展供需对接和项目建设，发挥典型示范带动作用。

重点领域

主要面向科学研究、工业制造、农业农村、智慧能源、交通运输、金融服务、医疗卫生、教育教学、商务领域、人力资源、文化旅游、应急管理、气象服务、绿色低碳、公共安全、城市治理等重点领域。

数据集要求

- 1. 质优量大。数据集建设完成并具备定期更新机制，具有实际的业务应用场景，具备一定先进性、创新性、规模性和示范性，适合向社会推广，能够支持开展大模型开发和训练。
- 2. 征集数据类型包括结构化数据、文本、视频、音频、图形图像、其他等。



政策解读



2025.4.14——湖北省财政厅印发《湖北省数据资产全过程管理试点方案》的通知

一、试点任务

- 时间阶段：2025年4月至2026年12月；
- 范围：有关省直部门、省属企业、部分地区财政部门；
- 总体目标：规范管理流程，形成工作指引，打造典型案例，完善制度基础；
- 阶段目标：到2025年底，总结试点地区、单位开展数据资产管理基础工作实践经验，同步形成交易流通、收益分配有效案例，探索形成我省数据资产台账及登记相关制度；2026年底，推动数据资产登记工作在全省铺开，探索初步形成我省数据资产授权运营、交易流通、收益分配相关机制。

二、试点范围

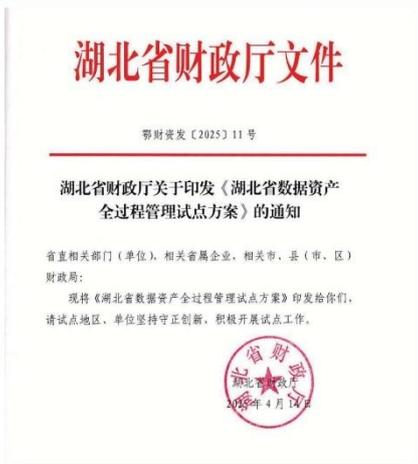
- 1. 市级(10个)：武汉市、襄阳市、黄石市、孝感市、荆州市、荆门市、随州市、鄂州市、咸宁市、十堰市。
- 2. 县级(20个)：洪山区、江夏区、黄陂区、硚口区、武汉市经济开发区、东湖高新技术开发区、樊城区、老河口市、夷陵区、点军区、枝江市、宜都市、茅箭区、竹溪县、郧西县、房县、荆州区、通山县、英山县、大冶市。
- 3. 省直部门(8个)：省发改委、省公安厅、省民政府、省住建厅、省交通运输厅、省农业农村厅、省卫健委、省农业事业发展中心。
- 4. 省属企业(7个)：湖北联投(湖北数据集团)、湖北交投、湖北铁路集团、湖北港口集团(湖北供应链物流公共信息服务有限公司)、湖北机场集团、湖北宏泰集团(湖北省融资再担保集团、华中文交所、武汉知识产权交易所)、湖北文旅集团。

三、工作机制

省财政厅会同省数据局、省发改委等单位建立工作协调机制，通过专题会议、实地调研、座谈研讨、经验交流等方式，指导试点相关工作。

四、试点内容

- 1. 建立数据资产登记机制(编制数据资产台账、开展数据资产登记、完善数据资产基础信息、开展数据资产入表)；
- 2. 推进数据资产授权运营(稳步推进数据资产授权运营、加强数据资产使用管理)；
- 3. 探索数据资产收益分配(探索数据资产收益分配机制、严格数据资产收益监管)；
- 4. 鼓励数据资产交易流通(推进数据资产交易流通、探索公共数据资产价格形成机制)；
- 5. 强化数据资产运行保障(加强数据资产制度建设、加强全流程动态监控)。



政策解读

2025.5.28——湖北省数据局印发《湖北省公共数据资源资产化改革方案》的通知

- 一、试点任务**
- 时间段：2025年4月至2026年12月；
 - 范围：有关省直部门、省属企业、部分地区财政部门；
 - 总体目标：围绕数据要素“贯通管用”，以数据清（理）、确（权）、入（账）、建（库）为重点，夯实数据资源资产化工作基础，充分释放数据要素价值，更好赋能高质量发展、高效能治理、高品质生活。
- 二、试点范围**
1. 市级（10个）：武汉市、襄阳市、黄石市、孝感市、荆州市、荆门市、随州市、鄂州市、咸宁市、十堰市。
 2. 县级（20个）：洪山区、江夏区、黄陂区、硚口区、武汉经济开发区、东湖高新区、樊城区、老河口市、夷陵区、点军区、枝江市、宜昌市、茅箭区、竹溪县、郧西县、房县、荆州区、通山县、英山县、大冶市。
 3. 省直部门（8个）：省发改委、省公安厅、省民政厅、省住建厅、省交通运输厅、省农业农村厅、省卫健委、省农业事业发展中心。
 4. 省属企业（7个）：湖北联投（湖北数据集团）、湖北交投、湖北铁路集团、湖北港口集团（湖北供应链物流公共信息服务有限公司）、湖北机场集团、湖北宏泰集团（湖北省融资再担保集团、华中文交所、武汉知识产权交易所）、湖北文旅集团。
- 三、主要任务**
1. 全面开展数据清查盘点。
 2. 稳妥开展数据确权确值。
 3. 扎实推进数据资源入卡入表。
 4. 夯实数据供给和管理基础。
 5. 推进数据确权。
 6. 推进数据共享流通。
 7. 推进数据资源开发利用。
 8. 加强数据资产化全过程管理。
 9. 统筹推进数据资源公共设施。
 10. 探索推出一批制度规范。

湖北省数据局文件

鄂数发〔2025〕39号

省数据局关于印发《湖北省公共数据资源资产化改革工作方案》的通知

各市、州、县数据主管部门，省直各单位，省属企业：

为贯彻落实《省人民政府办公厅关于印发〈湖北省深化国有“三资”管理改革 推动大财政体系建设走深走实总体工作方案〉的通知》（鄂政办发〔2025〕12号）要求，全面清查、有效盘活行政事业性单位和国有企业数据资源，加快形成数据资源资产化的有效路径，我们制定了《湖北省公共数据资源资产化改革工作方案》，经省人民政府同意，现印发给你们，请认真贯彻落实。



政策解读

2025.7.31——湖北省第十四届人民代表大会常务委员会第十九次会议通过《湖北省数据条例》

- 公共数据的定义**
- 公共数据，指本省国家机关、法律法规授权的具有管理公共事务职能的组织和供水、供电、供气、供热、交通运输、教育、医疗、文化旅游、体育、环境保护等公共企事业单位（以下统称公共管理和服务机构）在依法履行职责或者提供公共服务过程中收集、产生的数据，以及中央国家机关派驻本省的机关或者派出机构根据本省应用需求提供的数据。
- 数据权益保护**
- 第十四条 省人民政府数据主管部门按照国家规定统筹建设全省数据产权登记体系，推进数据产权登记工作。鼓励自然人、法人和非法人组织在依法设立的登记机构对数据持有、使用、经营等权益进行登记。经登记机构审查后取得的数据权益登记凭证，可以作为数据资产会计处理、数据企业认定、数据流通交易、融资担保等活动的证明。
- 数据流通**
- 第四十条 省人民政府数据主管部门应当会同有关部门推动建设省数据流通交易平台，健全数据流通交易平台体系，推进区域性、行业性数据流通交易平台互联互通。公共数据授权运营形成的数据产品或者服务、国有企业的产品或者服务以及财政资金保障运行的公共管理和服务机构采购数据的，应当通过省数据流通交易平台或者其他依法设立的数据交易所进行交易。鼓励市场主体通过数据流通交易平台或者其他依法设立的数据交易所进行交易。
- 生效时间**
- 本条例自2025年10月1日起施行。



解读《湖北省数据条例》新闻发布会



政策解读



2025.9.28——湖北省人民政府办公厅关于印发《〈湖北省数据条例〉贯彻实施工作方案》的通知

【七】构建数据产权登记体系。统筹建设全省数据产权登记体系，制定实施细则和相关指引，推动平台建设，引导企业开展数据产权登记，推动登记凭证作为数据资产会计处理、数据企业认定、数据流通交易、融资担保等活动的证明。（完成时间：持续实施；责任单位：省数据局、省地方金融管理局、省财政厅、湖北大数据集团等）

【十四】促进社会数据资源开发利用。实施国有企业数据赋能提升行动，深化国有企业数据资源开发利用，定期征集遴选发布湖北省高质量数据集与湖北省可信数据空间创新发展试点，充分发挥省数据流通交易平台作用，促进数据资源高效流通、融合利用，深化“数据要素×”行动，每年组织“数据要素×”创新大赛，发布典型案例，打造一批可感知的应用成果。（完成时间：持续实施；责任单位：省数据局、省有关部门、湖北大数据集团、各市、州、县人民政府）

【十五】健全数据交易服务体系。出台数据流通交易管理办法，加强对数据交易机构、第三方专业服务机构的管理。依托数据交易机构发布交易指引，推行数据流通交易合同示范文本，规范数据交易行为，提升中部数据流通服务中心能级，完善球务功能和市场监管体系，推动与省外其他数据交易市场互联互通，融入全国一体化数据市场。（完成时间：持续实施；责任单位：省数据局、湖北大数据集团）

【十六】建立价格形成和收益分配机制。运营机构基于授权派发的公共数据开发形成的数据产品和服务实行政府指导价管理，非公共数据产品由市场主体依法自主定价。充分发挥中部数据流通服务中心的价格发现和供需撮合作用，建立符合数据要素特性的价格形成机制和收益分配机制，探索公共数据资产付费使用，将相关非收益纳入预算管理。（完成时间：持续实施；责任单位：省数据局、省发改委、省财政厅、省政府国资委、湖北大数据集团、各市、州、县人民政府）

【十七】规范数据资产管理与数据采购。深入实施数据资产全过程管理试点和数据资源资产化改革，指导和规范专业服务机构开展数据资产评估业务，强化数据资产管理，规范公共管理和服务机构、国有企业通过省数据流通交易平台或者其他依法设立的数据交易市场采购数据产品和服务；严格审查公共管理和服务机构对非公共数据的采购需求，实行统筹采购、集约共享。（完成时间：持续实施；责任单位：省财政厅、省数据局、省政府国资委、湖北大数据集团、各市、州、县人民政府）

【十九】培育数据要素市场主体。建立数据企业认定机制，每年开展1-2批次数据企业登记入库，加强数据资源招商，完善数据企业公共服务体系，引导金融机构开发“数据贷”等产品，培育壮大数据企业。依托数据交易所，引进优质数据商和第三方专业服务机构。（完成时间：持续实施；责任单位：省数据局、省经信厅、湖北大数据集团、各市、州、县人民政府）

湖北省人民政府办公厅关于印发《〈湖北省数据条例〉贯彻实施工作方案》的通知

湖北省人民政府网 2025年10月09日 15:22 湖北

湖北省人民政府办公厅关于印发
《〈湖北省数据条例〉贯彻实施
工作方案》的通知

鄂政办函〔2025〕13号

各市、州、县人民政府，省政府各部门：

《〈湖北省数据条例〉贯彻实施工作方案》已经省政府同意，现印发你们，请认真贯彻执行。

2025年9月28日

(此件公开发布)





3.1 认识数据要素

数据为什么是石油？

《数据领域名词解释》



数据

是指任何以电子或其他方式对信息的记录。数据在不同视角下被称为原始数据、衍生数据、数据资源、数据产品和服务、数据资产、数据要素等。

数据资产

是指特定主体合法拥有或者控制的，能进行货币计量的，且能带来经济利益或社会效益的数据资源。通常含数据产品，具有可确权、可收益等特征。



数据流通

是指数据在不同主体之间流动的过程，包括数据开放、共享、交易、交换等。

数据交易

是指数据供方和需方之间进行的，以特定形态数据为标的，以货币或者其他等价物作为对价的交易行为。国家淡化所有权，推行持有权、经营权、使用权“三权”分置，现阶段数据交易更多的是使用权的授权使用，并非所有权转移。



数据为什么是石油？



数据为什么会成为新型生产要素？

数据能成为生产要素是“时代在召唤”，是人类社会演进至数字时代的必然产物，谁能掌控数据、谁就能掌控未来无限可能。

农业时代

- 以土地为根基，收成依赖土地肥力与气候条件。
- 劳动效率低下，经济呈自给自足模式。

信息时代

- 以技术为驱动，计算机普及与互联网兴起，信息成为战略资源，信息时代是软件的时代。
- 数据仅是“物理世界的投影”，服务于既有流程。例如，如ERP系统仅记录业务结果，不改变业务模式。

1

2

3

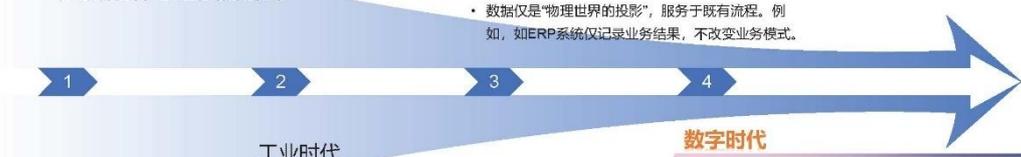
4

工业时代

- 机械化、电气化取代手工作坊，大规模机械化生产高度依赖巨额资本投入。
- 技术专利体系逐渐成熟，技术独立为生产要素。
- 劳动力出现专业化分工。

数字时代

- 日益强烈的智能化需求促使问题处理从被动响应升级为主动预测与创新，催动数据从“被动记录者”到“主动创造者”的质变。例如，智能网联汽车需融合路况、用户等数据，才能支撑其自动调整行驶策略。
- 推动决策逻辑从“人脑+有限信息”转向“算法+海量数据”，将人类从繁杂琐碎中解放出来。



数据为什么是石油？



**与其他传统生产要素不同
数据具有三大特性**

非竞争性

- 无限共享与零边际成本复制

同一份数据可被多主体同时、重复、无损使用（如气象数据可同时服务农业、物流、保险等领域的不同主体），且复制传输成本趋近于零，打破了传统要素因稀缺性导致的“排他竞争”模式，催生开放共享经济范式。



伴生性

- 数据是唯一非主动投入却可规模化增值的要素

传统生产要素需刻意投入才能参与生产，数据是自动伴随主体活动产生，且一旦产生，可低成本存储和无限复用。

数据在产生时未被预知其用途，价值在后期主动挖掘使用场景中显现。

价值动态性

- 数据价值高度依赖应用场景和分析能力

同一数据在不同场景下价值不尽相同，需通过“算法炼金术”提炼释放。（如用户浏览记录原始值≈0，经AI建模后成为千亿级广告收入基础）

数据价值在产生初期最高，随后随时间推移呈指数级衰减。（如网约车轨迹数据在实时调度场景中价值最高，但若延迟1小时，仅剩历史规律分析价值）

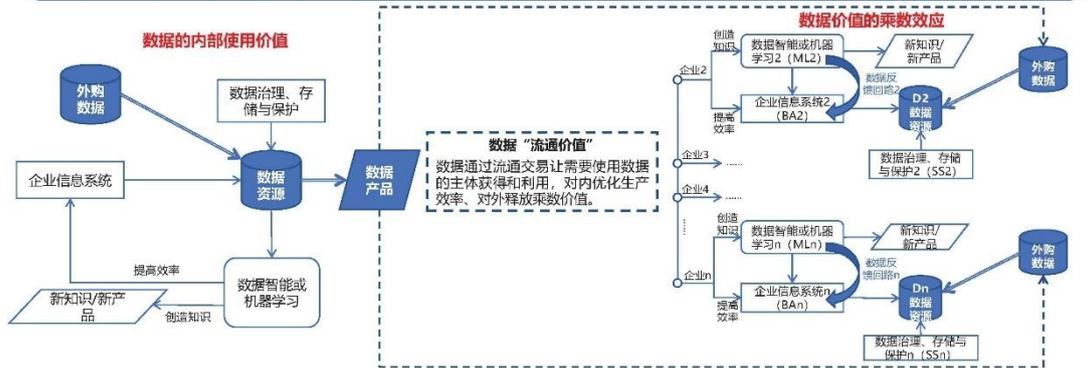
数据为什么是石油？

首席数据官

数据是新时代的“石油”

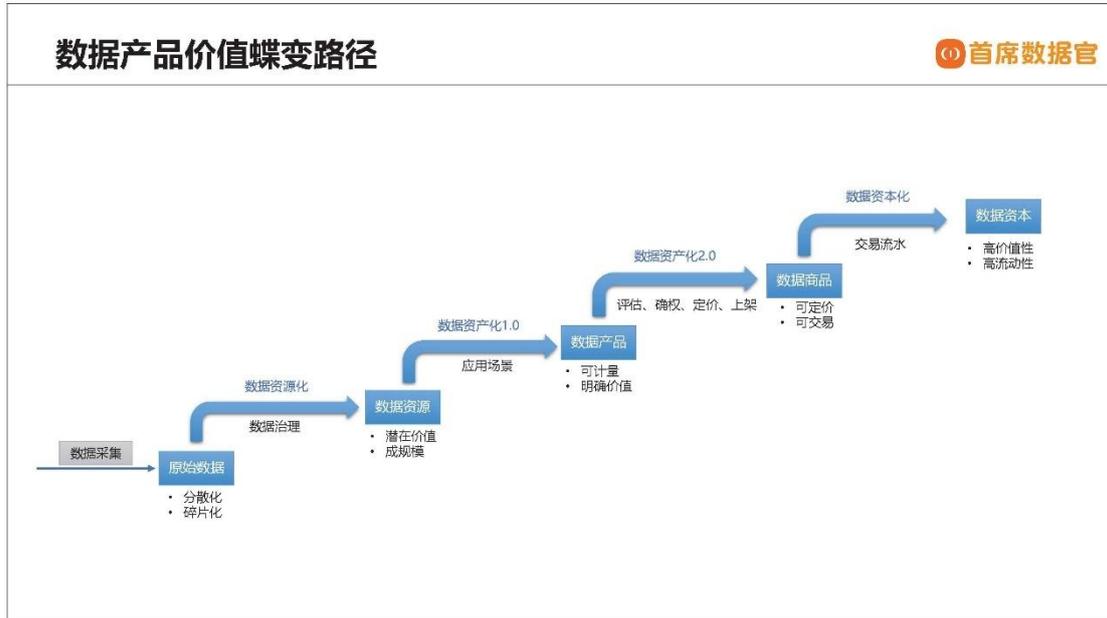
数据是数字融合时代的“燃料”，既是企业内部经营决策的关键要素，也是通过流通利用提高全社会资源配置效率，培育新产业新业态，释放对经济发展的乘数效应。

数据是新时代的“石油”，但却不止于此。石油会枯竭，**数据源源不断**，谁能掌控数据，谁就能像掌握可控核聚变一样，拥有未来的无限可能。



首席数据官

3.2 认识数据产品



04

数据资源资产化改革实践路径



首席数据官

4.1

数据资产入表

为什么要数据资产入表?

首席数据官

✔ 什么是“数据入表”? 一句话讲清!

◆ 通俗说:

“数据入表”就是把企业手里的数据变成可以在财务报表中体现的“资产”或“资源”，成为有账面价值的东西。

过去，数据是“看得见用得着，但不算账”的虚拟资产；

现在，**国家政策+资本市场+会计制度正在推动：让数据成为“资产负债表”上的一项正式内容。**

为什么要数据资产入表?

首席数据官

🚩 为什么说“数据不入表”，企业损失巨大?

✘ 1. 无法估值

数据再多，也不能“入账”，投资人无法评估你的真实价值，融资难!

✘ 2. 无法交易

很多地方政府和数据交易所都要求“确权+入表”，没有这一步，数据不能挂牌交易。

✘ 3. 无法入资

你想用数据作为资产出资、技术入股? 对不起，**没有“表格凭证”，工商系统不认可。**

✘ 4. 无法审计

IPO审核、年终审计时，数据资产无形、无凭证、无标准，无法作为合规资产核查。

👉 总结一句：**没有“数据入表”，就等于你白拥有了这些数据。**

为什么要数据资产入表?

首席数据官

🔥 数据入不了表，就只是“信息垃圾堆”!

现在不是有没有数据的问题，而是：

- ✅ 你的数据有没有“身份证”? (确权)
- ✅ 能不能被“合法估值”? (入表)
- ✅ 能不能对“外部变现”? (交易)

数据资产化三部曲：确权 → 入表 → 交易
其中“入表”就是价值释放的“临门一脚”。

📌 政策信号已明确：“数据资产入表”已成趋势!

- 财政部：正加快推进数据资产会计核算标准研究
- 国资委：推动央企开展“数据资产价值评估”试点
- 多地数据交易所：入表是挂牌、融资的“入场券”
- **2024年初，首批数据资产入表试点企业名单公布**
- 👉 意味着：**谁先让数据“上账”，谁就更有资格“变现”!**

数据资产入表的微观意义

首席数据官

依据：《企业数据资源相关会计处理暂行规定》

适用范围：第一条 适用于企业按照企业会计准则相关规定确认为无形资产或存货等资产类别的数据资源，以及企业合法拥有或控制的、预期会给企业带来经济利益的但由于不满足企业会计准则相关资产确认条件而未确认为资产的数据资源的相关会计处理。

微观意义和作用

有利于**显化数据资源价值，提升企业数据资产意识**，激活数据市场供需主体的积极性，增强数据流通意愿，减少“死数据”，为企业对数据进行深度开发利用提供动力。

有助于有效反映企业在**数据要素方面的投入和收益**，**增厚企业资产**，满足融资需求，并辅助国家加强对数据要素市场的统筹管理。

能够有效带动**数据采集、清洗、标注、评价、资产评估等数据服务业发展**，激发数字经济发展活力。

明确了数据作为资产的价值。目前拥有数据的企业可以盘活存量数据的价值，数据所有者**增厚资产负债表**。同时，**数据资源入表**将刺激企业进一步挖掘数据的价值，而数据加工是为数据增值的直接方式，看好数据加工服务商的市场定位。

政策依据

会计术语
存货

数据资源
入表

会计术语
无形资产



存货

指企业在日常活动中持有以备出售的产成品或商品、处在生产过程中的在产品、在生产过程或提供劳务过程中耗用的材料和物料等。

《企业会计准则第1号-存货》第三条

无形资产

指企业拥有或者控制的没有实物形态的可辨认非货币性资产。资产满足下列条件之一的，符合无形资产定义中的可辨认性标准：(一)能够从企业中分离或者划分出来，并能单独或者与相关合同、资产或负债一起，用于出售、转移、授予许可、租赁或者交换。(二)源自合同性权利或其他法定权利，无论这些权利是否可以从企业或其他权利和义务中转移或者分离。

《企业会计准则第6号-无形资产》第三条

数据资产入表的宏观意义

首席数据官

充分释放企业数据价值

- 借助数据推动智能化工作流程，优化企业生产模式，提升企业运营效率、推动企业降本增效。
- 使得企业数据资产的业务价值、经济价值和社会价值显性化，将数据资产转化为企业资产的一部分，通过与金融体系接轨进一步推进数据价值实现。

促进产业链升级转型

- 有助于重组传统产业结构、颠覆现有业务模式，催生新产业、新业态、新模式。
- 提升产业基础能力和产业链现代化水平提出了更高要求，同时也为产业发展带来更广阔的空间。



推动企业数智化转型

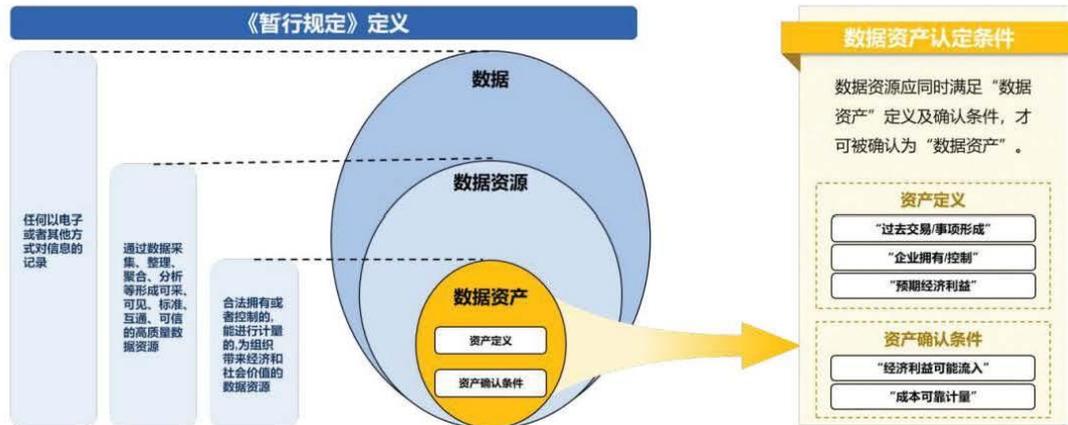
- 激活市场潜能，催生丰富的数据应用场景，提高企业开展面向数据生产与创新的积极性。
- 推进生态建立，推动服务型企业充分发挥各自独特的价值，促使企业深化数智化转型，实现数智化思维与设计、生产、管理、服务和运营等全过程深度融合，利用数智化方式重塑企业发展模式和竞争优势。

实现经济发展倍增效应

- 将通过从数据中挖掘有用信息，作用于其他要素，找到企业、行业、产业在要素资源约束下的“最优解”，进一步提升全要素生产率。
- 有效带动数据采集、清洗、标注、评价、资产评估等数据服务业发展，将进一步深化数字技术创新应用，激发数字经济发展活力，营造繁荣发展的数字生态。

什么数据可以入表？

首席数据官



合规评估要点?

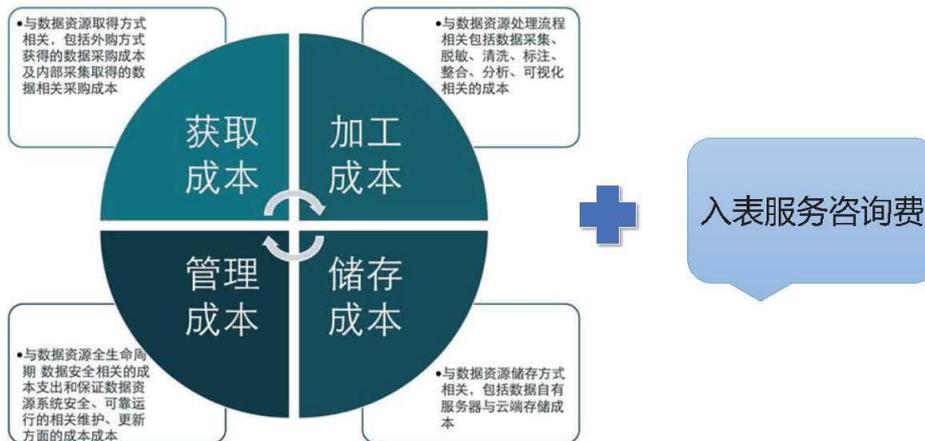
系统性数据合规评估

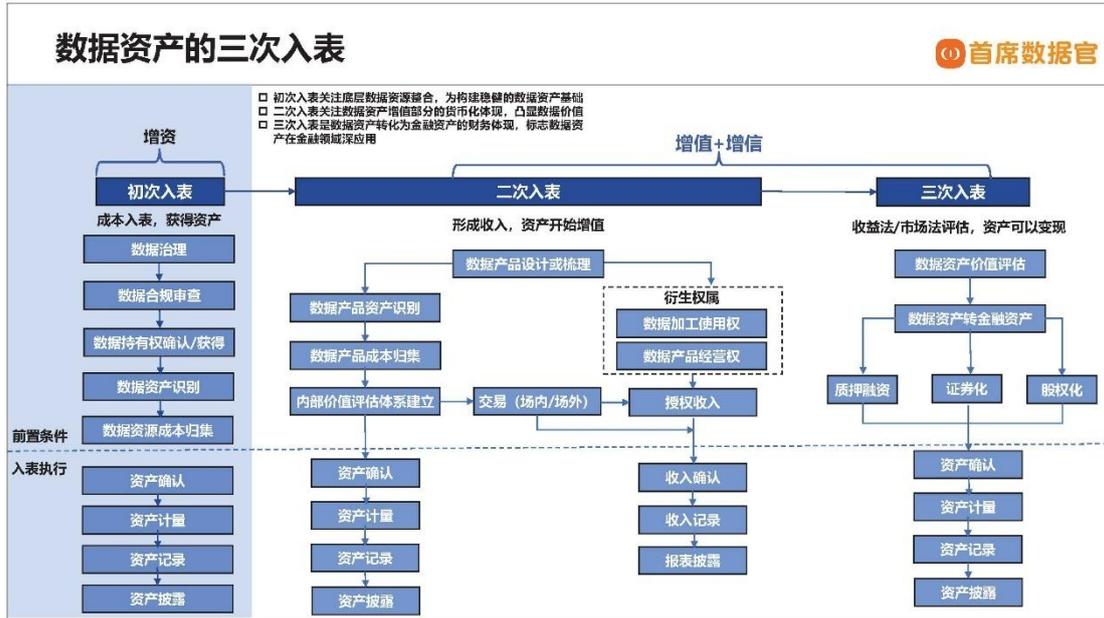
指在数据处理过程中，遵守相关国家法律法规、标准、政策、最佳实践、国际条约和规则、公司章程和规章制度，从数据主体、数据管理、数据来源、数据处理、数据内容、数据交易等6个方面主要维度梳理，建立企业数据合规管理机制，确保数据资源的合法、合规。

 主体合规	工商信息，经营异常，数据安全、个人信息保护行政处罚、涉诉信息，关键信息基础设施运营者，商誉、重大舆情、重大财务风险、可能影响持续经营的重大事项	数据合规体系法律要求 《数据安全法》 《数据安全法》第一条、第四条、第七天、第十八条和第十九条。明确要建立数据合规体系，加强培训、人才建设和合规认证。 《网络安全法》 《个人信息保护法》 对个人信息保护负责人、个人信息内部管理制度和操作规范、个人信息定期合规审计、个人信息保护影响评估、个人信息分类管理、个人信息加密等进行规定。 <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="background-color: white; color: #0056b3; padding: 5px; border-radius: 5px;">第十二条</div> <div style="background-color: white; color: #0056b3; padding: 5px; border-radius: 5px;">第五十一条</div> <div style="background-color: white; color: #0056b3; padding: 5px; border-radius: 5px;">第五十四条</div> <div style="background-color: white; color: #0056b3; padding: 5px; border-radius: 5px;">第五十五条</div> <div style="background-color: white; color: #0056b3; padding: 5px; border-radius: 5px;">第五十八条</div> </div>
 数据管理	需按照法律法规、规章等要求，建立数据合规相关管理制度，开展包括合规管理体系搭建、风险评估与处置等管理活动，对数据分级分类管理、数据跨境等领域建立相应的全链条监督管理机制	
 数据来源	获取数据行为不违反任何法律法规、国家政策和社会主义公德，不侵犯任何第三方合法权益，自行生产的合法性，相关协议的充分性和合法性	
 数据处理	处理数据行为不违反相关法律法规规定，符合合法、正当、必要原则	
 数据内容	数据的内容需真实、合法、合规，不违反法律、法规强制性规定，不危害国家安全、公共利益和第三方合法权益，符合社会主义公共秩序和公序良俗，数据性质视为公共、个人敏感、重要数据	
 数据交易	数据产品具有合法性、可控性、流通性，投入实质性加工和创新性劳动，数据产品应用场景与使用条件、限制，数据交易协议内容是否合格	

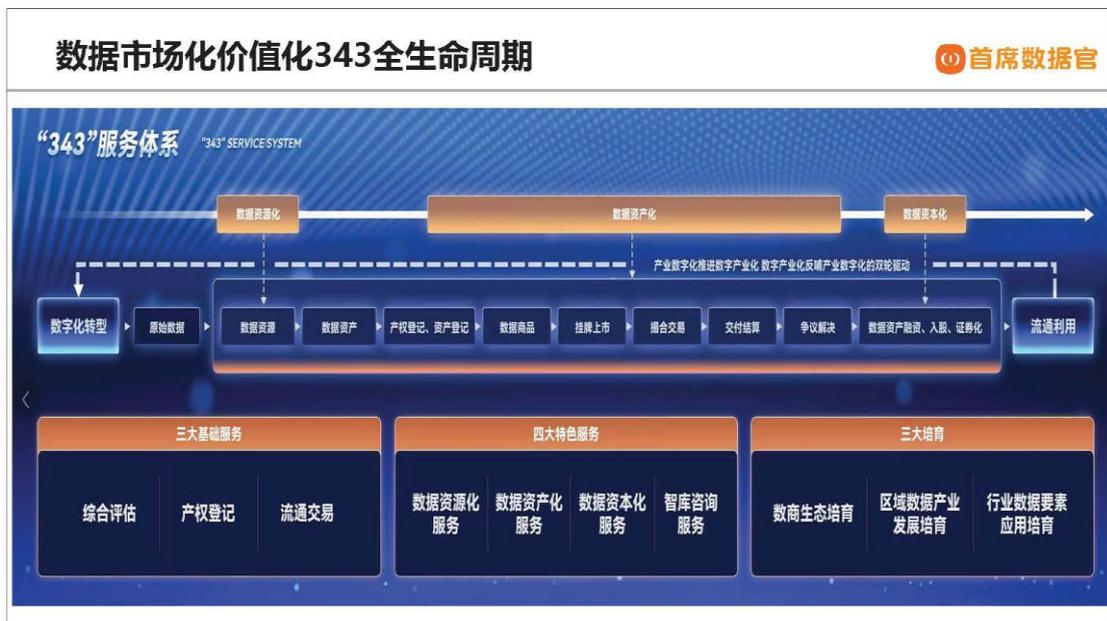
入表金额（成本）如何构成？

【成本归集与分摊】数据资源成本组成





4.2 数据流通交易

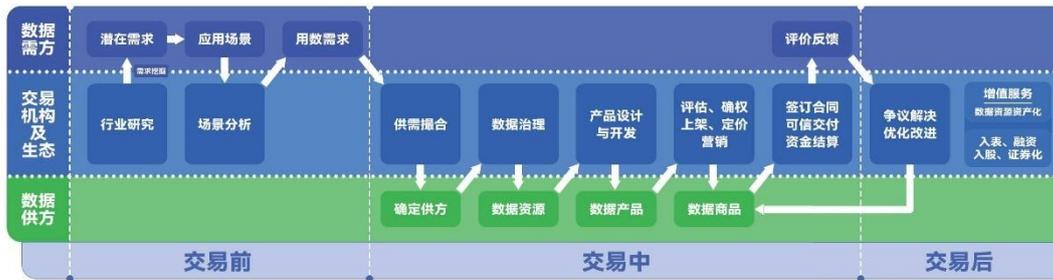


以“需求”为导向的数据交易全过程

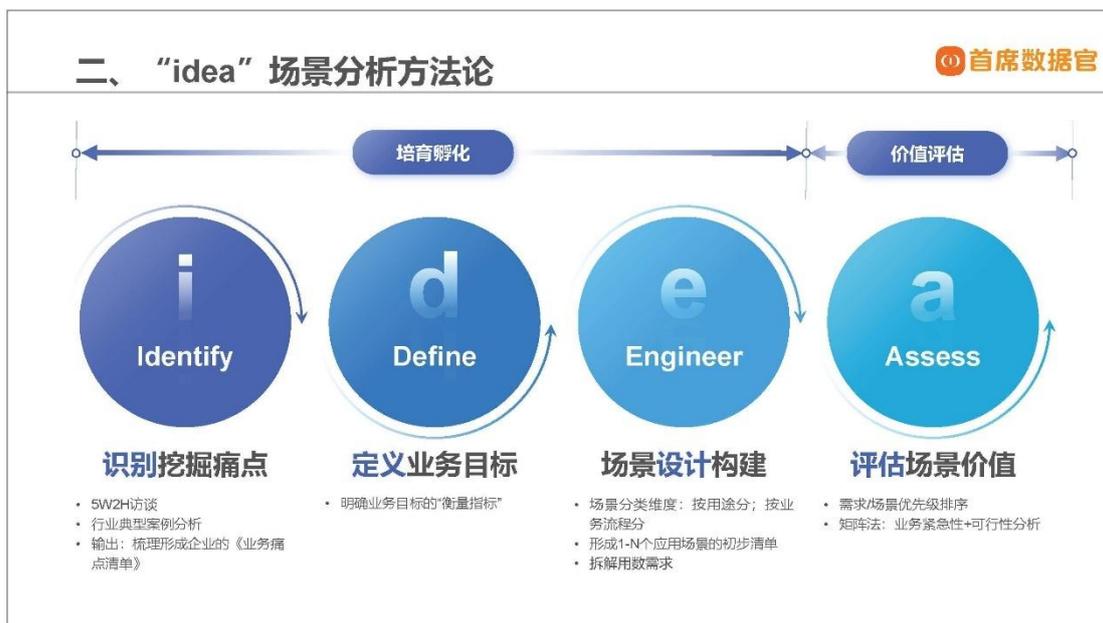
首席数据官

以需求为导向的数据交易全过程

数据交易的实质不是资产转移，而是使用付费，类似专利使用付费



数据场景分析 “idea” 方法论



“idea” 识别挖掘痛点

首席数据官

梳理形成企业的《业务痛点清单》

理解业务痛点

- 企业访谈：
 - 优先筛选行业领域内有一定影响力的头部企业开展调研；
 - 5W2H法开展提问调研，梳理需求方的《业务痛点清单》；
- 行业分析：针对金融/商贸/医疗/文旅/低碳等典型行业，梳理分析其已落地的应用场景案例，提炼“场景类型、用数需求”等关键信息。

精准定义问题

- 清晰的问题定义陈述：包含“业务环节、衡量指标、业务问题”等关键因素；
- 例如：市场部在广告投放优化时，因为广告数据与用户交易数据割裂，导致无法评估渠道的长期用户价值（LTV），只能优化短期点击成本，使得广告预算被大量低价值用户消耗，整体ROI偏低。

业务痛点清单

序号	访谈对象	问题描述			
		业务环节	衡量指标	业务影响	详细描述
1					
2					

- What: 要解决什么核心问题?
- Why: 这个问题是否重要且紧急?
- Who: 这个问题会影响哪些人? (管理层、客户、业务员)
- Where: 问题发生在业务流程的哪个环节?
- When: 决策/行动发生的频率?
- How: 现在是怎么处理的? 效果如何?
- How Much: 问题的影响多大? 期望的改善程度是什么?

“idea” 识别挖掘痛点

首席数据官

梳理形成企业的《业务痛点清单》

理解业务痛点

- 企业访谈：
 - 优先筛选行业领域内有一定影响力的头部企业开展调研；
 - 5W2H法开展提问调研，梳理需求方的《业务痛点清单》；
- 行业分析：针对金融/商贸/医疗/文旅/低碳等典型行业，梳理分析其已落地的应用场景案例，提炼“场景类型、用数需求”等关键信息。

精准定义问题

- 清晰的问题定义陈述：包含“业务环节、衡量指标、业务问题”等关键因素；
- 例如：市场部在广告投放优化时，因为广告数据与用户交易数据割裂，导致无法评估渠道的长期用户价值（LTV），只能优化短期点击成本，使得广告预算被大量低价值用户消耗，整体ROI偏低。

业务痛点清单

序号	访谈对象	问题描述			
		业务环节	衡量指标	业务影响	详细描述
1					
2					

- What: 要解决什么核心问题?
- Why: 这个问题是否重要且紧急?
- Who: 这个问题会影响哪些人? (管理层、客户、业务员)
- Where: 问题发生在业务流程的哪个环节?
- When: 决策/行动发生的频率?
- How: 现在是怎么处理的? 效果如何?
- How Much: 问题的影响多大? 期望的改善程度是什么?

① 首席数据官

“idea” 定义业务目标

明确业务目标的“衡量指标”

问题 >>> 市场部在**广告投放优化**时，因为广告数据与用户交易数据割裂，导致无法评估**长期用户价值 (LTV)**，只能优化短期点击成本，使得广告预算被大量低价值用户消耗，整体**ROI偏低**。



S

具体的

提升基于长期用户价值的广告ROI



M

可衡量

将“渠道长期用户价值ROI” (LTV/投放成本) 从日前的1.5提升到2.0



A

可实现

通过打通广告平台API和内部用户数据库来实现



R

相关联

直接关联市场部的核心KPI (ROI) 和公司的盈利目标



T

有时限

在3个月内为实现

目标 >>> 在**3个月内**，通过构建**广告渠道长期价值评估系统**，将“**渠道长期用户价值ROI**” (LTV/成本) 从**1.5提升到2.0**，从而解决广告预算因无法识别高价值用户而被**低效消耗**的问题。

① 首席数据官

“idea” 场景设计构建

形成《“场景+数据”清单》

1

分类进行场景设计

- 按输出形式分类：
 - 描述型**（发生了什么？）：**报表、仪表盘、自动化报告**（例如：为营销团队建立多渠道投放效果实时监控大屏）。
 - 诊断型**（为什么发生？）：**归因分析、下钻分析、路径分析**（例如：分析高LTV用户的来源渠道和行为路径）。
 - 预测型**（将会发生什么？）：**预测模型、预警系统**（例如：预测未来一个月哪些新用户可能成为高价值用户）。
 - 策略型**（我该怎么做？）：**推荐系统、优化策略、智能风控、动态定价**（例如：自动为高潜力渠道分配更多预算）。
- 按业务阶段分类：
 - 电商行业：获客阶段-转化阶段-留存阶段-变现阶段
 - 制造业：生产阶段-库存阶段-售后阶段
 -

2

初步列1-N个场景清单

- 行业领域+输出形式+目标成效，例如：
 - 电力数据助力经济分析，提升“即时”预测分析实效
 - “杭小忆”：数据赋能伴随型旅游新体验
 -

3

拆解用数需求

- 数据需求、数据源、数据质量：
 - 例如：某房地产开发企业需要有土拍市场风险预警提示，来优化项目决策——>土地成交面积、单价、类型、周边配套设施情况等数据.....
- 数据需求是否涉及多源数据：比如不同行业数据，抑或是行业数据叠加公共数据，需要做整合建设

“idea” 场景设计构建

《“场景+数据”清单》示例

60%

数据来源清晰

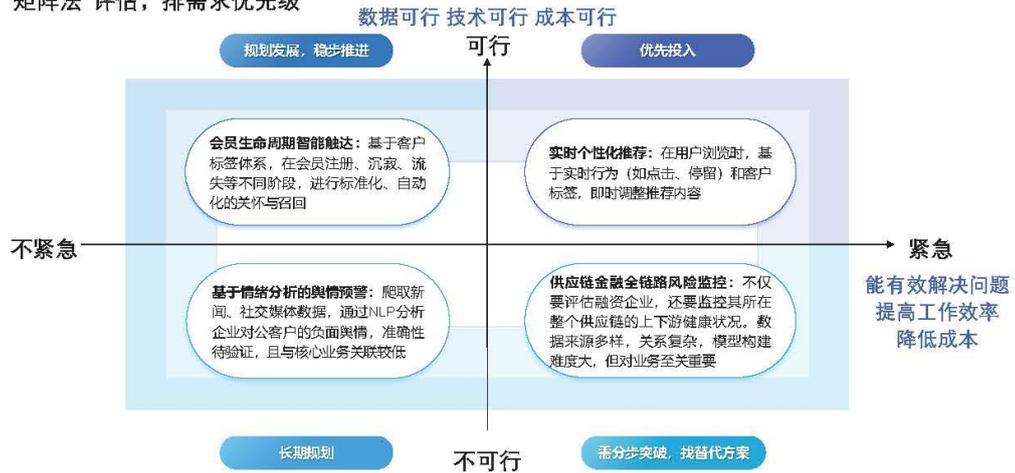
40%

数据来源不明

序号	行业领域	场景类型	输出形式	目标成效	数据需求	
					公共数据	社会数据
示例	城市建设	预测型	预警系统	决策效率提升3% 投资风险降低5%	<ul style="list-style-type: none"> 宏观经济数据：GDP、人均可支配收入； 房地产市场数据：新房/二手房成交量、去化周期、土地成交楼面价与溢价率 	<ul style="list-style-type: none"> 地块详情数据：地块规划条件、周边配套数据、竞品项目数据、竞品销售动态数据、客户舆情数据
1						
2						

“idea” 评估场景价值

“矩阵法”评估，排需求优先级



模拟案例解析
 首席数据官

某智能制造企业的“生产良率提升”应用场景培育

<div style="background-color: #004a99; color: white; padding: 2px 5px; margin-bottom: 5px;">01 i: 识别挖掘痛点</div> <p>当前生产质量的管控依赖于事后分析，无法在生产过程中实时发现异常并预警，导致良率92%（低于行业领先水平的96%），且返工成本高昂。</p>	<div style="background-color: #004a99; color: white; padding: 2px 5px; margin-bottom: 5px;">02 d: 定义业务目标</div> <p>通过实时监控关键工艺参数，异常数据及时预警，将整体生产良率从92%提升至95%。</p>
<div style="background-color: #004a99; color: white; padding: 2px 5px; margin-bottom: 5px;">03 e: 场景构建设计</div> <p>场景一：关键工艺参数实时监控与告警系统 在关键生产设备上部署传感器，实时采集温度、压力、转速等工艺参数，并设置控制上下限。一旦参数超限，系统立即在车间看板和中控室发出声光警报。</p> <p>场景二：基于实时参数的质量预测与自调节系统 在监控系统的基础上，构建机器学习模型。根据工序的实时参数，预测该零件最终成为不良品的概率。如果概率超过阈值，系统可自动通知设备微调参数，或标记该零件以待检。</p>	<div style="background-color: #004a99; color: white; padding: 2px 5px; margin-bottom: 5px;">04 a: 评估场景价值</div> <ul style="list-style-type: none"> • 业务紧急性：高：解决的是“有无”问题，从依赖人工巡检到自动监控；并且能立即阻止批量性质量事故的发生，避免巨大损失； • 可行性：中：需要部署传感器；能直接解决一线工人的痛点，接受度高。 • 业务紧急性：中：解决的是“好不好”问题，研发周期长； • 可行性：低：依赖场景一积累的高质量数据，初期数据样本量不足；可能改变现有生产决策流程，需要时间适应和信任。



一、数据产品概述与类型

首席数据官

根据国家数据局发布的《数据领域名词解释》，数据产品和服务定义为“基于数据加工形成的，可满足特定需求的数据加工品和数据服务”。

数据集产品

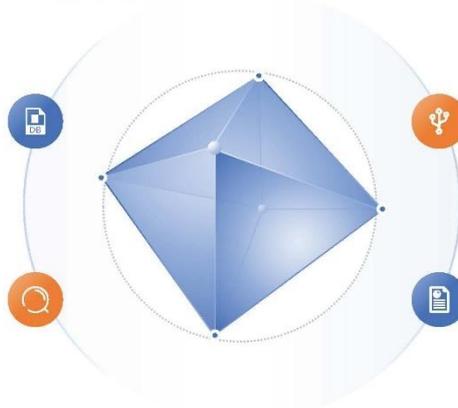
指经采集、收集后，汇总、整理及加工后的**数据资源集合**，可作为分析或模型训练的“原材料”。

典型场景：人工智能行业使用数据集尤其是高质量数据集产品进行训练。

数据应用及模型产品

封装好的**软件、工具或算法模型**，能直接用于解决特定问题，属于高附加值产品。

典型场景：金融领域根据用户风险偏好自动推荐投资组合的App，工业制造等领域的数据分析与BI工具，支付宝的实时交易反欺诈模型。



数据服务产品

通过API接口、分析报告等形式提供的**数据调用或咨询服务**。

典型场景：金融行业的征信数据API服务、企业信用信息公示查询服务，咨询行业提供《水上交通安全要素定向数据分析报告》等各种分析报告。

其他多元化产品

算力资源（算力形成过程中涉及的计算资源，含云存储、云安全及衍生服务）、**综合解决方案**等复合型产品。

典型场景：人工智能行业的算力资源产品，智慧城市如智慧政务领域提供从数据中台建设到“一网通办”应用的全套解决方案。

二、数据产品设计和开发核心原则

首席数据官

坚持四大原则

01

安全合规

合规先行

将**数据合规与安全评估**贯穿于产品生命全周期，从源头确保数据来源合法、处理得当、授权清晰



02

需求分析

价值导向

产品设计应**紧扣市场需求**，**解决实际业务痛点**，确保数据产品具备明确的市场价值与竞争力。



03

紧密配合

协同合作

鼓励**跨职能团队（业务、合规、技术）紧密协作**，采用敏捷迭代方式，快速响应市场变化。



04

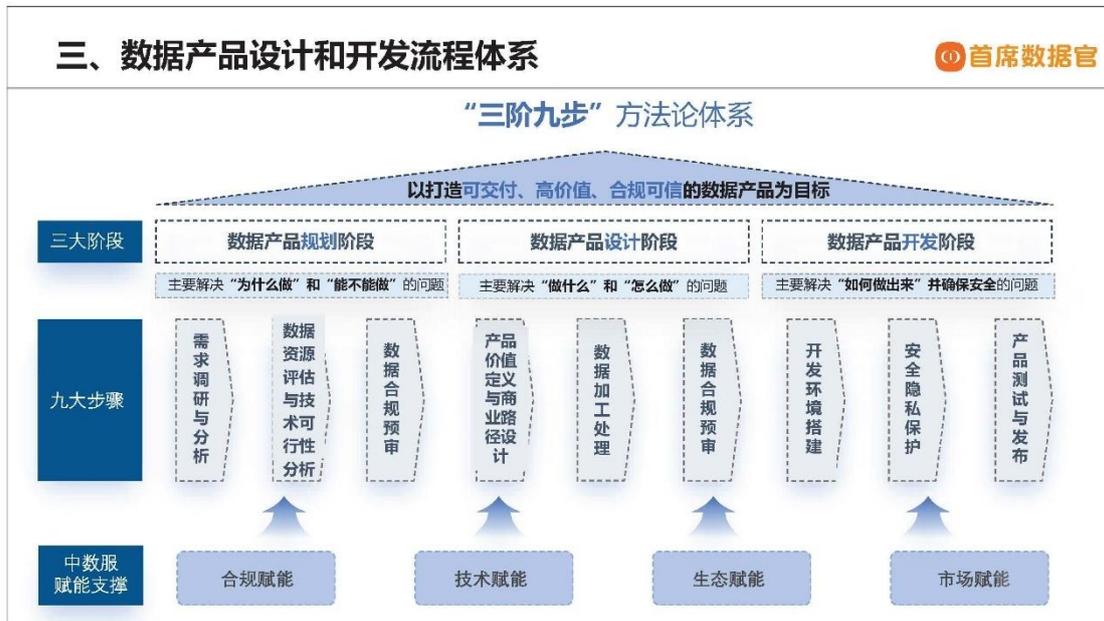
共创共享

生态共赢

依托中部数据流通服务中心提供的合规、技术、营销等**公共服务**，**降低开发门槛**，与**生态伙伴**共创共享数据价值。



三、数据产品设计和开发流程体系



第一阶段：数据产品规划阶段



第二阶段：数据产品设计阶段

首席数据官

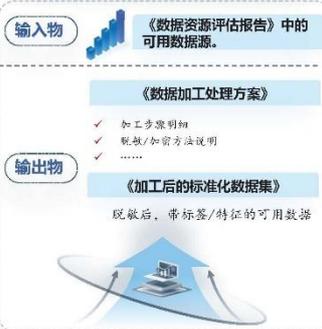
04：产品价值定义与商业路径设计

明确产品的业务定位、目标用户、核心价值主张，选择合适的定价模式，确定使用权限、付费流程，验证商业可行性。



05：数据加工处理

数据开发方从自有库、合作方、公共平台获取原始数据，开展数据采集、清洗与集成、数据脱敏与匿名化、特征工程与标签化等操作，将原始数据转化为安全、可靠、有价值的数据，实现产品增值。



06：产品形态与交付标准确定

围绕场景匹配分析（用户怎么用产品）、技术支撑检查（确认技术能不能做这个形态）等方面确认产品最终形态，基于产品形态设计交付关键指标标准，并把指标反馈给用户，说明标准落地细则。



第三阶段：数据产品开发阶段

首席数据官

07：开发环境搭建

根据产品需求选择隐私计算、数据沙箱等开发环境，或用自有服务器搭建环境，开展环境配置与测试，记录环境文档。



08：安全隐私保护

构建安全防护体系，对静态数据和传输中的数据进行全面加密，实施访问控制与强身份认证，记录所有数据访问和操作日志，并上传至中央数据流通服务中心区块链存证平台，确保操作可追溯、不可篡改。



09：产品测试与发布

开展数据产品多维度测试，确保产品符合需求和安全合规。对离中心数据流通服务中心的数据产品发布审核清单，提交数据产品审核资料，进行内部审查，通过审核后发布。





四、模拟案例-某银行对公信贷全流程智能风控数据服务产品 首席数据官

-----> 产品开发：对公信贷全流程智能风控数据服务产品落地 ----->

• 开发环境搭建
• 安全隐私保护
• 产品测试与发布

三级环境

开发环境
自由服务器

测试环境
数据沙箱

生产环境
隐私计算节点

融合联邦学习与可信执行环境

加密防护 访问控制

日志存证 安全测试

结论：符合金融数据安全等级保护三级标准，风险防控覆盖率 100%

- **多维度测试**：功能测试（12 个 API 接口全部达标）、性能测试（并发 1000 用户响应时间 180ms）、安全测试（无数据泄露风险）
- **内部审查**：对照中部数据流通服务中心审核清单，完成材料核验
- **发布**：提交审核资料（测试报告、安全评估等），通过后在中心平台上线

100%

《数据产品测试报告》：测试通过率 100%，性能指标优于交付标准

《数据产品说明书》：API 调用文档、终端操作指南、常见问题解答等

五、中部数据流通服务中心的赋能与支持 首席数据官

合规赋能

提供前置咨询、合规评审、合规培训等服务。

技术赋能

提供开发工具包、沙箱环境、隐私计算平台等技术支持。

生态赋能

链接数据经纪、法律、评估、审计等第三方专业服务机构。

市场赋能

推出品牌活动、供需对接会、渠道推广等方式。



参考案例

首席数据官

1. 数据产权确认

- 省住建厅：“湖北省房地产业监管数据”等三个高质量数据集数据产权登记
- 省大数据中心：省直行政单位首个数据资源“三确”工作
- 十堰茅箭区数据局：全国首个行政单位数据资源“三确”案例
- 协和医院：“公立医院财务制度数据集”数据产权登记
- 麻城城市发展运营公司：“停车场充电桩数据”数据产权登记
- 武汉地脉集团：“武汉轨道交通岩土工程地质数据”数据产权登记
- 同济医院：“心血管真实世界研究数据集”、“胃癌多模态数据集”数据产权登记
- 咸宁金投：公积金数据产权登记
- 东湖高新区、江夏区、麻城市开展全区数据资源盘点及数据产权登记

3. 数据交易

- 武汉燃气公司与普品盟公司：燃气数据服务数据交易
- 华中电力公司与工商银行：“电力供应合同融资数据服务”数据交易
- 武汉市汉阳控股集团与四川人才供应链公司：“四川人才岗位数据分析报告”交易
- 光谷建设集团与成都万国公司：全国首例“城市地下信息管网数据”数据资产入表及交易
- 宜昌城市大脑公司与中国移动宜昌分公司：湖北首笔CIM数据产品场内交易
- 武汉测绘院与武汉双碳产业公司：全国首笔测绘地理时空数据暨全国首笔碳计量数据场内交易

2. 数据资产入表

- 湖北港口：省属国企首单“多式联运运费数据”数据资产入表
- 湖北交投：“高速车流分析数据产品-路段拥堵（风险）分布”和“试验检测类-工地实验室数据集”数据资产入表
- 湖北联投：“工程现场管理风险预警数据集”“U-Insight青年安居智选”数据资产入表
- 光谷交通：全省首个“自动驾驶数据”数据资产入表
- 孝感市思创公司：“基于动态计算的人类蛋白质科学高质量数据集V1.0”数据资产入表
- 九州通医疗供应链公司：“九州医药链数字化赋能产业服务”数据资产入表
- 咸宁城投集团：“咸宁智慧停车动态数据集”数据资产入表

4. 数据资本化

- 中交长江河口大桥公司“高速公路+数据资产”ABS项目
- 十堰鑫业公司“全国商用车行业车型数据”数据资产入表项目
- 湖北交投集团4个数据资产获兴业银行2个亿授信
- 岚风创投：“停车分布模型”获光大银行岚风分行获得了高达1000万元的融资
- 洪山科技公司“楼宇智慧运营数据”获得银行600万元授信
- 首义科创公司“首义智慧房管洞察分析”获得兴业银行500万元授信
- 家家养的车公司“全渠道智能零售生态数据应用”获得武汉农商行300万授信

参考案例
 首席数据官

| **典型数据企业剖析**

美国帕兰蒂尔(Palantir)

基本情况

2003年成立，**全球数据领域市值最大公司**，全球大数据和AI分析巨头，股价两年翻五倍，市值超3000亿美元，2025年二季度营收10亿，同比增48%。

典型产品

Gotham (哥谭)：综合多源信息，面向政府和国防部门，提供军事和情感分析服务；
Foundry (锻造厂)：综合分析大型企业分散数据，发现异常现象如财务欺诈，供应链效率低下等；
Apollo (阿波罗)：Gotham和Foundry的部署与数据集成平台，适用于高安全环境；
AIP：人工智能平台，将大语言模型集成到Palantir的私有网络技术优势：**数据本体论**——定义业务对象与关系，语义建模。动态数据表示，跨部门数据共享；
 商业优势：**高度定制化的服务模式——极高的客户粘性**
 数据优势：**跨行业、跨领域的数据经验**

核心优势

中国民航信息网络股份有限公司(中航信)

2000年成立，是**国内航空旅游业信息化龙头**，凭完备数据生态主导国内市场，为行业信息化关键支撑；国际上参与航空数据标准制定，与全球主流服务商合作竞争。

航旅纵横：基于互联网向旅客提供行程管理、机票搜索、航班动态等服务，用户规模超过1亿；
航班控制系统 (ICS)：提供航班计划，座位管理和库存控制等功能
计算机分销系统 (CRS)：提供航班信息查询和预定服务，连接全球400多家可销售航空公司；
机场旅客处理系统 (APP)：持值机、登机、行李处理等机场运营流程，已覆盖国内257家机场（占比达99%）和157家海外机场

数据优势：中国民航业的事实**数据枢纽**，整合了中国民航所有的核心业务数据，包括航班计划、旅客信息、运营数据和结算数据等；
技术优势：行业深度集成与高可靠性

参考案例
 首席数据官

| **【典型案例——省住建厅】**

清
确
入
建
聚
通
用
管

01 清：数据资源清查盘点

梳理形成5大类288个数据资源目录，并制定了信息系统台账和数据资源目录清单。

03 入：数据资源进账入卡

在国有“三资”管理平台，填报227项数据资源，形成省住建厅数据资源“一本账”。完成3个数据资源集的存证登记和产权登记工作。

05 聚：数据汇聚治理

编制完善《数字住建大数据中心数据标准》等，更新修订省大数据能力平台已加载的数据资源目录，保持数据资源动态更新。

07 用：数据资源开发利用

探索推进房地产市场监督数据、公积金、人企资质和工程建设项目数据开发利用，助力数字住建、数字经济发展。

02 确：数据资源“三确”

梳理出227个可填报数据资源目录，框选18类数据资源，形成3个高质量数据集。委托律师事务所、会计师事务所出具评估报告，完成确权确值工作。

04 建：建立数据供给基础

将56个系统整合为21个系统，从系统建设、管理、应用、运维、安全等方面制定了33项制度规范，已建成标准化的住建数字底座。

06 通：数据共享开放流通

按照厅网信委和厅网信办工作机制，强化住建业务协同和数据共享开放，推进公共数据资源高效供给，做好数据资源发布和供需对接。

08 管：数据资产化全过程管理

组织标技处、计财处、行政审批办公室、信息中心等处室（单位），研究公共数据授权运营审批机制，落实岗位责任，实现数据授权运营全流程管控，推动住建数据资源以产品形式转化为可交易的数据资产。

参考案例

首席数据官

【典型案例——省住建厅】

穿透式作业，打造住落地框架

穿透式作业
 形成数据资源“三破”
 工作法和方法论
 全面盘点：288个数据资源目录
 梳理确认：227个“三资”平台数据
 资源登记：18类数据资源
 确权确值：3个高质量数据集

清单式推进，构建工作体系

清单式推进
 16项工作任务，3个
 平台匹配对应目录
 任务清单 目录清单
 工作专班

形成实践案例，开展经验推广

形成案例材料
 首批试点，省直单位
 第一家完成试点任务。
 已形成《省住建厅
 数据资源资产化改革工作全链条实践
 案例》，作为样板
 全省推介。

参考案例

首席数据官

【典型案例——省住建厅】

登记18类公共数据资源；3个高质量数据集存证证书、产权登记证书；3个高质量数据集确权确值报告。

数据存证证书

证书编号：202503010001
 数据资源名称：湖北省住房和城乡建设厅“三资”平台数据
 数据提供方：湖北省住房和城乡建设厅
 数据接收方：湖北省住房和城乡建设厅
 数据接收日期：2025-03-01 14:30:00
 数据接收地点：湖北省住房和城乡建设厅
 数据接收人：湖北省住房和城乡建设厅
 数据接收单位：湖北省住房和城乡建设厅
 数据接收地址：湖北省住房和城乡建设厅
 数据接收电话：湖北省住房和城乡建设厅
 数据接收邮箱：湖北省住房和城乡建设厅
 数据接收网址：湖北省住房和城乡建设厅
 数据接收二维码：湖北省住房和城乡建设厅
 数据接收二维码地址：湖北省住房和城乡建设厅
 数据接收二维码电话：湖北省住房和城乡建设厅
 数据接收二维码邮箱：湖北省住房和城乡建设厅
 数据接收二维码网址：湖北省住房和城乡建设厅
 数据接收二维码二维码地址：湖北省住房和城乡建设厅
 数据接收二维码二维码电话：湖北省住房和城乡建设厅
 数据接收二维码二维码邮箱：湖北省住房和城乡建设厅
 数据接收二维码二维码网址：湖北省住房和城乡建设厅
 数据接收二维码二维码二维码地址：湖北省住房和城乡建设厅
 数据接收二维码二维码二维码电话：湖北省住房和城乡建设厅
 数据接收二维码二维码二维码邮箱：湖北省住房和城乡建设厅
 数据接收二维码二维码二维码网址：湖北省住房和城乡建设厅

产权登记证书

证书编号：202503010001
 数据资源名称：湖北省住房和城乡建设厅“三资”平台数据
 数据提供方：湖北省住房和城乡建设厅
 数据接收方：湖北省住房和城乡建设厅
 数据接收日期：2025-03-01 14:30:00
 数据接收地点：湖北省住房和城乡建设厅
 数据接收人：湖北省住房和城乡建设厅
 数据接收单位：湖北省住房和城乡建设厅
 数据接收地址：湖北省住房和城乡建设厅
 数据接收电话：湖北省住房和城乡建设厅
 数据接收邮箱：湖北省住房和城乡建设厅
 数据接收网址：湖北省住房和城乡建设厅
 数据接收二维码：湖北省住房和城乡建设厅
 数据接收二维码地址：湖北省住房和城乡建设厅
 数据接收二维码电话：湖北省住房和城乡建设厅
 数据接收二维码邮箱：湖北省住房和城乡建设厅
 数据接收二维码网址：湖北省住房和城乡建设厅
 数据接收二维码二维码地址：湖北省住房和城乡建设厅
 数据接收二维码二维码电话：湖北省住房和城乡建设厅
 数据接收二维码二维码邮箱：湖北省住房和城乡建设厅
 数据接收二维码二维码网址：湖北省住房和城乡建设厅

参考案例



【典型案例——智品堂与武汉燃气场内交易】

以市场需求为引导，中部中心为智品堂（数据需求方）匹配了武汉燃气（数据提供方）的相关数据资源，促进了双方在湖北省数据流通交易平台开展数据产品交易，切实助力金融机构解决“贷前风险识别”和“贷后欠款追回”两大难题。



案例示范性

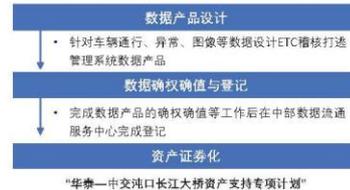
以需求为导向拉动了数据供需双方在场内开展数据流通交易，以数据流通交易激活了国有数据资源的价值，发挥了数据要素的乘数效应，为国有数据资源在多行业的应用与价值实现提供了借鉴。

参考案例



【典型案例——沌口大桥数据资产证券化】

“中部中心”前期为中交集团沌口大桥项目开展数据产品的设计，提供确权登记服务，后续通过证券化操作，成为市场上首例将高速公路数据资产融入ABS的创新实践，实现了“基础设施收益+数据资产增值”的双重价值机制。



数据资产价值体现

- ① 一是创新融资结构，把实体基建项目与数字化信息结合打包成金融产品，就像给传统公路/桥梁装上“数据导航仪”，让投资者能更清晰地看到项目价值，从而吸引证券市场的资金活水。
- ② 二是强化风控体系，数据的动态更新相当于给基建资产装了24小时监控摄像头，人流车流、设备状态一目了然，既提高安全管理水平，又让金融机构更愿意放贷，提升传统基建资产透明度和安全性，实现增值；
- ③ 三是释放溢价空间，通过精准评估数据价值，能帮助项目获得额外10%左右的信用加成，相当于用同样的底层资产撬动更多融资额度，实现资金规模的突破性增长。



可信数据空间&高质量数据集智能化应用

IITC工信人才岗位能力评价培训
数据产业高层次专业人才培养课程

 首席数据官
GOVCDO

可信数据空间&高质量数据集 智能化应用

讲师：胡俊 中国电子

2025年10月

可信数据空间

目/录 CONTENTS

01	02	03	04
建设思路	核心技术	运营机制	高价值场景

01

可信数据空间建设思路

建设背景：国际竞争驱动

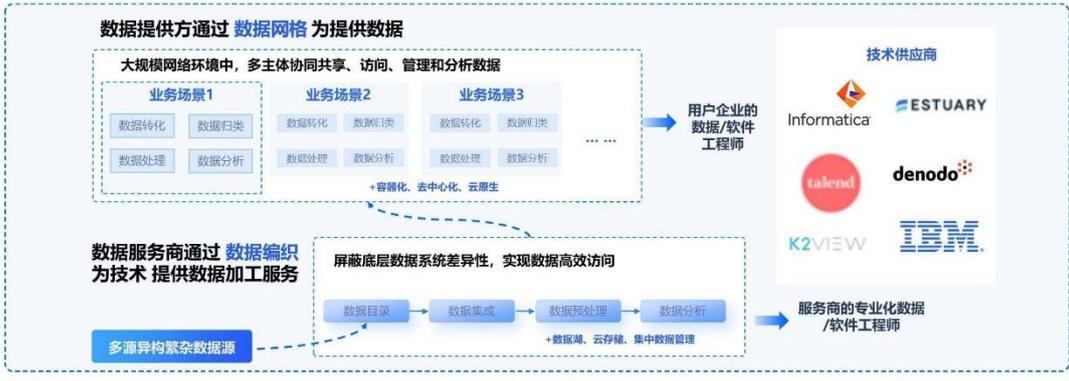
首席数据官 GOVCDO

	美国	欧盟	中国
空间模式	市场驱动平台化模式	联邦式去中心化模式	政府主导混合式模式
主导力量	市场+科技巨头	政策+行业联盟	政府+龙头企业
技术特点	云平台+API开放	联邦架构+数据主权	隐私计算+区块链
治理模式	行业自律	法律强制+多方协同	国家统筹+标准引领
典型应用	FDX (金融)	Catena-X (工业)	工业互联网 (卡奥斯)

建设背景：美国可信数据空间技术路线的研究



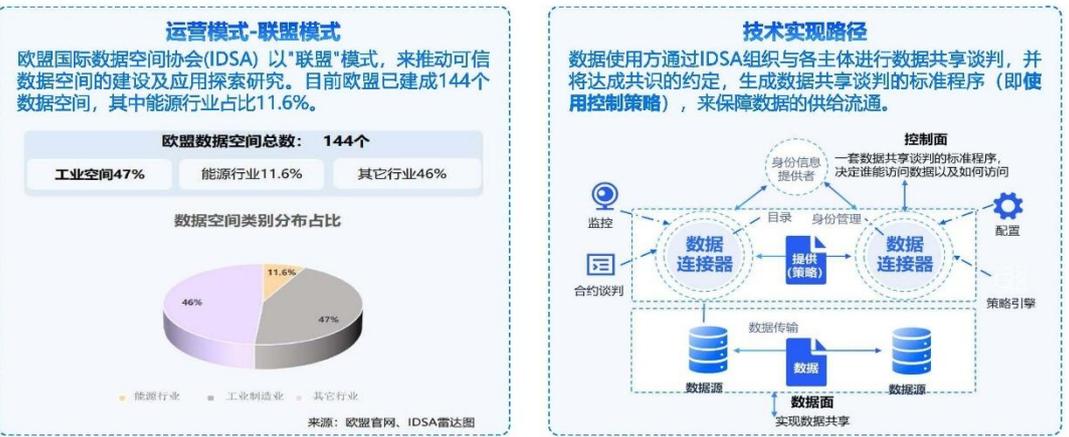
技术路线：美国侧重提高数据资源开发效率，利用**数据编织和网格技术**，协助数据提供方利用数据网络技术提供数据，委托数据服务商开展数据加工服务。
运营模式：企业自主，亚马逊、微软等云厂商积极加入欧盟Gaia-X和Catena-X等组织，研发并在云和本地部署数据空间连接器，实现数据可控流通的基础上提供数据分析。



建设背景：欧盟可信数据空间技术路线



技术路线：欧盟通过“**使用控制**”确保所有数据的数据主权，可信的参与者之间能够共享敏感和有价值的**数据资产**。
方案不足：欧盟数据空间方案在数据传输加密方面存在不足，原始数据须以明文方式出域，进行**数据共享交换**。



建设背景：中国可信数据空间技术路线

技术路线：中国结合自身的国情，在充分保障数据安全的前提下，提出三大自主创新的技术路线。
运营模式：国家数据局统筹，2024年-2028年开展城市、行业、企业、个人、跨境可信数据空间的试点建设，2025年完成63家第一批申报试点名单公布，2026年开展第二批试点申报。

- 在可信环境下，利用数据连接器技术，通过数字合约、数据使用控制引擎，打造具备数据跨域使用控制能力；
- 通过数据跨域使用控制、数据沙箱等多种技术融合应用，支持更广泛多来源、结构化数据可控使用需求，降低用户数据使用成本；
- 通过使用管控技术与标识、区块链等多种技术协同应用，形成数据资源发布发现、共享使用过程可信存证与追溯体系。

交换数据空间

数据流：——→
控制流：- - ->

该图展示了数据交换空间的架构。中心是“数据市场（数据offer）”，连接着“提供方（连接器）”和“消费方（连接器）”。

- 提供方（连接器）：**包含社会数据、国企数据、公共数据、委办局数据。数据通过“数据目录”（目录1、目录2）和“数据合约”（2）进行发布。提供“数据”（6）和“使用控制策略”（7）。
- 数据市场（数据offer）：**接收数据（5）并分发数据（4）。
- 消费方（连接器）：**接收数据（4），进行“数据共享开放数据目录”（8）和“消费方自有数据”（7）。生成“结果数据”（9），供“数据消费方”（9）进行“查看/分析/训练”。

建设背景：国家战略引领域

习近平总书记：数字经济正在成为重组全球要素资源、重塑全球经济结构、改变全球竞争格局的关键力量。
 ——《不断做强做优做大我国数字经济》

- 2019** **数据成为生产要素**
10月，十九届四中全会首次将数据纳入生产要素，与土地、劳动、资本等传统要素并列，标志着数据在数字经济时代的核心地位确立。
- 2022** **数据二十条发布**
12月，《关于构建数据基础制度更好发挥数据要素作用的意见》，系统提出数据产权、流通交易、收益分配和安全治理等制度框架。
- 2023** **国家数据局成立**
10月25日，国家数据局正式挂牌，负责统筹数据基础制度建设、数据资源整合共享及数字中国建设，推动数据要素市场化改革。
- 2024** **数据要素行动计划**
1月，国家数据局联合17部门印发《“数据要素×”三年行动计划》，聚焦12个重点领域，推动数据要素乘数效应，目标到2026年数据交易规模倍增。
- 2024** **可信数据空间行动**
11月，国家数据局发布《可信数据空间发展行动计划》，提出到2028年建成100个以上可信数据空间，构建安全可信的数据流通生态。
- 2025** **公共数据共享条例**
5月，国务院颁布《政务数据共享条例》，规范政务数据分类分级共享机制，打破“数据孤岛”，提升政府数字化治理能力。

建设思路：城市可信数据空间（1+3+5）



以“技术+制度+场景”互为支撑，打造城市数据要素现代化综合枢纽，成为区域产业链高质量发展新引擎。

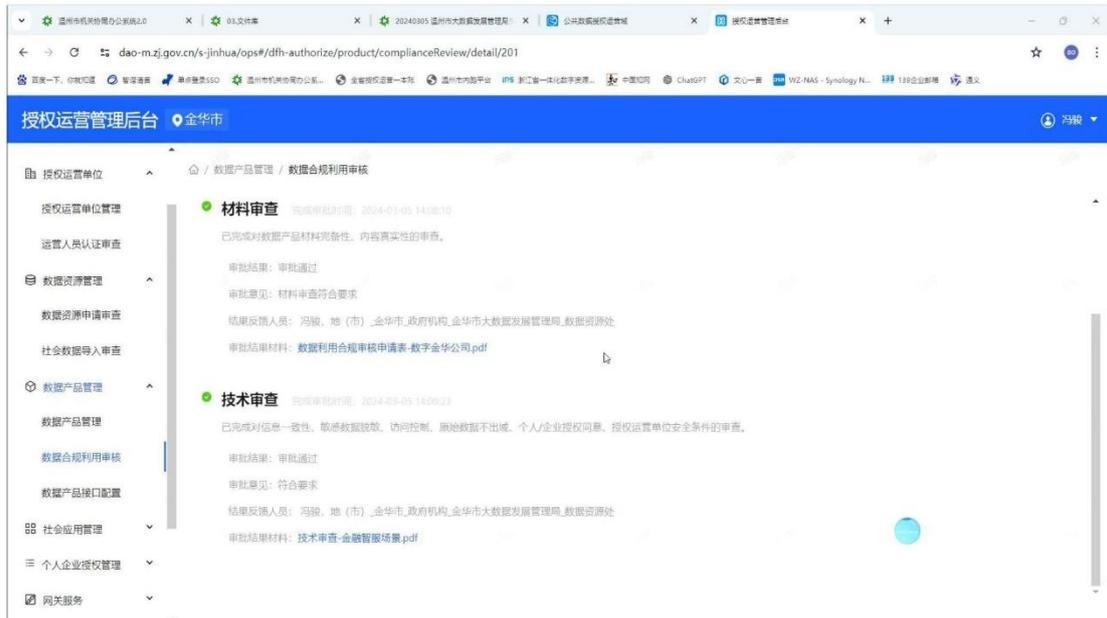


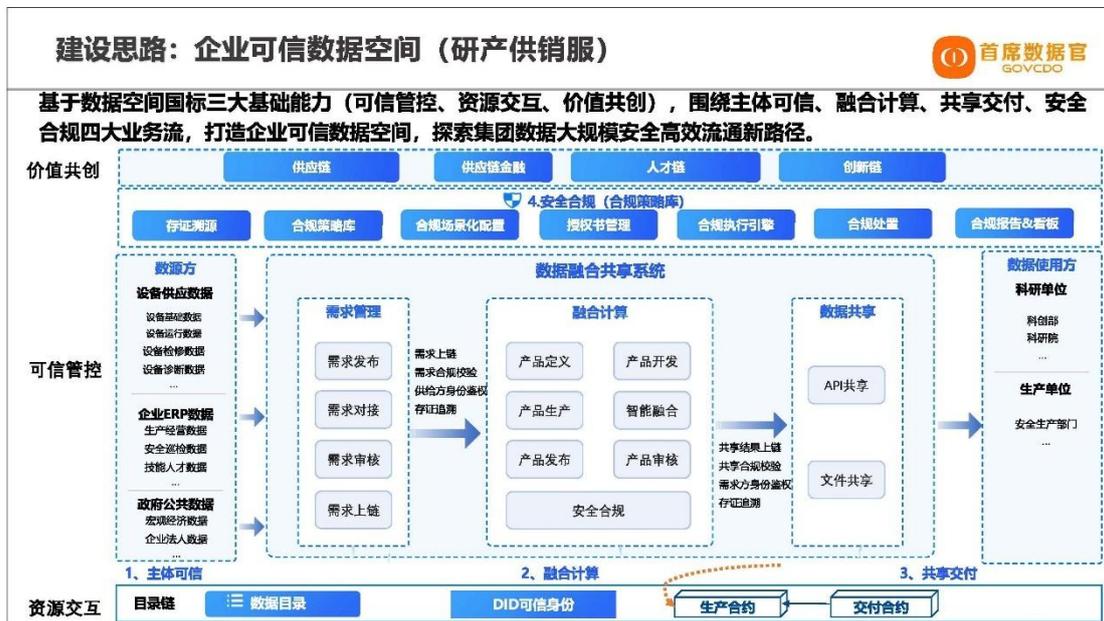
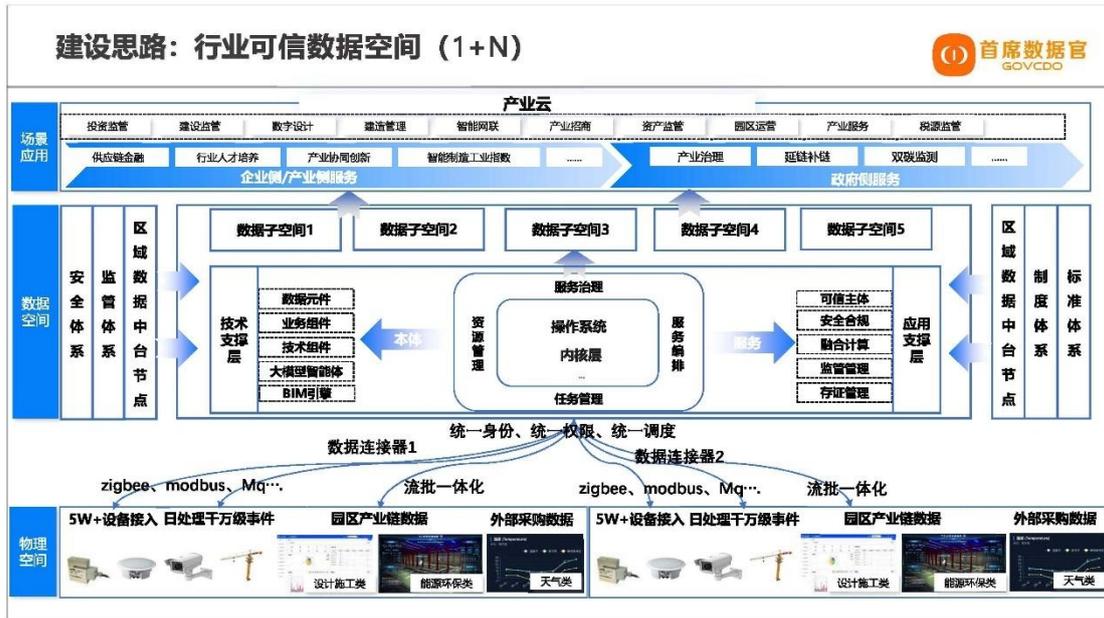
建设思路：城市可信数据空间（1+3+5）

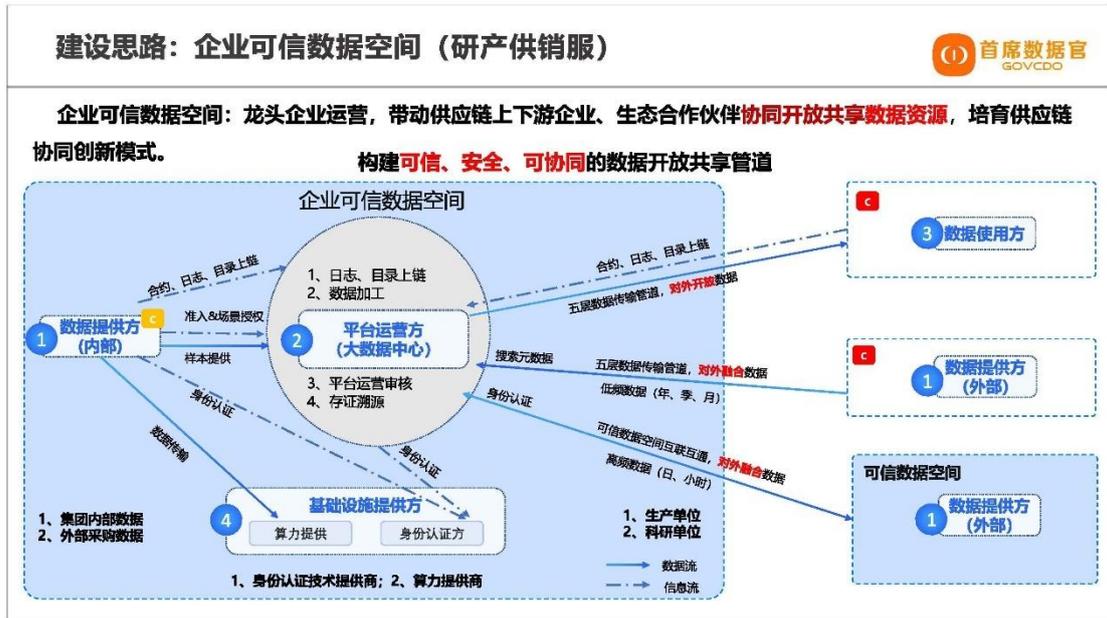


授权运营模式	整体授权	领域授权	场景授权
授权方式	<p>地方政府授权综合运营主体（一般为国有企业），面向全领域公共数据，牵头组织其他开发主体，共同开展公共数据开发利用。</p>	<p>根据不同行业或领域的的数据特点或特定行业或领域的应用需求情况，分别选择具备行业属性的运营机构开展工作。</p>	<p>该模式实行“一场景一审批”，鼓励“先有场景设计规划，后有授权运营”，准确地实现数据资源与需求相匹配。</p>
职责分工	<p>数据主管部门：负责综合运营主体征集、管理、评价。 综合运营主体：负责公共数据开发利用平台建设运营。负责其他开发主体管理。 其他开发主体：开展公共数据加工生产。</p>	<p>数据/行业主管部门：负责分领域运营主体管理；牵头公共数据开发利用平台建设。 分领域运营主体：负责本领域公共数据开发利用，如（金融、教育、医疗等领域）。</p>	<p>数据主管部门：负责综合运营主体管理。 综合运营主体：重点负责跨领域、跨区域的综合应用场景建设。 分领域运营主体：本领域公共数据开发利用。</p>

首席数据官 GOVCDO

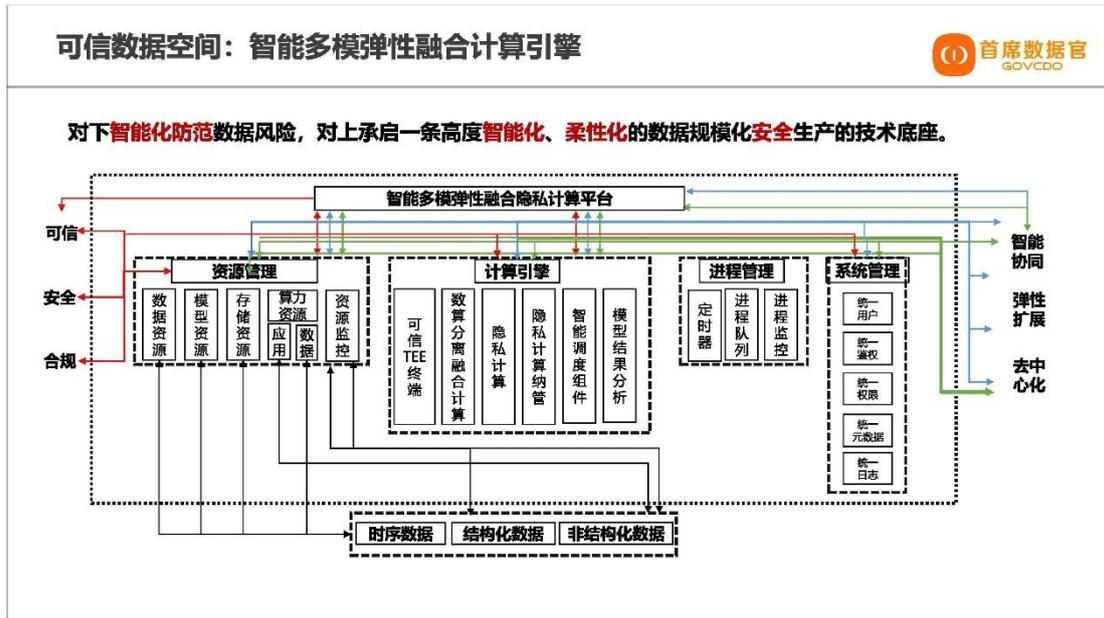






02

可信数据空间核心技术

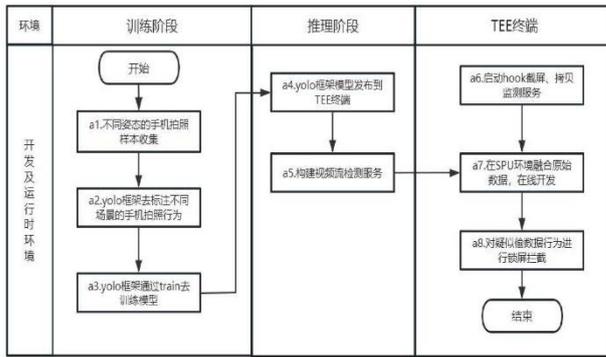


可信数据空间：可信TEE终端



核心技术：人体行为感知CV模型-实时监测手机拍照，可信终端自动锁屏；操作系统级HOOK+NLP-实时对截屏、拷贝U盘行为，进行内容解析，实时拦截。

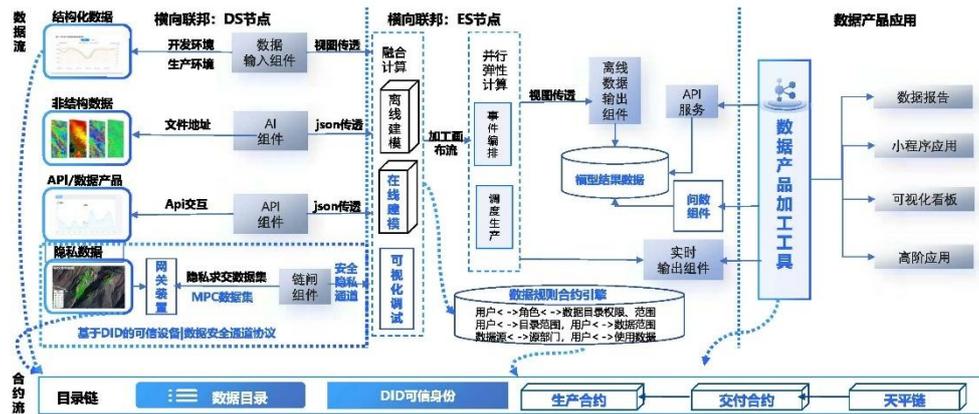
SPU卡：最大数据传输速度：7.88GB/s，密钥交换算法采用国密SM2，密钥加密算法采用国密SM4，密钥长度256BIT。



可信数据空间：数算分离模式的融合计算



在数据区、算力区分离的前提下，通过AI模型、联邦学习技术，构建智能多模弹性融合计算引擎，在开发方不接触原始数据的安全前提下，全自动完成算法模型安全扫描、打包、发布；并通过共享式智能调度引擎降低算力资源需求，从而实现大规模、低成本数据融合计算。



可信数据空间：数算分离模式的融合计算

元件名称	创建人	创建时间	信息领域	元件状态	操作
肺炎支原体流行性分析数据元件	developer1	2023-12-05 17:36:46	医疗卫生	发布审核中	撤回 审核评价
新冠病毒研究数据元件	developer1	2023-12-01 01:47:14	医疗卫生	元件开发中	元件开发 元件发布
Rpa元件测试	developer1	2023-11-09 05:46:12	教育科技	元件开发中	元件开发 元件发布
北京市小学教师平均工资	supplier	2023-11-02 08:02:24	教育科技	元件开发中	元件开发 元件发布
1012核酸元件	developer1	2023-10-12 06:28:48	金融财经	元件开发中	元件开发 元件发布
核酸元件	developer1	2023-10-12 06:18:33	金融财经	元件开发中	元件开发 元件发布
1012用电量查询元件	developer1	2023-10-12 06:04:47	能源能源	元件开发中	元件开发 元件发布
保险公司承保数据元件	developer1	2023-10-09 01:56:26	金融财经	元件开发中	元件开发 元件发布
IQ测试元件结果同步元件	developer1	2023-09-19 07:25:23	教育科技	元件开发中	元件开发 元件发布
zboeblde在线模型数据元件	developer1	2023-09-15 08:17:18	教育科技	元件开发中	元件开发 元件发布

可信数据空间：联邦平均模式的隐私计算

高质量数据集构建的场景：通过联邦模式进行遥感图像解译训练，在保障“原始数据不动，模型动”的原则下，进行多方模型联合训练，将准确率由90.14%提升到95.65%。

联邦平均算法

Algorithm1 FederatedAveraging

The K clients are indexed by k ; B is the local mini batch size, E is the number of local epochs, and η is the learning rate.

Server executes:

```

initialize  $w_0$ 
for each round  $t = 1, 2, \dots$  do
 $s_t \leftarrow \text{rand}(C, K)$ 
 $S_t \leftarrow \{\text{random set of } m \text{ clients}\}$ 
For each client  $k \in S_t$  in parallel do
 $w_{k,t}^0 \leftarrow \text{ClientUpdate}(k, w_0)$ 
 $m_t \leftarrow \sum_{k \in S_t} m_k$ 
 $w_{t+1} \leftarrow \sum_{k \in S_t} \eta w_{k,t}^E$  // Erratum
ClientUpdate( $k, w$ ): // Run on client  $k$ 
 $\beta \leftarrow \{\text{split } p_k \text{ into batches of size } \beta\}$ 
for each local epoch  $i$  from 1 to  $E$  do
for batch  $b \in \beta$  do
 $w \leftarrow w - \eta \nabla \mathcal{L}(w; b)$ 
return  $w$  to server
    
```

- K 客户端 client 在本地迭代聚合 E 次; (E 一般取 5-20, 小写的 k 表示下标)
- 然后传送给 server 进行对参数 w 最终聚合, m_k 代表每个客户端数据长度, m_t 代表所有客户端数据长度, 即聚合是对 $w_{k,t}^E$ 做加权平均

然后就是重复上述过程, 直到训练结束



可信数据空间：“四模融合”共享技术



数据产品在安全区内完成场内交付，其交付的服务形式多元化，可以AI agent智能体或API推拉模式满足不同场景的实际需求，如：数据集推送，消息流实时API推送服务，API请求服务等。

阳江青洲气候数据

智能助手

为您找到5个数据产品，请问您找的是【阳江青洲降雨数据】【阳江青洲风力数据】【阳江青洲气温数据】【阳江青洲空气质量数据】【阳江青洲温度数据】？

阳江青州降雨数据

智能助手

以下是【阳江青洲降雨数据】



请【点击】查看

智能助手

- 意图识别模型
- 多轮对话模型
- 智能问数模型
- 大模型
- 知识库&规则库

自然语言交互

智能服务

自然语言理解 意图识别 场景域服务

场景模型适配

语料库规则库 知识图谱 统计分析模型

AI

模型训练

语义依存树

看板

可视化数据

统计模型

可信数据空间：“四模融合”共享技术



28°C / 33°C 多云转晴
小E电力行业-现货交易助手
2025年06月28日 星期六

您好，小E智能助手为您服务，我可以为您提供如下几点服务

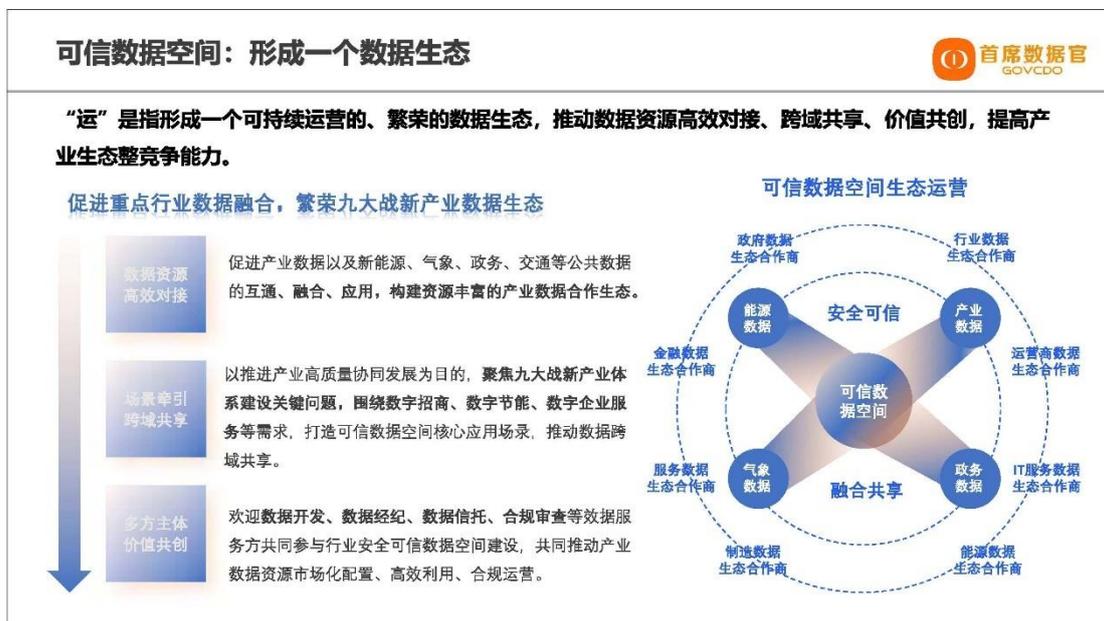
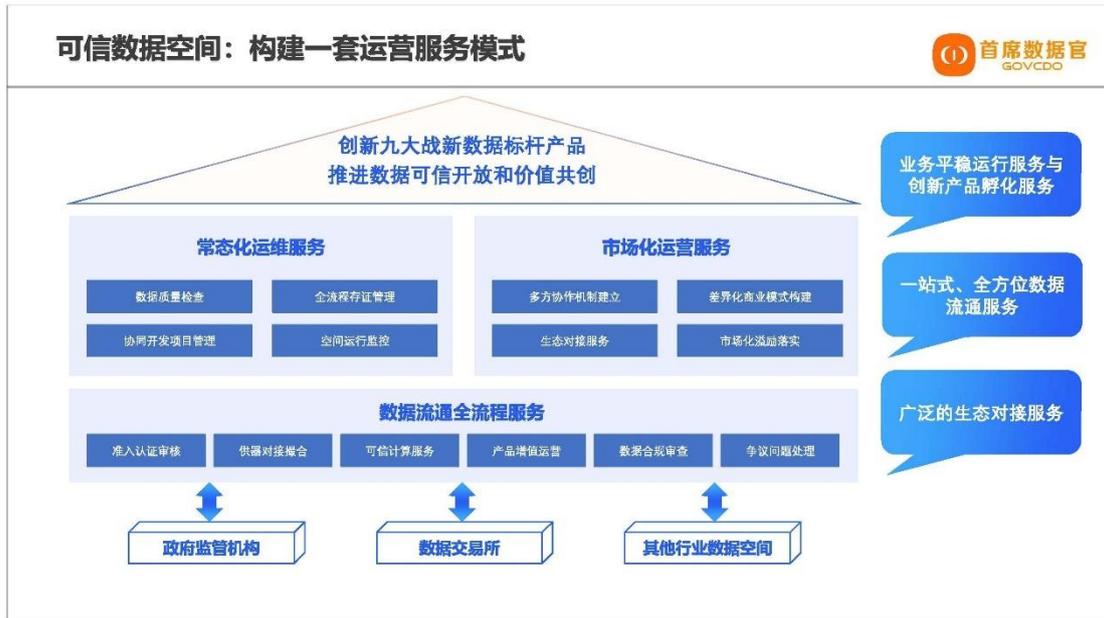
- 1、信息查询服务，涉及现货交易日报生成报告，现货交易日报披露
- 2、现货市场电价预测
- 3、报价方案策略推荐、报价方案提交审核





03 可信数据空间运营机制





04

可信数据空间高价值场景

应用场景：依托共性场景引领，赋能千行百业新篇

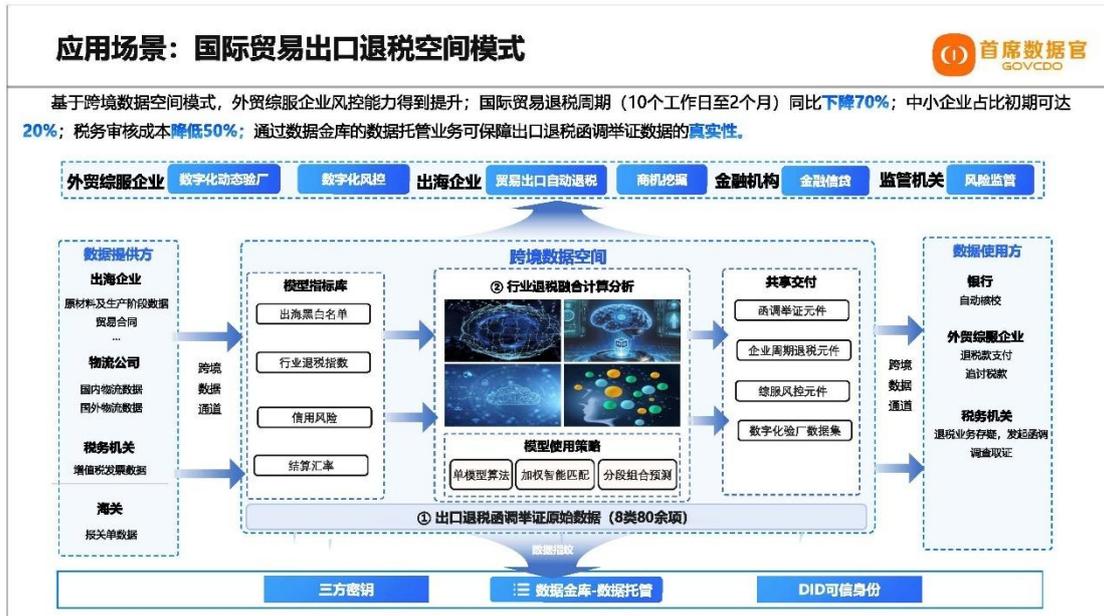
首席数据官 GOVCDO

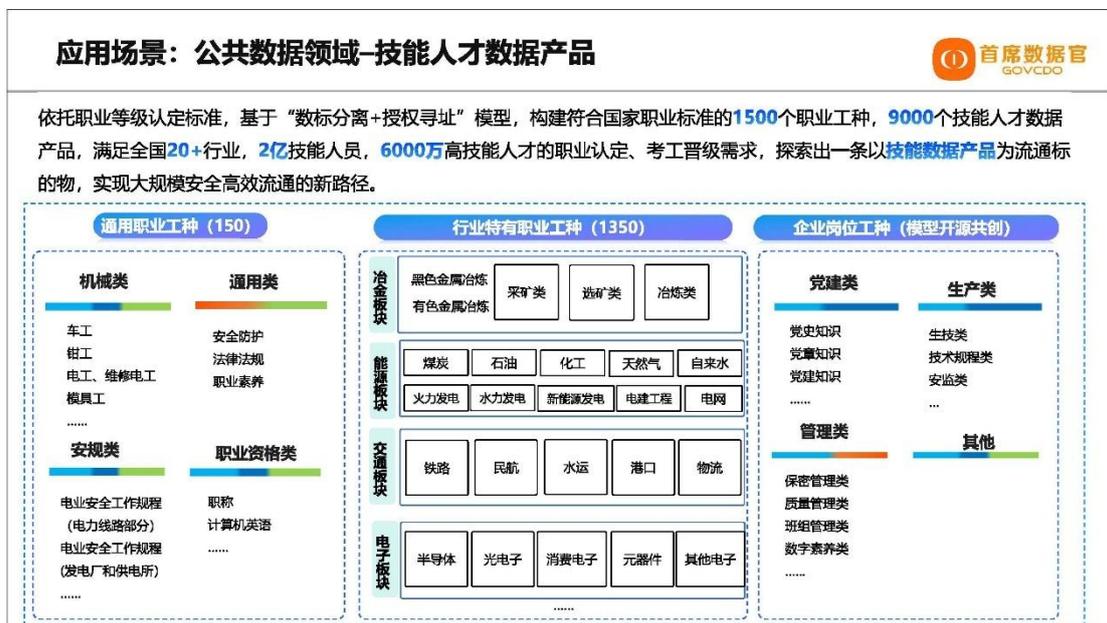
围绕可信数据空间价值共创能力提升，面向共性应用场景，支持可信数据空间运营者部署应用开发环境，为参与各方开发数据产品和服务创造条件。

本技术方案秉承“以数为媒、场景牵引”的价值共创建设思路，通过可信数据空间来保障数据安全高效流通利用，实现数据在开发利用和共享交付过程中的可控、可管、可防。相关成果将作为通用能力赋能到企业内外部数据安全使用的场景中，为高价值场景提供安全可控的多种算法调试及运行环境。

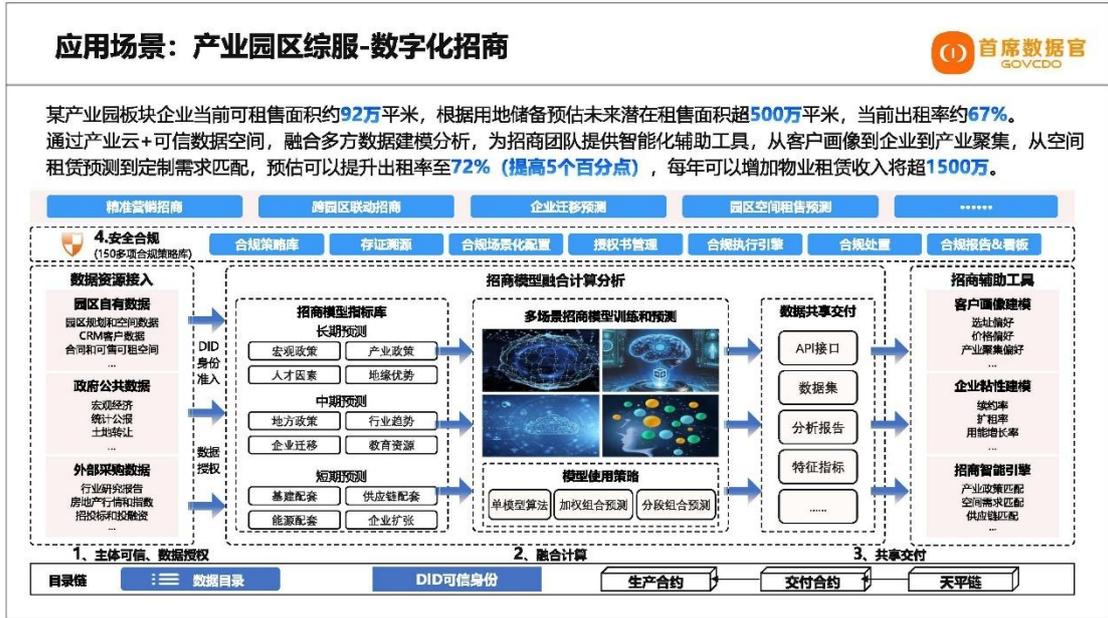
量子密码与可信数据空间的融合应用，充分利用了量子密码的高安全特性整体提升了数据共享安全，从数据采集、传输、计算、分析、存储、访问、共享等各环节，全方位、全流程地实现端设备鉴别、用户身份认证、传输加密、防篡改校验等安全防护。

应用场景









高质量数据集智能化应用

目/录 CONTENTS

01	02	03
建设思路	智能体	应用场景

建设背景：国家政策



1. 国家战略层面：《新一代人工智能发展规划》明确提出“人工智能+”重点任务，要求2025年实现政务领域AI应用规模化落地。

2. 地方推进层面：北京、上海、广东等省市出台专项政策，如《上海市人工智能赋能城市治理行动方案》，设立**政务AI专项基金**。

3. 行业推进层面：电力、金融等行业基于大模型+垂域模型，已在智能+生产领域已落地，从而降低了人工成本，促进企业人员转型。



典型案例：

1. 深圳“秒批”系统：通过**AI审核材料**，企业开办时间从3天压缩至1小时内。这一系统的应用，极大地优化了营商环境，吸引更多企业在深圳落地，促进地方经济发展。

2. 杭州“城市大脑”：**AI调度政务资源、交通违章处理、社保查询等服务**响应速度提升3倍。“城市大脑”整合了城市的各类政务数据，实现资源的优化配置，让市民能够更快地办理相关业务。

政策核心导向：

效率提升

服务下沉

数据安全

建设背景：AI需要千行百业的高质量数据集



高质量数据集定义

1. 核心要素

高价值、高密度、标准化数据，需通过采集、清洗、归类和标注等处理，可直接用于开发和训练人工智能模型，能有效提升模型性能的数据的集合。

2. 格式

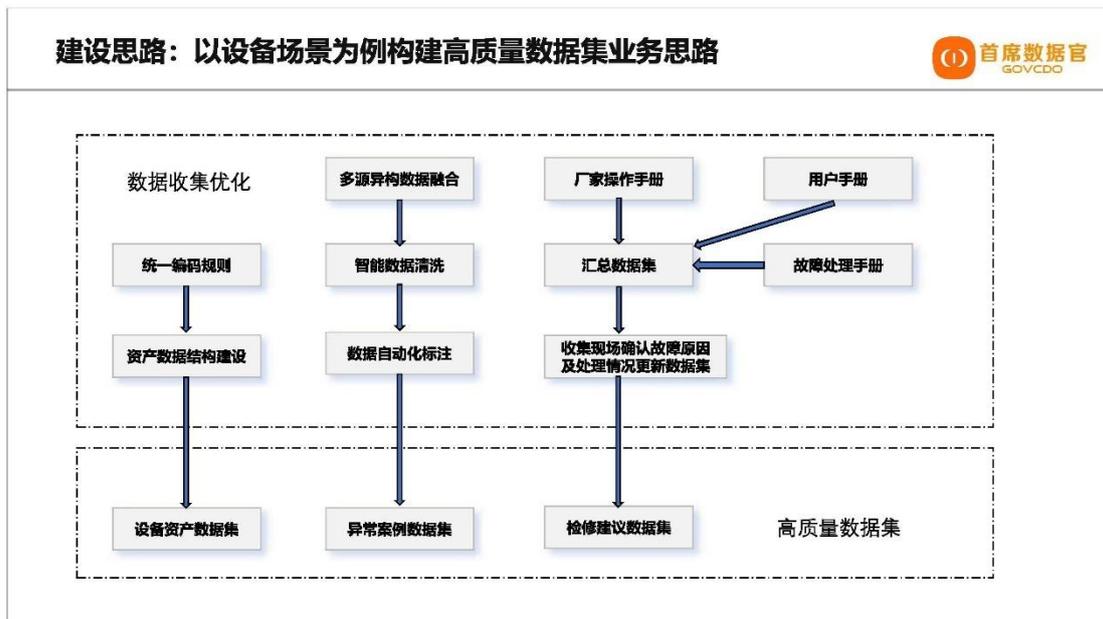
数据集标识、内容、标注、时间、版本及授权等核心元数据的结构化框架，特别针对多模态数据（如文本、图像）和标注信息制定了专项格式规范。

3. 分类

分为通识数据集、行业通识数据集、行业专识数据集。

4. 数据集质量评估

文档完整性、质量合规性、场景适用性。



建设思路：第一步-构建高质量数据集的资产台帐



当前生产环境所用编码规则

二级类码	二级类扩展码说明
002	风力发电机组系统：使用 0000 代表唯一（例如只有一个），或者 0001 代表 1 号风力发电机组，0002 代表 2 号风力发电机组

三级类码	三级类扩展码说明
003	变桨
004	变频柜
005	齿轮箱
007	发电机
008	机舱
012	液压系统
扩展规则	若设备唯一使用 0000 表示，若同一设备有多组，分别用 0001、0002 表示，例如变桨有 1、2 号变桨，则分别命名为 0001、0002

数据类码	说明
00	不区分
01	温度
02	速度
03	电压
04	电量
05	电流
06	功率
07	压力

建设思路：第一步-构建高质量数据集的资产台帐



1501 Y0 111 0 B1 002 0001 007 0000 06001

分级序号	0级	1级	2级
分级名称	管理域编码	类型域编码	业务域编码
占位	4位	2位	25位
注	管理域编码不做编码，信息化层面处理，关联、引用集团元数据系统内的组织机构代码。		

编码结构	管理域编码	类型域编码	业务域编码
分级名称	公司代码	发电类型编号、扩展编号	分类号、数据码
字符标记	A ₁ A ₂ A ₃ A ₄	T ₁ 、T ₂	全厂码 (H ₁ H ₂ H ₃)、前缀号 (F ₀)、一级类码 (F ₁ F ₂)、二级类码 (F ₃ F ₄ F ₅)、二级类扩展码 (F ₆ F ₇ F ₈ F ₉)、三级类码 (F ₁₀ F ₁₁ F ₁₂)、三级类扩展码 (F ₁₃ F ₁₄ F ₁₅ F ₁₆)、数据编码 (F ₁₇ F ₁₈)、扩展位号 (F ₁₉ F ₂₀ F ₂₁)
字符类型	ANAN	A、N	N(A)N(A)N(A)N(A)、N、AN、NNN、NNNN、NNN、N(A)NNN、NN、N(A)NN
示例	111N (可门电厂)	B (火电)、1 (燃煤)	111 (一期1号机组)、0 (内部凝聚)、B1 (生产运行)、001 (燃烧系统)、0001 (1号燃烧系统)、005 (屏式过热器)、H000 (屏式过热器后屏)、03 (管壁温度)、020 (第20管屏温度)

建设思路：第二步-构建异常案例高质量数据集



设备状态异常类

设备状态功能对风电、光伏、储能、电气四种设备类型进行预警统计、趋势分析，磁贴展示模型执行的状态。



建设思路：第三步-构建检修建议高质量数据集

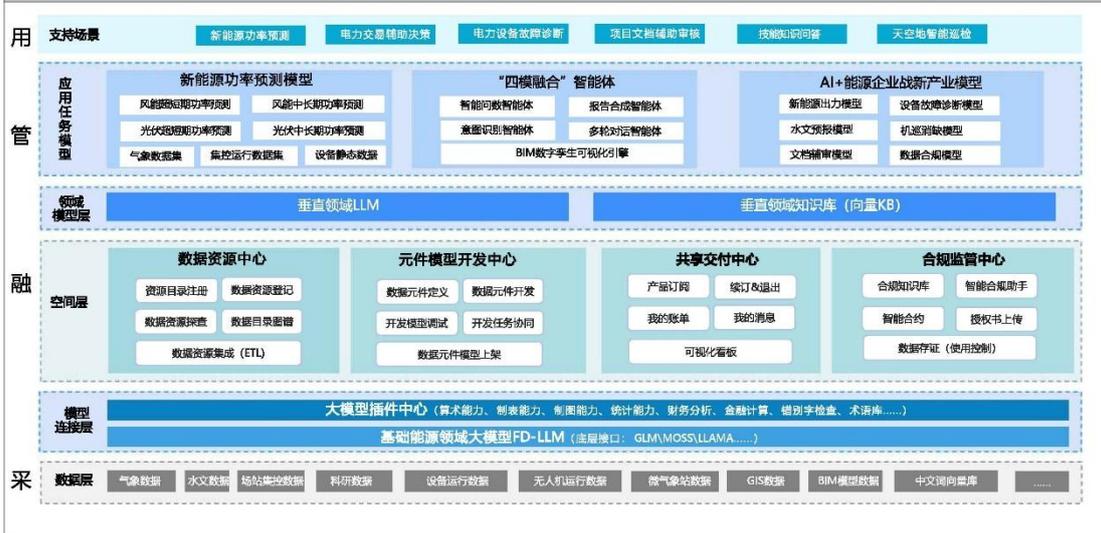


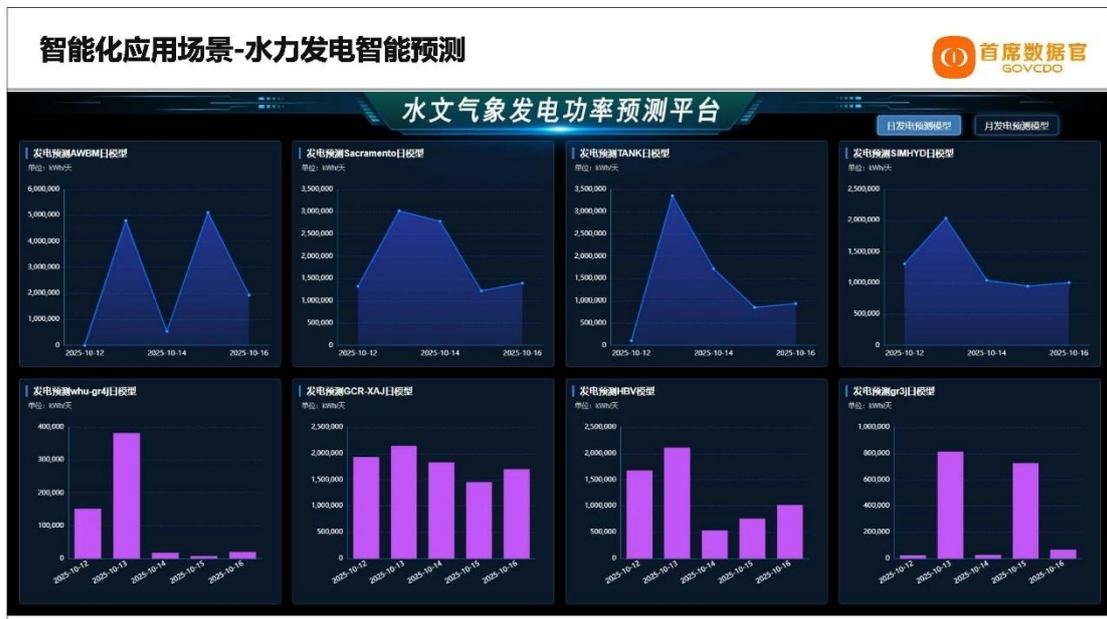
检修建议数据集更新

收集现场确认故障原因及处理情况，形成文本形式资料库，更新迭代检修建议数据集，动态优化检修建议库，标注好标签的数据进行特征分析，细化现场检查情况。



智能体：基于可信数据空间构建能源领域智能体功能架构图







数据治理体系

IITC工信人才岗位能力评价培训
数据产业高层次人才培养课程

首席数据官
GOVCDO

数据治理体系

讲师：曾新科 数字理政

2025年10月

关于CDO的理解

首席数据官
GOVCDO



刘备

CDO 应当是**主公、军师** ✓



关羽
张飞

~~CDO 别抢**猛将兄**的活~~

关于CDO的理解



以《广东省首席数据官制度试点工作方案》为例：原则上首席数据官由本级政府或本部门分管数字政府改革建设工作的**行政副职及以上领导兼任**。



(二) 明确首席数据官职责范围。

- 1. 推进数字政府建设。** 组织落实省、市数字政府改革建设工作领导小组决定事项、部署任务；组织制订本级政府或本部门数字政府发展规划、标准规范和实施计划。
- 2. 统筹数据管理和融合创新。** 组织制订数据治理工作的中长期发展规划及相关制度规范；统筹管理数据普查登记、规范采集、加工处理、标准规范执行、质量管理、安全管控、绩效评估等工作；统筹协调内外部数据需求，统筹推进数据共享开放和开发利用工作，推动公共数据与社会数据深度融合和应用场景创新，积极推进重点领域应用落地实施。
- 3. 实施常态化指导监督。** 协调解决本级政府或本部门信息化项目建设中的重大问题，对信息化项目的立项、验收工作拥有一票否决权。对数据治理运营、信息化建设等执行情况进行监督，及时发现、制止及纠正违反有关法律法规、方针政策和可能造成重大损失的行为。
- 4. 加强人才队伍建设。** 试点市、县(市、区)政府首席数据官负责推动本级数据运营机构建设，组织开展本级数据技能与安全培训工作。试点部门首席数据官负责推进本部门数据治理及运营团队建设，并组织开展本部门全员数据技能与安全培训。

关于CDO的理解



数据治理产业如此复杂，分工如此细致，**如何做好主公与军师？**



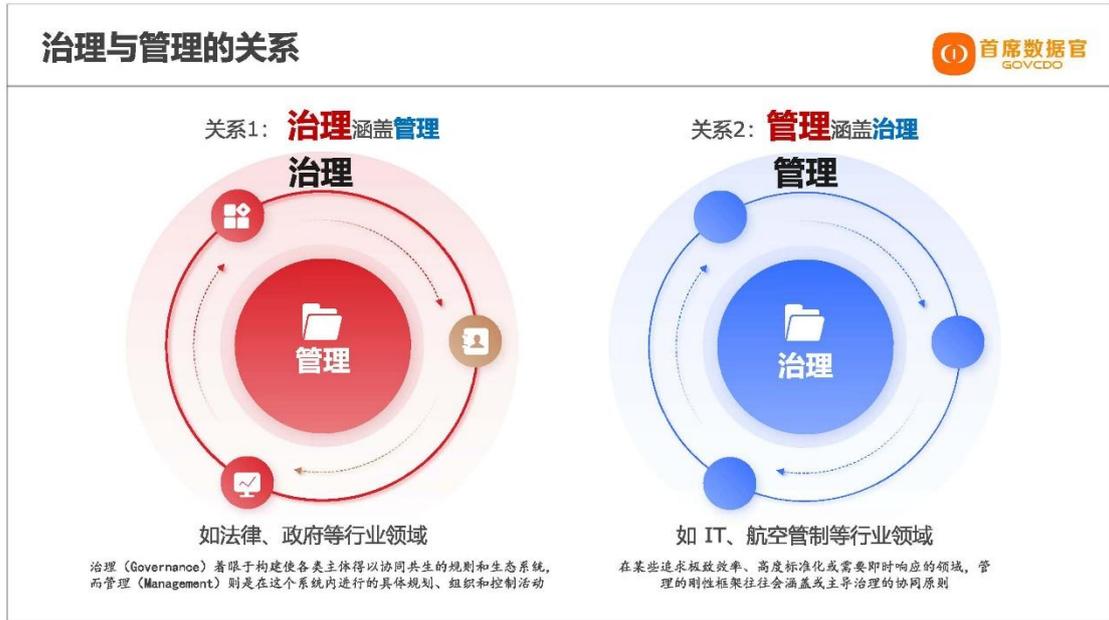
关于CDO的理解



要靠见多识广、判断力!

不要焦虑





关于数据治理与数据管理关系的理论研究



国际标准中数据管理与数据治理概念的比较研究

20250830 修改稿

孙智莹¹ 安小米^{1,2}

¹中国人民大学信息资源管理学院, 北京, 100872; ²数据工程与知识工程教育部重点实验室, 北京, 100872;

³中国人民大学智慧城市研究中心, 北京, 100872

¹ School of Information Resource Management, Renmin University of China, Beijing 100872, China, 2 Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), Ministry of Education, Beijing 100872, China, 3 Smart City Research Centre, Renmin University of China, Beijing 100872, China

摘要: [目的意义] 本研究旨在厘清数据管理与数据治理的概念与概念关系, 解决当前二者概念混淆与适用问题, 为我国建立健全数据管理与数据治理的标准体系、推动国际标准和标准制定提供参考, 促进我国参与数字领域国际规则和标准制定。[方法/过程] 本文采用文本内容分析方法, 按照 ISO 704:2022 的概念构建原则, 对 ISO、IEC、ITU-T 国际标准化组织在线数据库中数据管理和数据治理的定义进行分析, 从中识别出二者的概念对象、特征和关系类型并构建核心概念关系图, 从应用层次、概念对象、概念特征和概念关系四个维度比较二者的概念异同。[结果/结论] 研究发现, 数据治理涉及战略和中观层面, 包括数据治理、过程管理和价值化实现的综合活动; 数据治理覆盖宏观、中观和微观层面, 强调质量管理、权责界定、战略协同及政策制定; 二者共同关注过程要素和数据价值化实现, 呈关联关系。

关键词: 数据管理; 数据治理; 核心概念; 概念关系; 国际标准

分类号: G251

Comparative Study on the Concepts of Data Management and Data Governance in International Standards

表 3 数据管理与数据治理概念的核心概念比较

Table 3 Comparison of Core Concepts Between Data Management and Data Governance

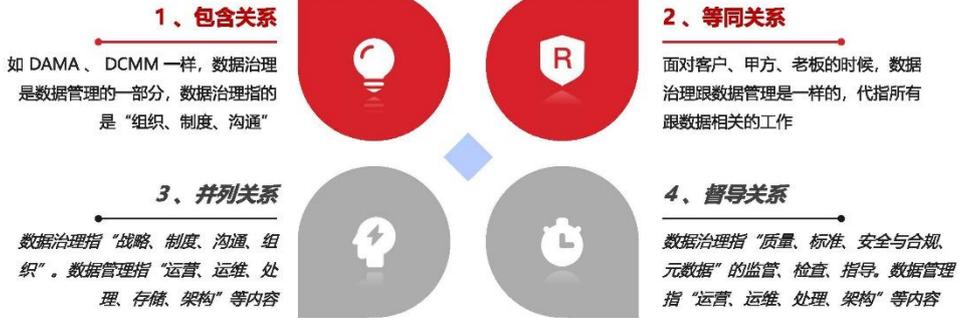
比较维度	数据管理	数据治理
概念对象	数据处理功能、数据价值化实现功能 数据库管理 数据处理活动、大数据架构实施活动 数据追踪过程、数据处理流程或系统	数据资产管理活动协调能力 数据管理政策 数据资产管理活动 数据管理活动 数据管理过程
概念特征	具体操作与功能实现	战略协调与规划 政策制定 权责界定 质量把控
概念关系	关联关系 数据治理的具体实施 为数据管理提供宏观框架和指导原则	
应用层次	中观、微观	宏观、中观、微观

[作者简介] 孙智莹, 女, 2002 年生, 中国人民大学信息资源管理学院硕士研究生。安小米, 女, 1965 年生, 中国人民大学信息资源管理学院教授 (通讯作者)。

数据治理与数据管理在实际工作中，当前概念也不统一



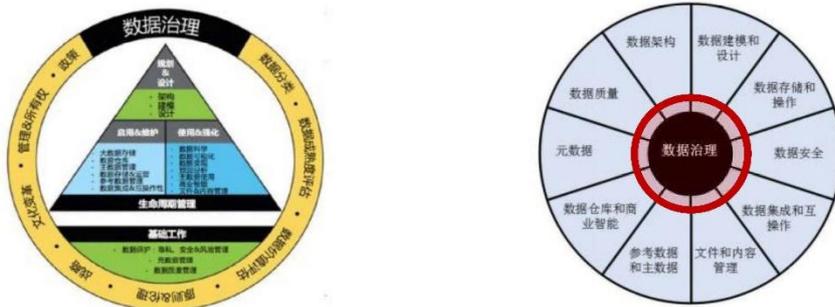
在工作中，要知道“包含关系”与“等同关系”



DAMA(Data Management Association 国际数据管理协会)



- DAMA 车轮图定义了数据管理知识领域。它将数据治理放在数据管理活动的中心，因为治理是实现功能内部一致性和功能之间平衡所必需的。
- 其他知识领域（数据体系结构、数据建模等）围绕车轮平衡。它们都是成熟数据管理功能的必要组成部分，但根据各组织的需求，它们可能在不同的时间实现。



DAMA 发布《DAMA 数据管理知识体系指南》(DMBOK)，涵盖数据治理、数据安全、数据质量等 11 个核心领域

DCMM(Data management Capability Maturity Model 数据管理能力成熟度模型)



GB/T 36073-2018 数据管理能力成熟度评估模型 (Data Capability Maturity Model, 简称DCMM) 是一个针对**组织数据管理、应用能力的评估框架**, 通过数据管理能力成熟度模型, 可以清楚的**定义**组织数据管理**当前所处的发展阶段以及和未来发展方向**。DCMM是国家大数据重点标准之一, 主要由8个能力域、28个能力项、5个成熟度等级组成。



数据管理能力成熟度评估模型框架图



适用对象	特征
数据拥有方 (甲方)	具有较好的信息化基础, 积累了大量的数据, 拥有数据的所有权、管理权。
信息技术服务方 (乙方)	拥有数据管理产品, 为数据拥有方提供解决方案, 输出数据管理服务能力、咨询能力。

数据治理核心概念



数据治理不关心具体的技术和流程。

数据治理不关心具体技术和流程



← 同理 →

技术是工具, 流程是载体, 治理只问两件事: 谁负责? 按什么规矩?

数据治理核心概念

首席数据官 GOVCDO

数据治理不直接管数据，而是管“怎么管数据”

数据治理管“怎么管数据”

不是动手干活，而是让该干的活有章可循、有人担责

数据治理核心概念

首席数据官 GOVCDO

数据治理（指挥） vs 数据管理（执行），谁才是“老大”？

对比维度	数据治理	数据管理
定义差异	不是简单“管数据”，而是“管怎么管数据” 数据管理的统筹者	具体执行数据相关工作，如架构、建模等
地位角色	位于 DAMA 车轮图中心，类似大脑、指挥官 制定规则、分配责任、监督执行	围绕在车轮图外围，是具体工作执行者
类比说明	如同教练，不亲自下场，制定战术等 确保数据可用、好用、安全用	如同球员，负责上场踢球执行

数据治理掌舵，数据管理划桨，协同方能驶向数据价值彼岸

数据治理的核心活动



数据治理的核心是确保数据被恰当地管理，而非直接管理数据，且需将监督和执行的职责分离



概括起来就是：“管人、管规矩、管沟通”

政企行业推进数据治理的典型问题



要警惕“葫芦现象”：两头大、中间细，各搞各的，协同不畅，业务与IT技术脱节，价值难有效发挥！



理论借鉴：智慧城市的数据治理与利用能力框架



安小米教授等专家学者在《我国政府数据治理与利用能力框架构建研究》、《基于国际标准的“数据利用”核心概念及概念体系研究》、《中国式现代化视角下的数据利用新诠释：内涵辨析与能力框架构建》等一系列论文中，以及《信息技术 智慧城市数据利用 第1部分：框架 ISO/IEC PW1 10267-1 Information technology—Data use in smart cities— Part 1: Framework》、《城市全域数字化转型 城市数据有效利用水平评估模型》等系列国家、国际标准中，从**数据可用、数据有用、数据易用、数据善用、数据赋能智能应用和服务**五个维度提出了智慧城市数据利用框架，旨在促进数据作为城市战略资源和资产的有效、可持续、全面和创新利用

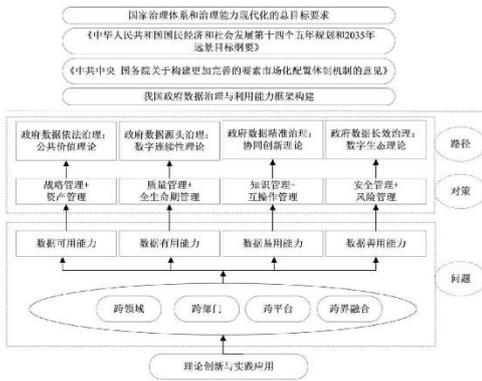
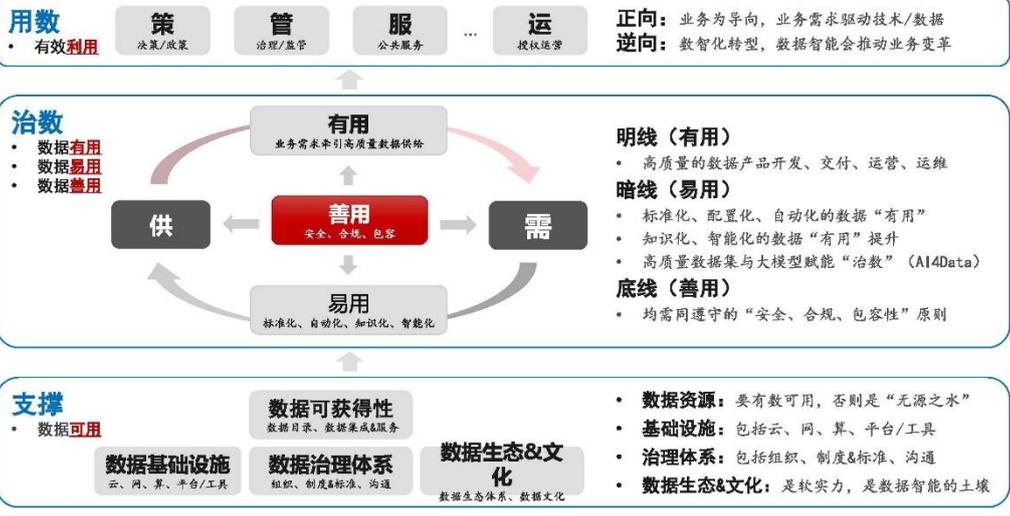


图1 我国政府数据治理与利用能力框架构建的总体思路



图 智慧城市数据利用总体框架

数据治理框架：数据治理“五用”方法论

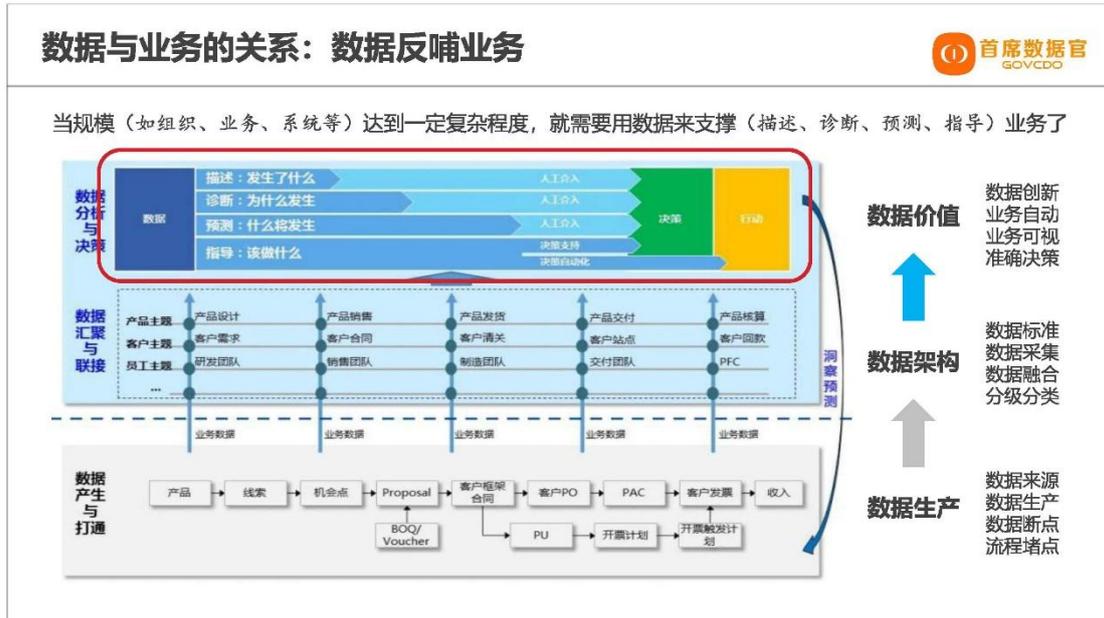




01 用数：有效利用

数据有效利用能力（数据赋能智能应用和服务）：指为特定目的对数据进行处理和管理，使数据增加价值以实现其目的的行为过程、方法、手段和服务的一组能力。

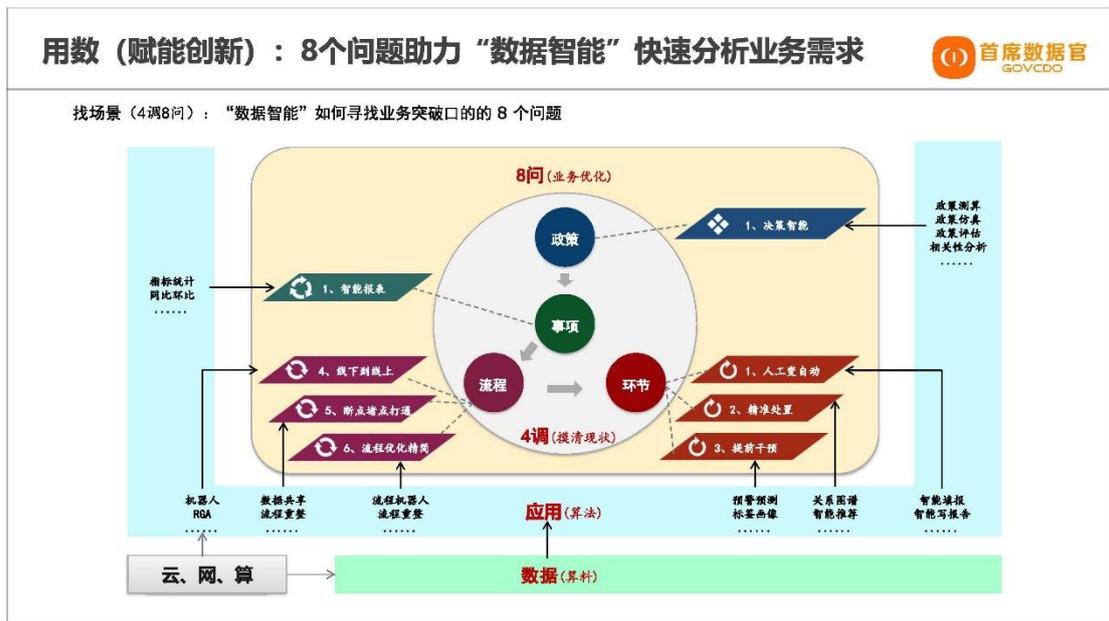
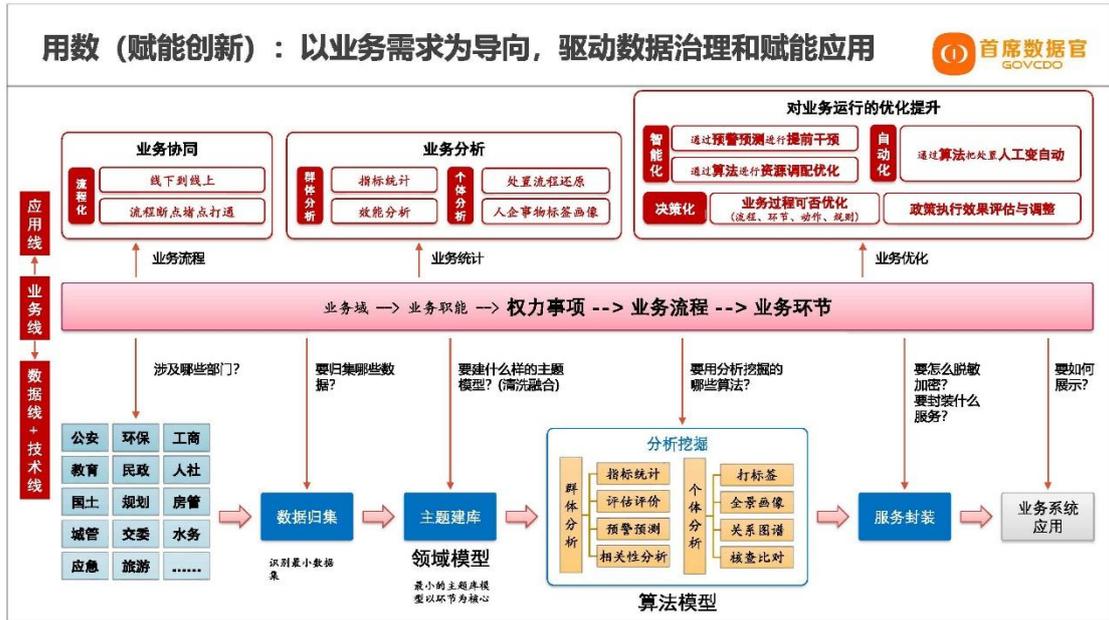




数据与业务的关系：数据反哺业务

□ 举例：财务数据运营从价值、风险、效率、精准四个方面对公司资源的记录、市场化配置、评价和报告披露提供支持

	收入管理	支出及结算管理	资产管理	资金管理	财务管理
价值 以企业价值增长为目标，推动各业务线的价值增值	应收账款分析	支出节约时效分析	资产价值收益分析 资产折旧预测分析	资金使用效率分析 现金流预测分析	财务管理绩效评估 税收筹划预测
风险 识别潜在的财务损失可能性，并采取措施加以避免	收入保障风险分析	费用支出风险分析	资产核算风险分析	资金使用风险分析	财务综合风险分析 财务审计支撑
效率 提高财务业务线效率，提高财务业务响应速度，降低人员消耗	收入核算时效分析	支出核算时效分析	资产核算时效分析	资金核算时效分析	税务核算时效分析 财务综合时效分析
精确 核算准确性是财务业务线最首要的目标，关系到监管与评价	应收账款科目追溯	应付账款科目追溯	资产科目追溯	资金科目追溯	税务科目追溯 成本追溯分析



用数（赋能创新）：业务需求与技术实现要匹配



场景分类分级（智能化水平）：从自动化到智能化，5个等级量化数据智能的技术实现目标

业务场景根据处理者的权责属性，可以分为操作型、分析型、核对应三类。

- 操作型场景：与事件处置相关的场景。如无人驾驶、智能客服、智能巡检等；
- 分析型场景：与群体性分析相关的场景。如客户行为分析、营销效果分析等；
- 核对应场景：为确保数据或流程的一致性、准确性与合规性而进行数据核对的场景。如财务对账、发票核对、订单和库存核对等；

场景分级	说明	举例
智能决策级 (T5: Intelligent decision-making)	业务的全过程高度智能化，最后的步骤（最终决策环节）是“智能决策、无人干预”。	1、无人驾驶（传感器技术、深度学习等）； 2、金融领域的智能交易系统（大数据、AI）； 3、智能化交通调度（如红绿灯）；
智能推荐级 (T4: Intelligent recommendation)	业务的全过程高度自动化，并在关键环节应用“智能推荐”，最后的步骤（最终决策环节）是“人工决策”。	1、智能分拨； 2、智能问答； 3、智能化医疗诊断系统；
自动处置级 (T3: Automated processing)	业务的全过程基本达到自动化处置、自动化流转等程度，需要“人工干预”。	1、政务服务办事“最多跑一次”；
系统支撑级 (T2: System support)	事件处理的主要过程实现系统/平台/工具的信息化支撑。	1、网格员巡查； 2、政务服务办事“综合窗口”；
人工处置级 (T1: manual processing)	事件处理的全过程主要靠人工执行。	

用数（赋能创新）：一个业务需求与数据产品不匹配的负向案例



(四) 一个负向案例：“好弓不拉满”，警示过度追求技术完美可能降低标准适用性，需保留迭代空间



(1) 审计发现：2023 年对区内高科技企业的补助发放有误。

(2 家企业已经注册地迁移，但是依然发放了 2023 年度的高科技企业补助)

追责

(3) 大数据中心暴怒：

(1) 数据治理的责任！(2) 类似的问题数据还有多少？



追责



(4) 数据治理公司回应：

- ① 问题数据是市建系统提供的；
- ② 我们已经反馈问题多次，市局系统不予理会；
- ③ 马上对出问题的数据集组织排查；



追责

(2) 财政反馈：“问题数据是大数据中心提供的。”

逆向：技术升级、数据利用，最终会推动业务变革

显示健康码每日得能量 5g

04-21 07:16:04

绿码

防疫行程卡
核酸检测
抗原检测

其它服务

最多跑一次

免申即享

提高惠企政策落实效率

靶向 监管

政策
直达

DD打车的案例：数据价值化 ≠ 数据交易

DD打车表面是互联网出行平台，但其核心竞争力实为数据资产驱动的用户黏性。2021年7月，DD因数据安全问题被国家网信办处罚80.26亿元并暂停新用户注册。此后半年间，市场涌现出超100家打车应用，但DD解禁后市场份额仍维持绝对优势：截至2023年，DD占据中国网约车市场75.5%的交易额（GTV 1924亿元），2024年上半年市占率稳定在73%。

这一现象表明，DD的护城河并非单纯的平台流量，而是海量数据沉淀与精准用户运营能力。

数据佐证与模式解析

- #### 1、数据驱动的高黏性机制

DD通过12年运营积累的5.8亿用户行为数据（如出行时间、路线偏好、支付习惯等）构建了深度用户画像。基于此，平台实现了动态定价、智能派单、路线优化等功能，将司机接单效率提升30%，用户平均候车时间缩短至3分钟。这种“数据闭环”使DD能持续优化体验，形成用户依赖。例如，2024年DD用户单日人均使用时长达10.7分钟，启动频次6.6次，均居行业首位。
- #### 2、市场份额的“反脆弱性”

尽管受罚期间（2021-2022年）涌现大量竞争对手，但DD凭借数据生态壁垒迅速收复失地：2023年订单量同比增长24%，2024年Q3国内单量突破40亿单创历史新高。同期，曹操出行、T3出行等第二梯队平台虽接入高德、百度等聚合流量，但过度依赖外部入口（曹操73.2%订单来自聚合平台），佣金成本激增，难以撼动DD的端到端服务闭环。
- #### 3、技术赋能的护城河

DD的“数据大脑”整合了AI调度算法、实时路况预测和司机信用体系，使平台能动态匹配供需。例如，其智能调度系统将车辆空驶率从传统出租车的40%降至15%，司机收入提升25%。此外，DD通过用户评价系统（日均处理超千万条反馈）持续优化服务标准，形成“数据-体验-留存”正循环。

核心结论

DD的商业模式表面是共享经济平台，本质是数据驱动的出行服务运营商。其成功源于将用户行为数据转化为运营效率与黏性壁垒，即使面临监管冲击和竞争分流，仍能通过数据资产的不可替代性维持市场统治地位。





有用（明线）：高质量的数据供给很关键

(一) 质量治理（2次加工3次检测）：从原始数据资源到最终数据产品，存在2次加工、3次检测

原始数据

问题导向：6性

治理前的数据质量

清洗融合前的数据质量

用六性来评价

为精准化的质量增强

无多
错少
重乱
缺错
孤旧

→

治理后数据

结果导向：5级（可信度）

治理后的数据质量

支撑场景应用的数据质量

用五级来评价

为精细化的使用管控

清洗融合

- 清洗
- 融合
- 标注

项目	详细描述	达标	待提升	不达标
1级	数据完整性：数据源头的完整性、数据链路的完整性、数据内容的完整性	✓	✓	✓
2级	数据准确性：数据内容的准确性、数据链路的准确性、数据内容的准确性	✓	✓	✓
3级	数据及时性：数据更新的及时性、数据链路的及时性、数据内容的及时性	✗	✓	✓
4级	数据可用性：数据内容的可用性、数据链路的可用性、数据内容的可用性	✗	✗	✓
5级	数据安全：数据内容的安全性、数据链路的可靠性、数据内容的完整性	✗	✓	✗

3. 评测分级 《高质量数据治理及运营白皮书(2025)》

数据质量评估分为A、B、C、D、E共5个等级，每个等级的描述及分值范围。

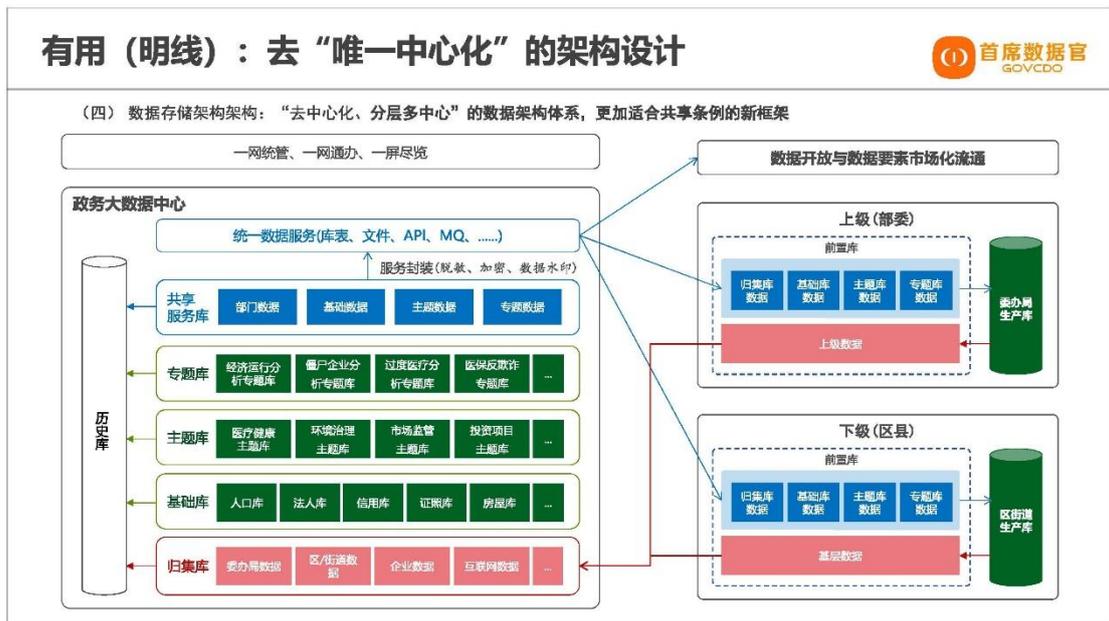
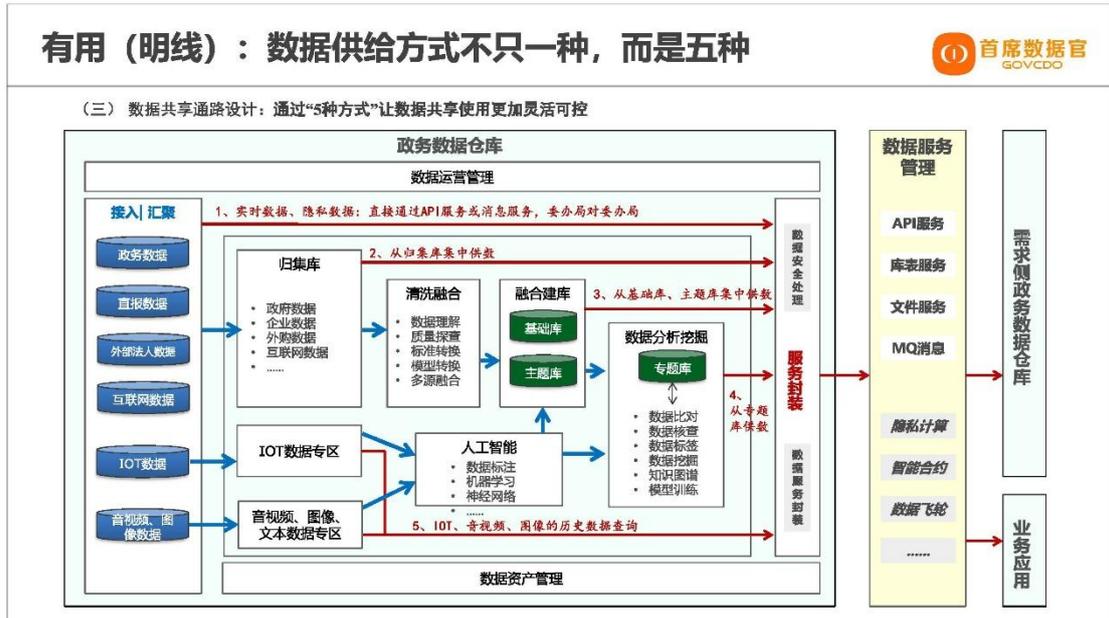
表 2-5 质量评级表

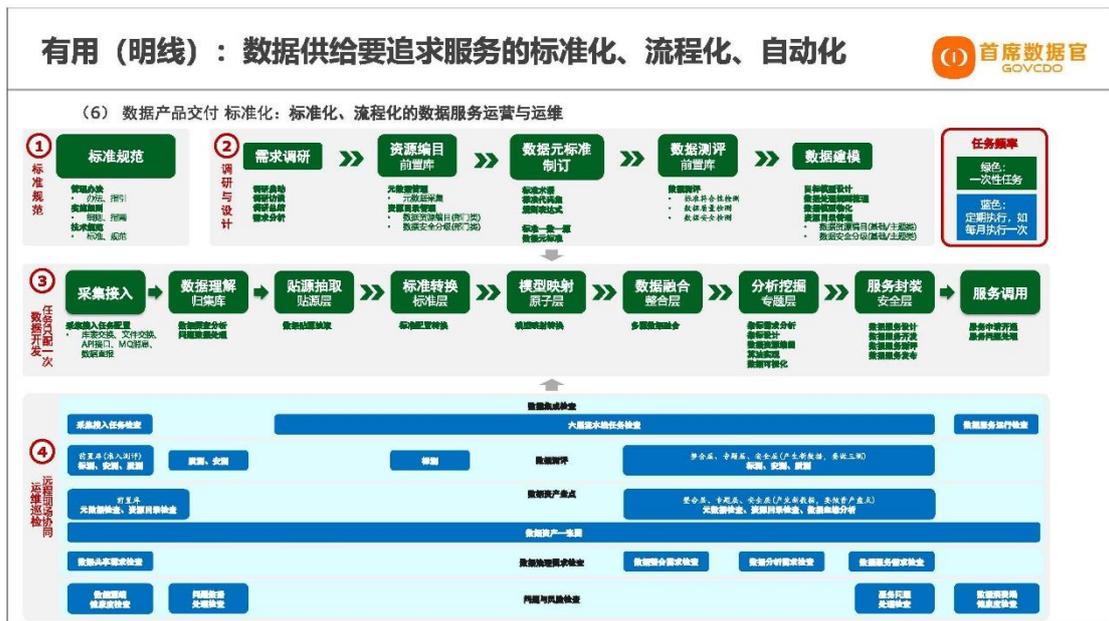
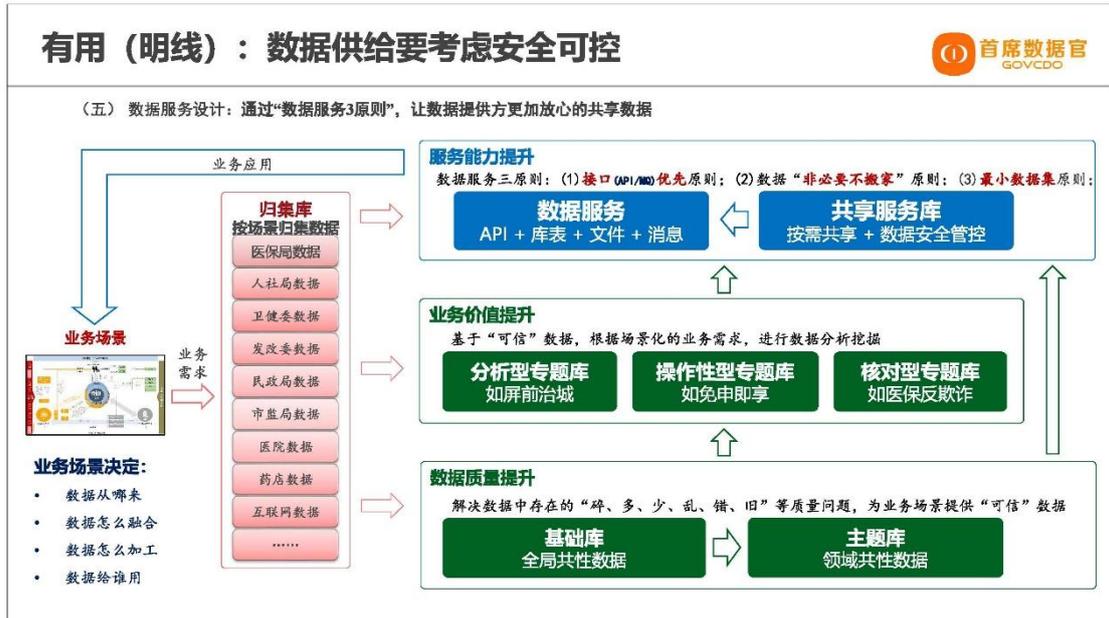
等级	数据质量等级描述及可用情况	分值和等级
A	数据质量高，可直接使用	80-100
B	数据质量较高，经简单处理后可使用	60-80
C	数据质量存在数据问题，经简单处理后可使用	40-60
D	数据质量存在一般问题，经处理后数据质量较差	20-40
E	数据质量存在重大问题，需从源头进行数据整改	0-20

有用（明线）：不同的场景需求决定数据的供给方式

(二) 数据使用场景梳理：政府内的数据需求有9类，而这些数据可能分布在不同部门、不同层级、不同系统、不同网络、不同领域

供给侧	资源类型	数据供需场景	需求方式	需求侧	
政府 委办局 上级单位 下级单位 法人单位 企业 事业单位 社团 互联网站	数据资源 政务数据 直报数据 法人单位数据 互联网数据 IOT数据 音视频、图像数据	1、物联网IOT数据。 举例：防台风时(应急指挥)旅游大巴的实时位置 2、个体隐私数据。 举例：社保缴纳、个税缴纳、疾病诊疗等个人隐私数据 3、电子文件数据。 举例：企业政务服务办事时，查询企业的电子营业执照 4、基础信息数据。 举例：如个人全景画像、企业全景画像等应用，调取基础库数据 5、业务主题数据。 举例：疫情防控场景下，调取居家隔离人员名单、疫情排查任务执行清单	要数据 自己加工	数据赋能业务场景	
		1、统计分析。 举例：进行城市人口总量、企业总量、城市事件总量等群体性统计 2、风险预警。 举例：复工场景下，获取疑似僵尸、疑似跑路、疑似外迁、劳资纠纷预警等风险预警企业 3、关系图谱。 举例：政法委需要戒毒人员的社会关系、家庭成员、固定资产、住房、工作经历等数据，加强对特殊人员的监管 4、智能识别。 举例：公安局通过人脸识别技术，将部署在机场、高铁站、体育场馆等重点场所采集的图像进行智能比对分析，识别在逃人员			要结果 自己展示







03 治数：数据易用

数据易用能力：指在特定的利用环境中，特定主体能够以互操作性、可关联性和提升用户体验为特定目标利用数据的能力。

易用（暗线）：知识与规则，可以提升服务供给水平



(一) 知识沉淀、规则固化

完整性 **准确性**
唯一性 **规范性**
时效性 **一致性**

来源于数据标准和规范

规则名称	规则描述	规则类别	示例
合同签订日期	YYYYMMDD	规范性	符合年月日的自定义规范
民族数据元值域	C?	规范性	民族字码值为代码，且存储长度超过2

来源于业务逻辑和经验

规则名称	规则描述	规则类别	示例
身份证规则检查	身份证号码规则校验	规范性	1. 号码长度非15或18位 2. 数字号码和行政区划不符
字符枚举判断	字符枚举值超出“A, B, D, AB”之中	规范性	血型字段中枚举为C
关键业务键空值校验	统一社会信用代码、执法证件、合同编号等信息为空	完整性	法人信息表中统一社会信用代码为空
身份证号重复	身份证号重复	唯一性	死亡信息表中人尸体化信息，身份证号重复，火化时间不一致
医保补助中预缴证已注册	医保相关补助信息中，对比预缴证号，按预缴证号办理注册信息	准确性	医保补助申请表预缴证号为CA0000，该预缴证号无效，且日期晚于证件注册日期
男女方证件号相同	出生、婚姻等场景夫妻双方证件号应不一致	准确性	出生信息表中，父亲身份证号与母亲身份证号相同



标准化转换：把来自各部门的数据转换成标准数据元字段

源字段 (如身份证号)

目标字段 (如身份证号)

业务属性

- 业务名称
- 业务类型
- 业务用途
- 业务规则

技术属性

- 数据类型
- 数据长度
- 数据类型
- 数据类型

管理属性

- 数据来源
- 数据更新
- 数据维护

1. 数据源的格式规范、字段的命名规范
 2. 数据源字段的格式规范、和数据规范
 3. 数据源格式自动转换
 4. 数据源的格式规范(如身份证号: "18" 格式: "18")
 5. 数据源的格式规范(如身份证号: "18" 格式: "18")
 6. 数据源的格式规范(如身份证号: "18" 格式: "18")

易用（暗线）：新技术和新领域，可以提升服务供给水平



（二）面向Data for AI的企业高质量数据集开发框架

基于企业数据资产管理业务需求，通过数据采集、主数据、元数据、数据标准、数据质量、数据安全、数据资产管理、数据服务、数据可视化等全流程治理，实现数据价值提升。



传统数据治理一般指结构化数据治理，非结构化数据如何治理、高质量数据集如何建设，是热点之一。

传统数据治理普遍实现了**支撑系统化、任务配置化、过程协同的系统流程化**，但**知识库、产业智能程度不足**，当前业内普遍在关注**大模型、agent与数据治理的结合**，“AI for Data”成为热点。

AI+背景下
数据资产管理内涵的新变化

数据编织 (Data Fabric)、数据网格 (Data Mesh)、本体论、知识图谱、业务智能等技术、理念与数据资产管理的实践如何结合，越来越受关注。

传统数据治理主要服务于“策、管、服”等场景，典型的决策应用是**统计报表等数据提供**，AI背景下，决策支撑从领导层关注的“宏观决策支撑”走向执行层**智能推荐和智能决策**，“Data for AI”成为热点。

易用（暗线）：高质量数据集可以提升服务供给水平



（三）高质量数据集

价值

随着人工智能技术向纵深演进，高质量数据集的研究正涌现**“以少样本推动高质量发展”**的前沿趋势。
 • 少样本学习依托先验知识、元学习及迁移学习等技术框架，深度挖掘数据的潜在特征与内在关联，显著降低模型对海量数据的依赖。
 • 即使在样本量有限的条件下，该模式仍可实现模型的高效训练与精准推理，缓解了数据采集与标注的成本压力及资源消耗，将数据利用的深度与效率提升至新高度。
 这一趋势助力突破传统数据驱动模式**对大规模样本的路径依赖**，使高质量数据集在资源受限场景中持续发挥核心支撑作用。

“算力焦虑”
“参数焦虑”
“数据焦虑”
“内容焦虑”

主要特征	目标与内容
高质量	<ul style="list-style-type: none"> 价值维度：准确性、完备性、唯一性、一致性、真实性、时效性、精度、代表性等 使用维度：针对数据存储格式、编码形式、封装方式强化用户支持，例如在建设数据集时，需考虑数据库软件、数据恢复软件等辅助工具。 管理维度：从数据脱敏、数据安全、网络安全等角度考虑，如保密性、依从性、可追溯性等
强合规	<ul style="list-style-type: none"> 权属合规性：来源合法/合规授权/合规登记 内容合规性：无涉及危害国家安全、法律法规禁止、歧视、垄断和不正当竞争、个人合法权益侵犯 等中毒数据)
主题性	<ul style="list-style-type: none"> 以对象为核心，以程序员、建筑物等对象为主题整合多源多模态数据集 以场景为核心，以安全生产、精准营销、反欺诈等场景为主题整合多源多模态数据集
效用性	能够为各种活动和决策提供的支持与帮助，具体可细化为易用性、可理解性、可再利用性
可信性	数据可追溯，流程透明，多主体平等参与和协同联动生态
复合性	类型覆盖结构化数据、文本、视频、音频、图形图像、知识库等多元形式，涵盖政府、企业、个人等不同主体数据
市场性	作为数据产品，具有重要的经济社会价值，属于竞争性产品，可估值定价，可通过数据交易所等平台实现数据要素市场化流通

高质量数据集的内涵与外延：质量与价值逻辑统一



内涵：高质量、AI 外延：高价值、业务

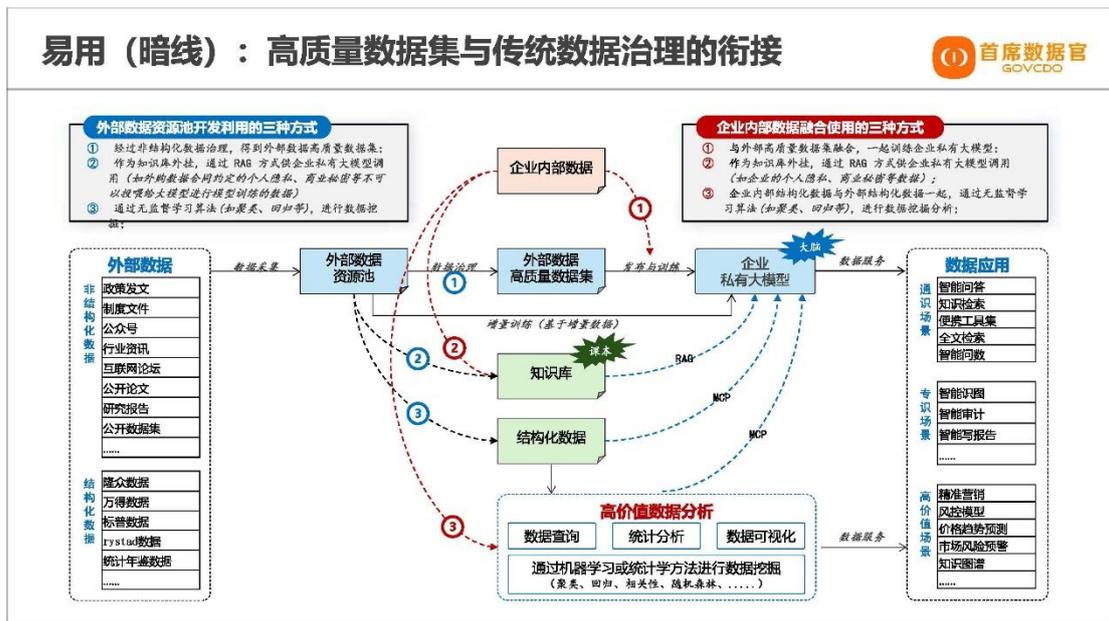
强调“开发和训练AI人工智能模型”应用 强调“业务应用的价值”

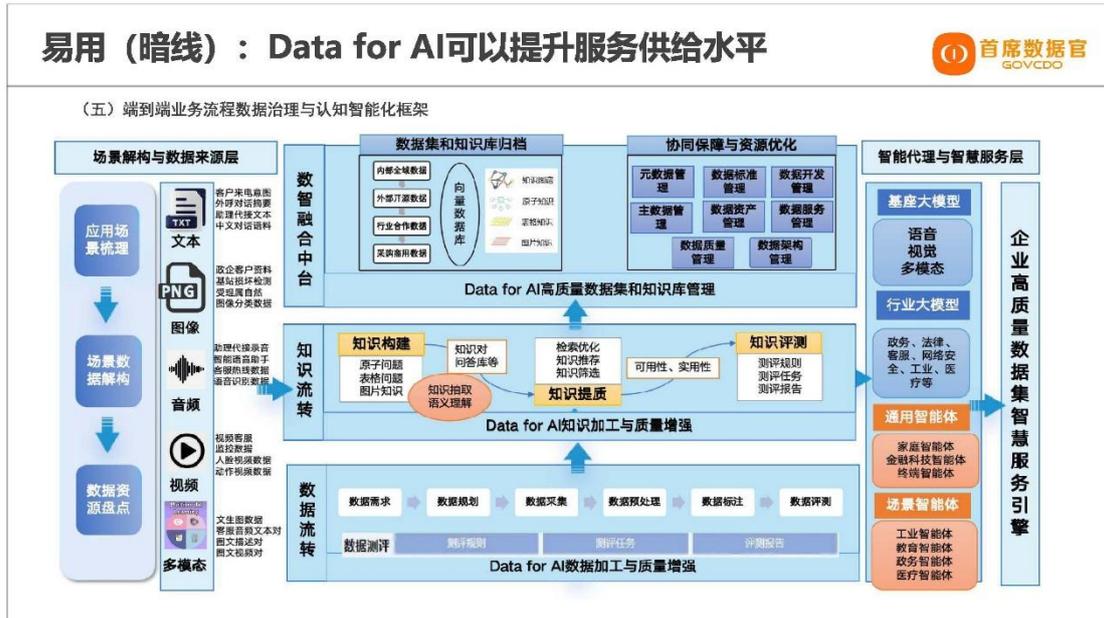


https://mp.weixin.qq.com/s/wGQ_doWE4XkPIUFWYO0Pqg



https://mp.weixin.qq.com/s/6Dvq3NIXopP-Nood1mg7Mw?scene=1&click_id=16





善用：伦理、偏见、歧视、隐私、恐暴、区别对待、不当使用等问题 

村镇银行“取款难”多名储户被赋红码 郑州12345：大数据出问题

更新新闻客户端 2022-06-14 18:30

天目新闻记者 于璐 李沐子

6月13日，有多名网友在社交平台反馈称，前往郑州沟通村镇银行“取款难”的储户被赋“红码”。

随即，天目新闻记者联系到其中一位储户唐女士，唐女士表示自己一直在安徽，老家也没有疫情，但是通过扫描储户维权群里发的郑州高铁站的二维码后，自己也被赋予了红码，“但是我的家人扫这个二维码就是正常的绿码。”唐女士告诉记者，由于他们“取款难”一事一直没有进展，所以部分储户商量要在6月13日到河南郑州当地去维权，“有些储户在12日这天到达郑州，然后就发生了被赋红码的事情。”



06-13 12:39:16 06-13 11:57:57 06-13 11:42:59

06-13 12:07:29 06-13 11:31:33

2022年06月13日 20:41:40

2022年6月13日 20:37:59

存在异常

善用（底线）：安全与合规也是大模型的底线 

《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》

“数据二十条”

二、建立保障权益、合规使用的数据产权制度

(三) 探索数据产权结构性分置制度。

(四) 推进实施公共数据确权授权机制。

(五) 推动建立企业数据确权授权机制。

(六) 建立健全个人信息数据确权授权机制。

(七) 建立健全数据要素各参与方合法权益保护制度。

序号	业务对象	违规内容
产权合规	1 来源合法	从违法违规渠道获得的数据
	2 授权合规	公共/企业/个人数据无合法合规的授权协议
	3 登记合规	数据无合规的产权登记证明

《生成式人工智能服务管理暂行办法》

第四条 提供和使用生成式人工智能服务，应当遵守**法律、行政法规**，尊重**社会公德和伦理道德**，遵守以下规定：

(一) 坚持社会主义核心价值观，不得生成煽动颠覆国家政权、推翻社会主义制度，危害国家安全和利益，损害国家形象，煽动分裂国家、破坏国家统一和社会稳定，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情，以及**虚假有害信息**等**法律、行政法规禁止的内容**；

(二) 在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、性别、地域、性别、年龄、职业、健康等**歧视**；

(三) 尊重**知识产权、商业道德，保守商业秘密**，不得利用算法、数据、平台等优势，实施**垄断和不正当竞争行为**；

(四) 尊重**他人合法权益**，不得危害他人身心健康，不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益；

(五) 基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。

序号	业务对象	违规内容
内容合规	1 国家安全	涉危害国家安全的内容
	2 公共秩序	涉法律法规禁止的内容 涉歧视的内容
	3 企业	涉垄断和不正当竞争的内容
	4 个人	涉侵犯个人合法权益的内容

监管算法：大数据伦理管控与审计

对各业务部门的大数据应用，进行大数据伦理审计，覆盖大数据杀熟、信息茧房、精准欺诈、信息诱导、不良排行榜、个人隐私侵犯、特定人群排斥、垄断竞争、差别待遇等有违法法规、社会公平、公序良俗的情况

数据的无界应用
有界
灰度边界

	商用	不当商用	违规	违法	
	精准营销	精准杀熟	精准内容	精准欺诈	
	推荐商品	推荐服务	推荐内容 信息茧房	推荐受害者	
	民生	公益	宗教	政治	
	精准服务	精准扶贫	精准传教	精准竞选	



监管数据：数据操作行为监管与审计

对DBA等乙方人员，进行数据操作行为的审计，覆盖数据的增删改查、导出、展示等操作行为

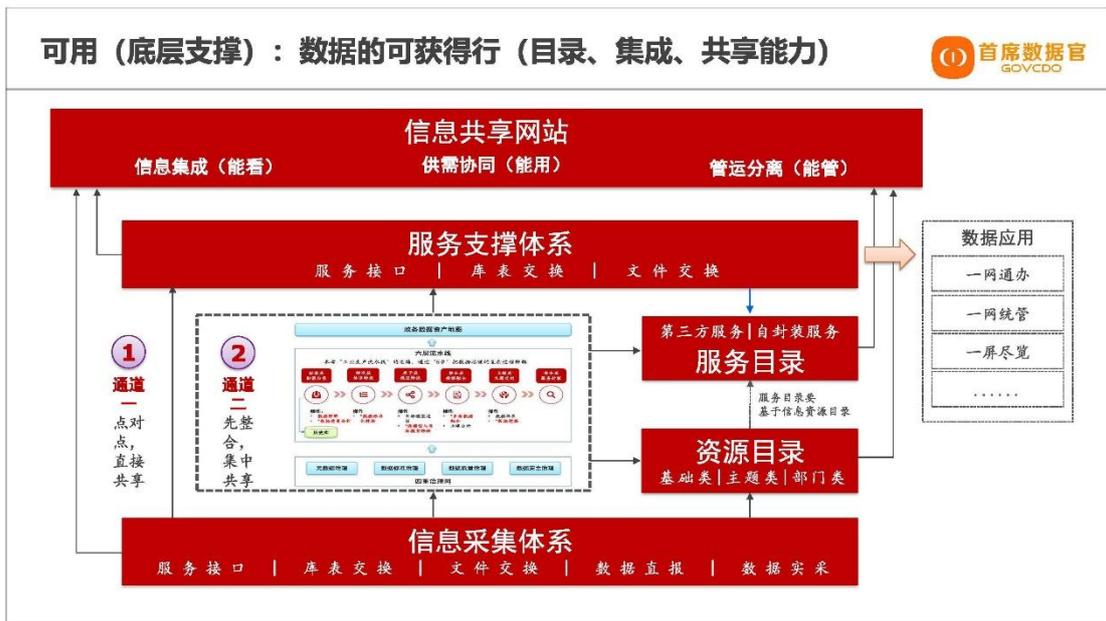
查询	搜索	修改
删除	导出	展示

数据出境安全评估制度配套法规金字塔



用于监视并记录对数据库服务器的各类操作行为，并提供详细的行为日志。通过对网络数据的分析，实时地、智能地解析对数据库服务器的各种操作，并记入审计数据库中以便日后进行查询、分析、过滤，实现对目标数据库系统的用户操作的监控和审计。具备强大的风险发现能力，实时监控平台系统自身数据库及第三方库的操作记录，并对此操作进行风险评估与告警，为数据库安全提供全面的防护。审计内容包括数据库、表、字段、用户、操作时间、风险类型、风险级别、行为日志、操作热度、性能指标等。





可用（底层支撑）：数据基础设施 首席数据官 GOVCDO

中华人民共和国中央人民政府 www.gov.cn

国家发展和改革委员会 工业和信息化部 关于印发《国家数据基础设施建设指引》的通知

发改数据〔2024〕1853号

发布日期：2025年01月01日

四、总体架构

(一) 技术架构

图1 数据基础设施及网络、算力设施总体架构图

可用（底层支撑）：数据治理保障体系 首席数据官 GOVCDO

数据治理“铁三角”：组织 × 制度 × 沟通

组织
搭台子、定角色（谁来做）
组织虚设→没人干活

制度
立规矩、定方圆（怎么做）
制度缺失→无章可循

沟通
通信息、聚合力（如何协同）
沟通不畅→各自为战

治理体系

终极目标：有效处置→格式化→规范化→管好数据全生命周期

可用（底层支撑）：数据生态与数字文化



数据治理培训与文化培育

数据治理培训
(基础认知建设)

数据文化培训
(文化理念渗透)

双训协同实施
(落地保障)

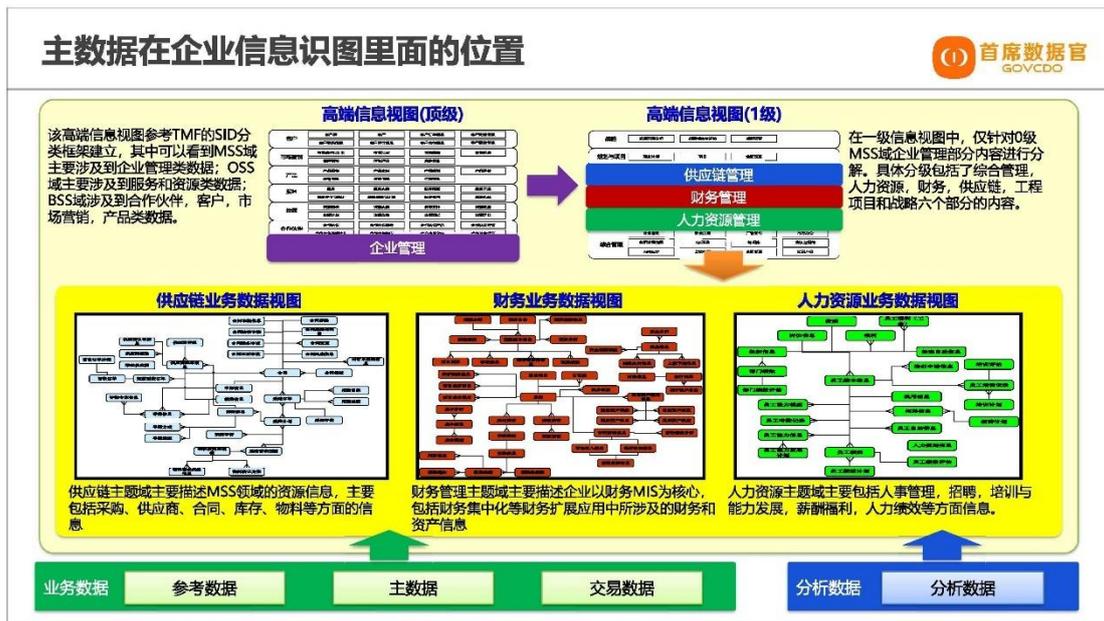
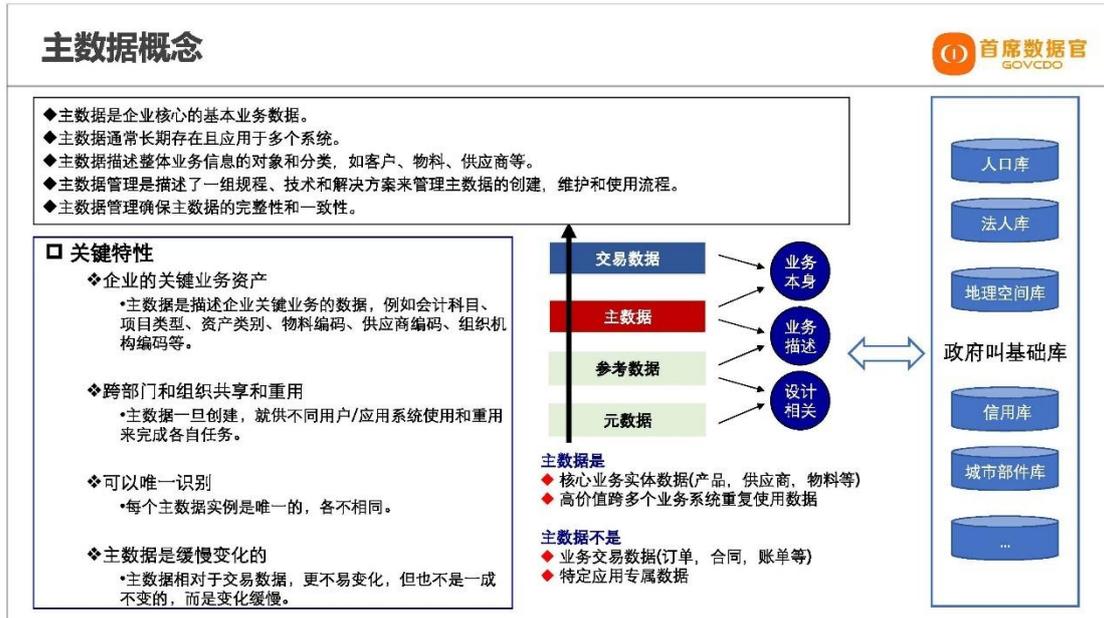
双训协同关系
(共生引擎)

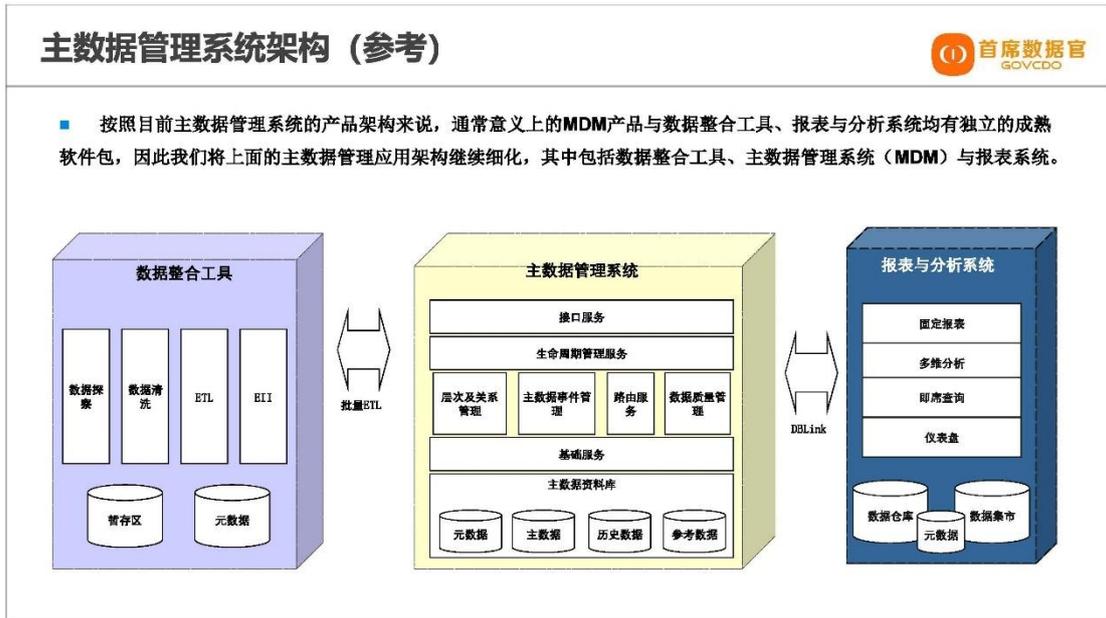
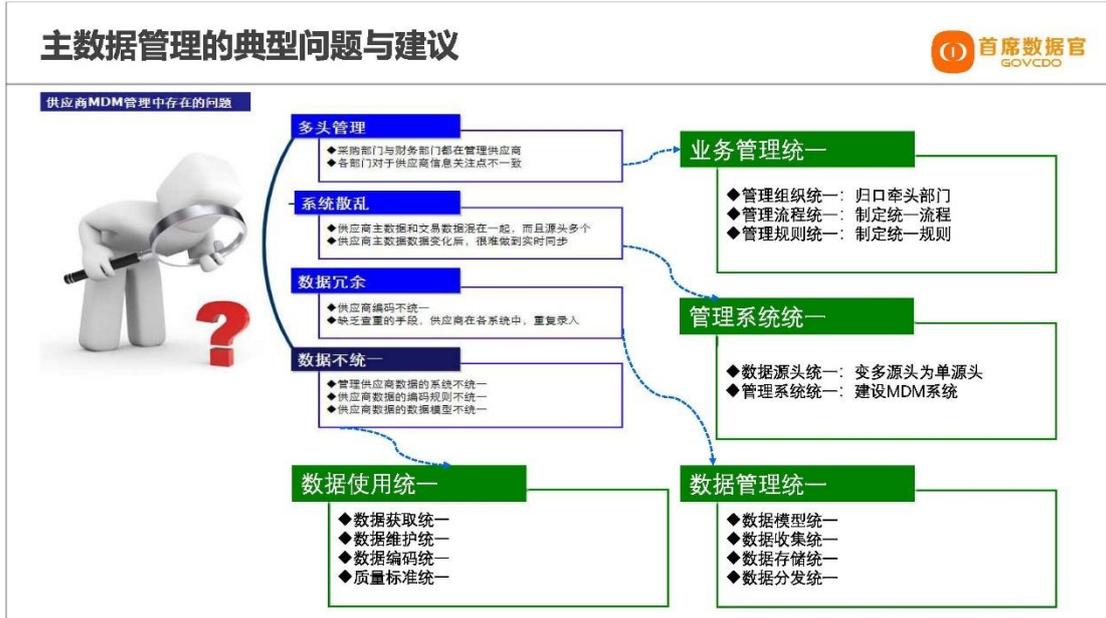


硬核定规筑基，软体铸魂融脉，双螺旋催生数据基因

02 主数据

以一个企业主数据项目为案例





一个主数据系统升级的案例：项目建设背景



领导高度重视 + 实际业务需求，是新建/改造主数据管理系统的驱动力

集团公司领导高度重视数据治理工作，多次做出明确部署，要求加快数字化转型，加快标准体系建设，全力推动数据治理，完善数据管理体系，打通数据价值链，主数据贯穿与业务的主线或资源，跨业务和部门被重复使用，是信息系统互联互通的基石，更是数据治理的坚实基础。

董事长：要进一步强化创新意识，积极应对时代挑战，把握产业变革大势，着力推动技术创新和管理商业模式创新，加快数字化转型和智能化发展步伐，围绕数字化转型关键技术攻关，加快数据标准体系建设，积极探索数字化转型智能化发展新模式。

——集团公司2022年工作会议主题报告

总经理：要全力推动数据治理，完善集团公司数据治理体系，建立健全数据标准，实现底数摸清、账上统一，同时持续推进内部数据融合共享，打通数据价值链，形成规范体系，保证数据可信、完整、准确，有效服务业务发展。

——集团公司2022年工作会议生产经营报告

数据治理专项工作要求

治理工作的重要组成部分，随着数据治理工作的开展，新制定的主数据标准、新识别主数据的安全管理以及更多的主数据共享需求将会对主数据管理系统的技术支撑能力提出更高的要求。数据治理工作的开展也为主动的主数据管理带来机会，基础数据平台需要提前做好技术支持准备。

经营管理数据治理工作要求

《关于成立集团公司经营管理数据治理专项工作组的通知》
中国能源科信，2023，8号

为推动集团公司经营管理数据实现源头唯一、质量可靠、协同共享，推进各业务域的数据认领、确权 and 源头治理工作。

技术支持组负责经营管理数据资产管理，开展数据资产目录管理、数据架构管理、数据标准管理、数据质量核查以及数据需求分析等技术支持工作；并负责经营管理基础数据平台建设和技术保障工作。

主动的主数据管理



建设思路



主数据平台新建/升级的三大任务：数据治理、技术升级、国产替换

主数据管理系统建设重点方向是继承已有成果，吸收成熟技术方案，充分考虑现有主数据管理现状、需求和数据治理工作的需要，建设安全自主可控的主数据管理系统替代现有主数据管理平台，实现主数据标准、主数据的纵向贯通、横向协同。

主数据管理系统未来定位

- 支撑集团及所属单位全域主数据管理
- 为业务系统提供权威、一致的主数据服务
- 实现自主可控

主数据管理系统建设方向

1. **选用新型架构，实现弹性部署**，满足纵向贯通、横向协同的主数据管理，满足多源异构接入需要，满足服务需求不断扩展的需要
2. **补齐功能**，满足主数据安全管理和统一共享的需要，满足进一步提升数据质量的需要
3. **硬件、应用软件、数据库软件国产化**，满足自主可控的需要。



主要功能 2：主数据生命周期管理



建设主数据生命周期管理用于管理主数据在系统中的活动，从申请创建开始，到归档删除结束。现有主数据管理平台已初步形成主数据的全生命周期管理，但在各个环节在功能细节上需要重点加强。



数据生命周期重点内容

主数据创建	主数据校验	主数据审核	主数据查询	主数据同步	主数据归档
<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid #ccc; padding: 2px;">自动分类</div> <div style="border: 1px solid #ccc; padding: 2px;">权威数据</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <div style="border: 1px solid #ccc; padding: 2px;">大批量数据</div> <div style="border: 1px solid #ccc; padding: 2px;">页面设计</div> </div> <ul style="list-style-type: none"> • 利用自动分类，引用权威数据简化数据填报 • 利用灵活的页面定义，提高用户体验 • 提升大批量数据导入的性能 	<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid #ccc; padding: 2px;">校验规则</div> <div style="border: 1px solid #ccc; padding: 2px;">权威数据</div> </div> <div style="border: 1px solid #ccc; padding: 2px; text-align: center; margin-top: 5px;">相似校验</div> <ul style="list-style-type: none"> • 利用相似校验，提高校验准确度 • 利用权威数据完成数据校验 • 丰富校验规则，形成校验规则库 	<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid #ccc; padding: 2px;">智能审核</div> <div style="border: 1px solid #ccc; padding: 2px;">审核规则</div> </div> <div style="border: 1px solid #ccc; padding: 2px; text-align: center; margin-top: 5px;">大批量审核</div> <ul style="list-style-type: none"> • 提升大批量数据审核的效率和便利性 • 利用审核规则、AI引擎，智能审核 	<div style="border: 1px solid #ccc; padding: 2px; text-align: center; margin-bottom: 5px;">智能查询</div> <div style="border: 1px solid #ccc; padding: 2px; text-align: center;">查询效率</div> <ul style="list-style-type: none"> • 提供智能查询，提高查询的效率和便利性 	<div style="border: 1px solid #ccc; padding: 2px; text-align: center; margin-bottom: 5px;">数据服务</div> <div style="border: 1px solid #ccc; padding: 2px; text-align: center;">数据同步</div> <ul style="list-style-type: none"> • 利用数据服务提高数据共享的便利性，降低集成成本； • 提高数据同步效率； 	<div style="border: 1px solid #ccc; padding: 2px; text-align: center; margin-bottom: 5px;">归档公示</div> <div style="border: 1px solid #ccc; padding: 2px; text-align: center;">数据归档</div> <ul style="list-style-type: none"> • 数据归档前发布公示，征求使用单位的反馈 • 根据公示结果开展归档，并同步

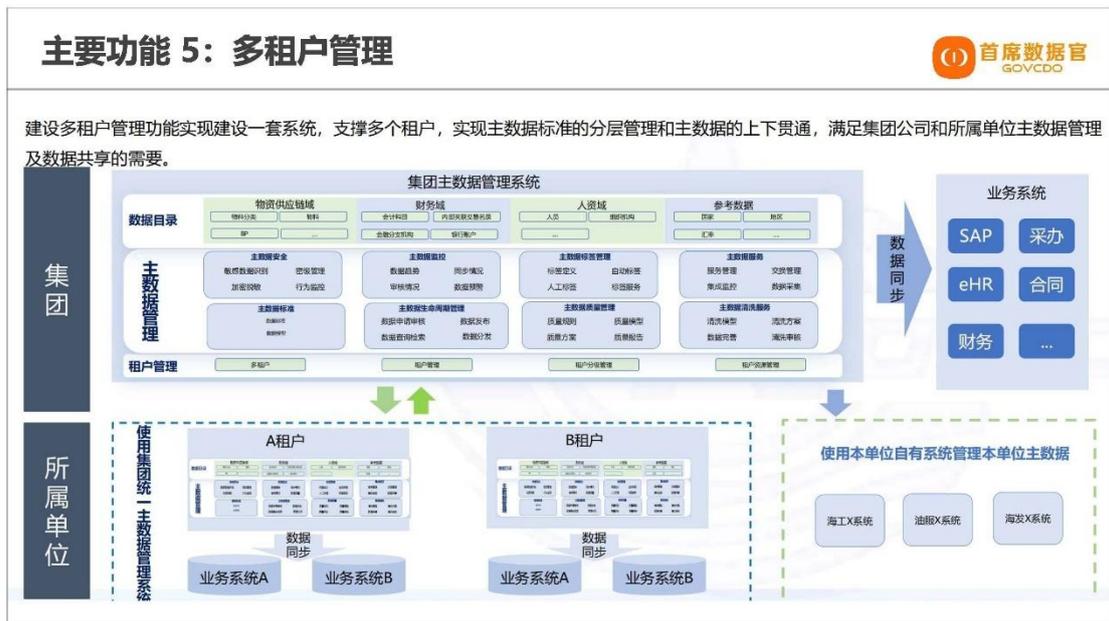
主要功能 3：主数据质量管理、清洗管理及标签管理

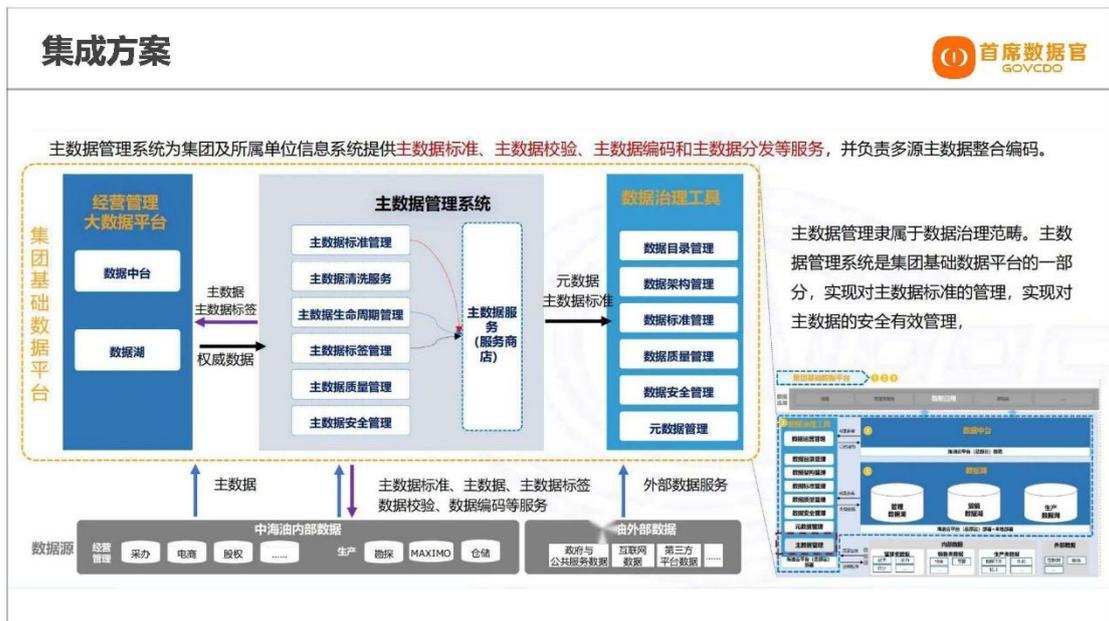


建设主数据质量管理功能支持根据业务管理需要和质量标准要求开展事前和事后检查，并出具质量报告，为质量评价提供支撑。建设主数据清洗功能支持根据质量标准和业务规则对数据进行清洗加工。建设主数据标签管理功能，对主数据进行灵活标记，满足业务需要。



主数据质量管理	主数据清洗管理	主数据标签
<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid #ccc; padding: 2px;">质量模型</div> <div style="border: 1px solid #ccc; padding: 2px;">质量规划</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <div style="border: 1px solid #ccc; padding: 2px;">质量规则</div> <div style="border: 1px solid #ccc; padding: 2px;">质量报告</div> </div> <ul style="list-style-type: none"> • 提供对于内外部主数据的质量稽核能力 • 跟踪业务、标准的变化，主动开展数据质量稽核，形成质量报告 • 为质量评价提供支撑，推动数据治理工作 	<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid #ccc; padding: 2px;">清洗模型</div> <div style="border: 1px solid #ccc; padding: 2px;">清洗方案</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <div style="border: 1px solid #ccc; padding: 2px;">清洗审核</div> <div style="border: 1px solid #ccc; padding: 2px;">清洗结果</div> </div> <ul style="list-style-type: none"> • 提供对于内外部主数据的清洗能力 • 跟踪业务、标准的变化，根据质量报告，发起主数据清洗 	<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid #ccc; padding: 2px;">标签定义</div> <div style="border: 1px solid #ccc; padding: 2px;">标签规则</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <div style="border: 1px solid #ccc; padding: 2px;">自动标签</div> <div style="border: 1px solid #ccc; padding: 2px;">人工标签</div> </div> <ul style="list-style-type: none"> • 非主数据的自然属性，为满足统计分析或业务应用需要对主数据进行标记（例如：物料是否是集采物料，供应商是否是制裁单位等） • 实现标签的自主定义 • 利用规则、算法自动打标签





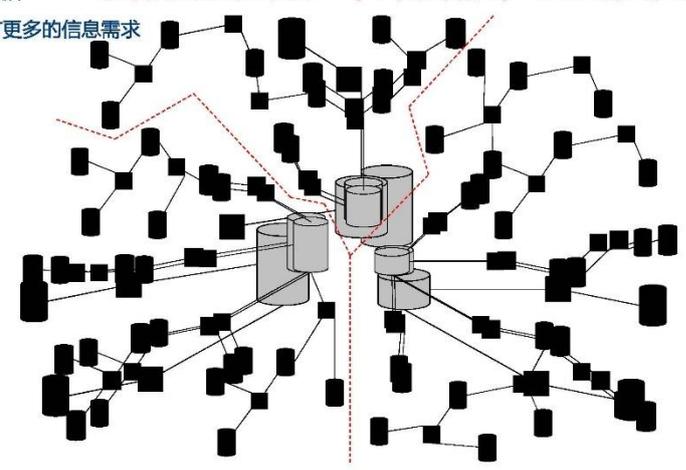




信息化系统越来越多，数据越来越乱

最初，信息化系统很简单
然而，一段时间以后
然后，我们会有更多的信息需求

我们的系统在不断膨胀，直到变成一张蜘蛛网



首席数据官 GOVCDO

数据质量问题是怎么产生的？

数据质量问题的必然性：条块分割的业务现状，必然导致数据质量的“多、少、乱、错、旧”

1、数据获取难

条块分割 → 导致 → 各自为政 → 导致 → 烟囱林立 / 信息孤岛

2、数据使用难

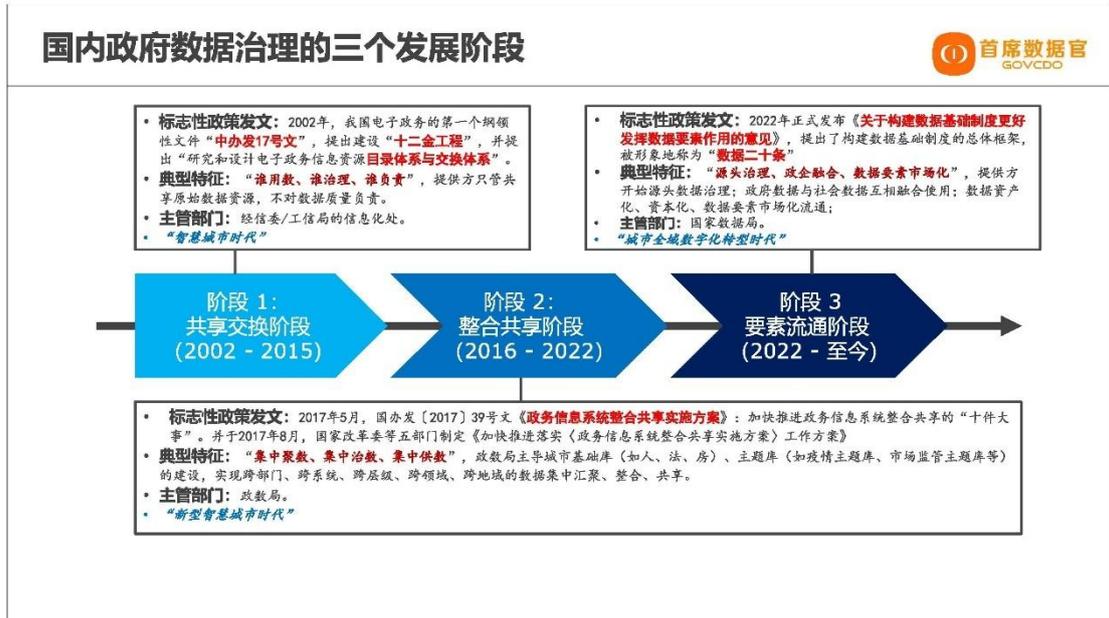
各自为政 → 导致 → 碎 / 数据碎片化问题 → 导致 → 数据不一致 → 表现为

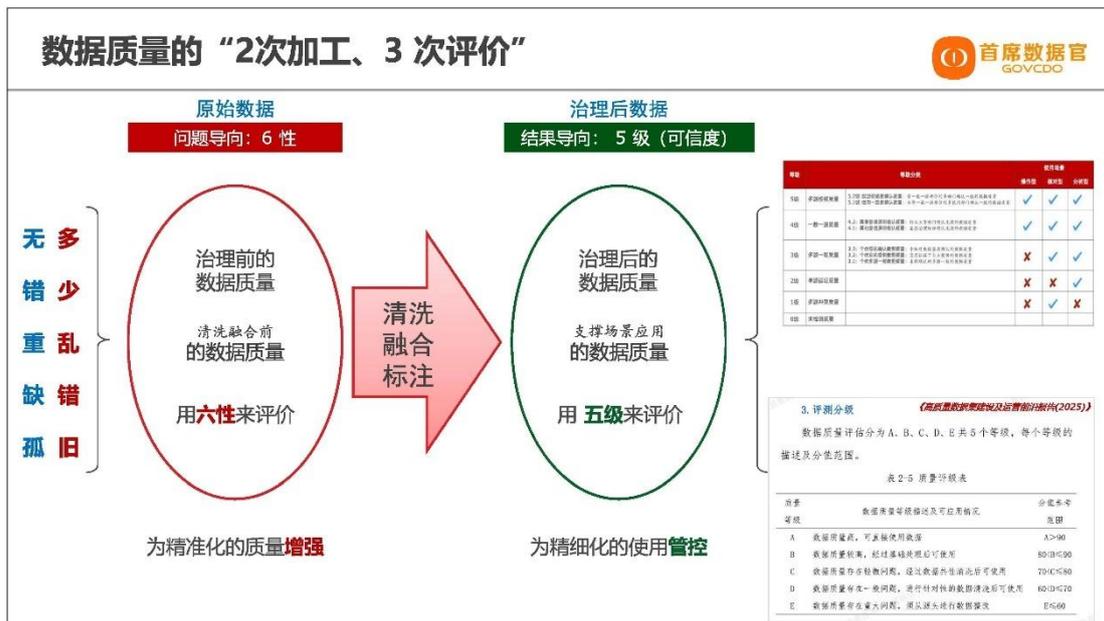
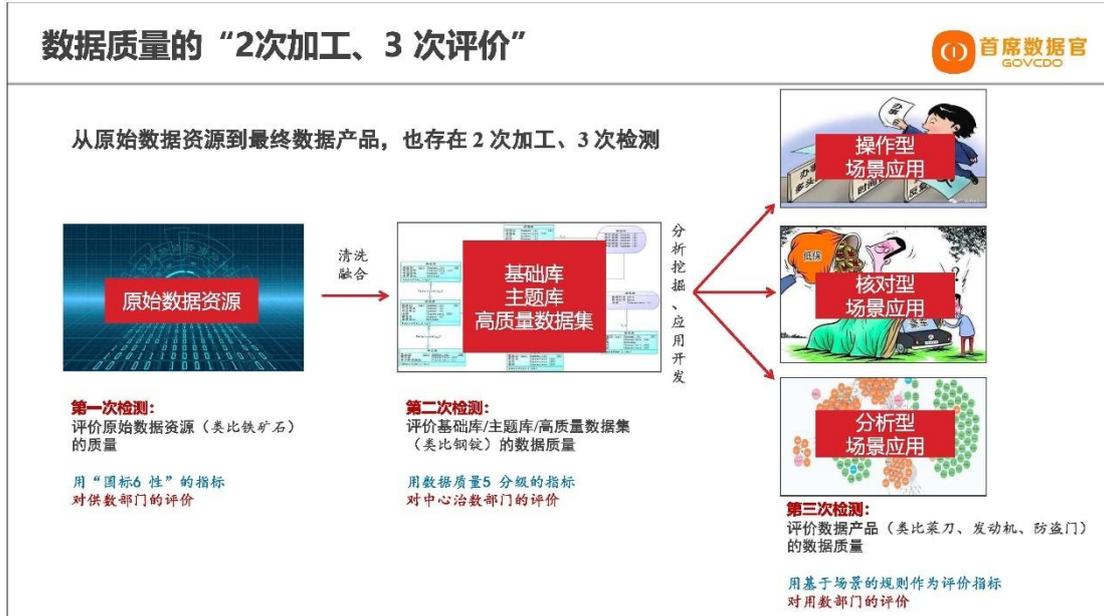
- 多：数据重复性问题
- 少：数据完全性问题
- 乱：数据混乱问题
- 错：数据冲突问题
- 旧：数据时序问题

如何质量增强？

首席数据官 GOVCDO





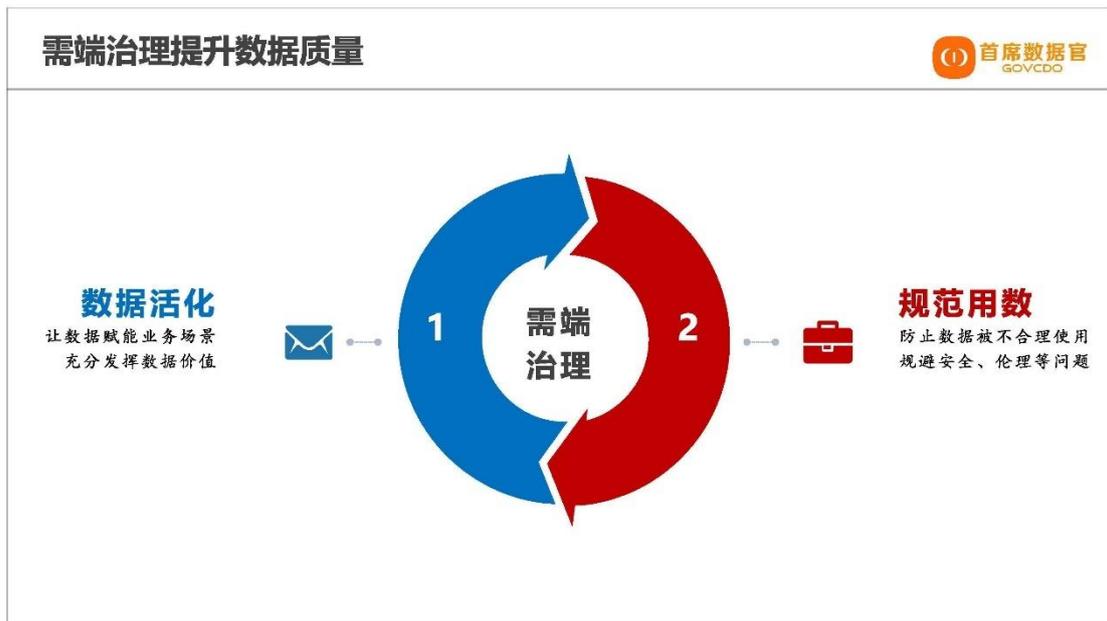


如何提升数据质量

目录

CONTENTS

- 01 需端治理**
数据使用方如何提升数据质量
- 02 中心治理**
数据管理方如何提升数据质量
- 03 源头治理**
数据提供方如何提升数据质量
- 04 能力提升**
数据质量治理能力有几个方面



用 AI 能力赋能业务举例：（1）事件智能发现



视频图像AI识别：通过AI人工智能，分析城市的各类视频与图像数据，及时发现城市治理中安全、环保、卫生、交通、城管、市监、水等各类典型事件，提升事件的感知、研判、分析能力。

手工流转
管理效率低

人力成本高
工作量大

定时巡查
发现不及时

算法：25+AI算法能力场景

准确率：平均准确率达90%以上

部署地点：苏州工业园区、江西宜春市、石家庄新华区等

价值：人力节省50%，城市管理效率和市民幸福感大幅提升

- 视频+AI，实现事件自动检测和预警
- 视频+AI，全维度自动核查、结案

用 AI 能力赋能业务举例：（2）事件智能分拨



数据挖掘&神经网络：“自动分拨+推荐分拨+人工分拨”结合，提升分拨效率与准确率

痛点

- 1.分拨员需掌握**40+**部门三定方案，**1000+**权责清单/事项清单，**2000+**法律法规
- 2.疑难工单处置周期长，识别滞后，热线中心事后被动管理

方案

实时工单 → **AI模型推理**

- 疑难工单识别
- 自动派单
- 相似案例推荐

特事、大事

→

人工分拨

非疑难工单
置信度 > 0.95

→

自动分拨

非疑难工单
置信度 > 0.85

→

推荐分拨

基于AI模型迭代，让更多人工派单转为机器自动派单

智治增效实践案例

广州市白云区热线中心

Before

人工派单

- 平均每月受理工单**2W+**
- 需**8名**派单员
- 平均每单用时**4分30秒**

After

2021年6月试运行热线智能派单方案

- 6月累计受理**22468**单
- 机器人累计派单**18119**单
- 占受理总工单的**80%**
- 机器派单准确率**88%**
- 平均每单用时**1分15秒**
- 可减少**6名**派单员

价值

- 1.机器人秒级派单，准确率**88%**以上，可完成**80%**的派单工作，节省**80%**人力
- 2.疑难工单提前识别，指挥中心提前干预，降低疑难工单处置时长，提高满意度

用 AI 能力赋能业务举例：（3）城市热点智能研判



通过NLP与神经网络技术，分析城市运行的“高发热点、突发热点、潜在热点”等热点态势，让政府第一时间了解老百姓最急、最盼、最忧、最怨的问题，识别潜在风险，关键预警

热点(高发热点&突发热点&潜在热点)研判

态势研判

- 事件总量统计
- 处置效能分析
- 热点趋势分析
- 发生时段分析
- 热点主体标签
- 热点主题标签
- 热点地址标签
- 一人多诉标签
- 多人同诉标签
- 久而未决标签
- 潜在热点标签
- 突发热点标签
- 高发热点标签
- 群体高发标签
- ...

热点挖掘

主体热点
高发/突发/潜在

① **法(法人)**: 如企业、学校、商场等

② **基(基层)**: 如街道办事处、居委会、物业等

③ **人(人群)**: 如特殊人群、快递员、保安等

主题热点
高发/突发/潜在

如健康码赋红码问题

地址热点
高发/突发/潜在

如街道、社区、小区、楼栋、核酸检测亭等

事件主题	被投诉主	处置单位	处置情况
社区深圳北站	投诉疾病...	深圳北站...	— 结案: 250 件; 正常未结案: 120 件
社区樟坑一区	求助疾病...	樟坑社区...	— 结案: 201 件; 正常未结案: 22 件
社区富联二区	梅林多利产业园、泰源社区服务中心、粤省事健康码	—	结案: 228 件; 正常未结案: 4 件;
社区民乐一区	求助疾病...	梅林多利...	— 结案: 160 件; 正常未结案: 45 件
社区沙博小区	求助疾病...	深圳市龙... 龙华区观...	结案: 135 件; 正常未结案: 1 件;
	求助疾病...	教育机构... 龙华区观...	结案: 141 件; 正常未结案: 1 件;

城市运行事件数据

↑

热点研判(NLP自然语言处理&神经网络)

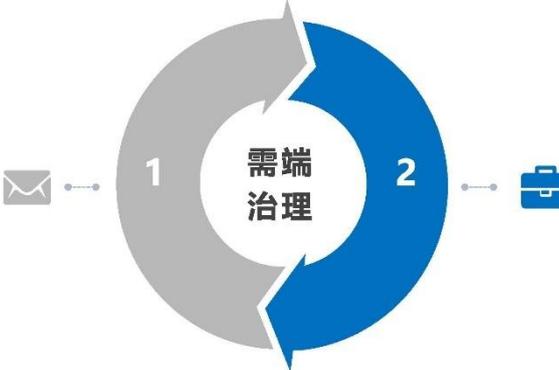
↑

需端治理提升数据质量



数据活化

让数据赋能业务场景
充分发挥数据价值



需端治理

规范用数

防止数据被不合理使用
规避安全、伦理等问题

规范用数：适应症/用药参考 vs 地表水功能 vs 数据适类使用



用药须知卡
珍菌（注射用特立帕肽）

适应症
绝经后妇女骨质疏松症的预防。

用法用量
本品需皮下注射给药，具体用法用量请参见说明书。如果患者不能耐受本品，应当暂停给药并密切观察其反应。

注意事项

1. 本品可能引起头痛、头晕、胃肠道不适感、恶心、注射部位一过性疼痛等反应。
2. 本品含有人体生长激素类似物成分，避免在儿童和青少年以及在成年男性使用本品。
3. 避免与含雌激素类药物同时使用。
4. 避免与含维生素D类药物同时使用。
5. 避免与含维生素K类药物同时使用。
6. 避免与含维生素E类药物同时使用。

药品保存
本品应在2-8℃冷藏保存，不得冷冻。

生活护理

1. 注意饮食卫生，避免暴饮、暴食，少饮辛辣及酒精类饮品。
2. 避免长时间暴露于阳光下，避免日光浴、桑拿、日光浴床等，保持皮肤清洁。
3. 避免吸烟、饮酒、咖啡、浓茶、辛辣、油腻、生冷、刺激性食物。
4. 避免剧烈运动，适当休息。

健康就是幸福
扫码了解更多药品

血液透析的适应症和禁忌症

图1 地表水功能划分

分类	地表水功能
I	源头水、国家自然保护区
II	饮用水水源地（一级）等
III	饮用水水源地（二级）、渔业、游泳
IV	一般工业用水、娱乐用水
V	农业用水、一般景观
劣V	丧失使用功能

水资源
饮用水功能
使用功能
污水
有害



如何提升数据质量

目录 CONTENTS



需端治理

数据使用方如何提升数据质量



中心治理

数据管理方如何提升数据质量



源头治理

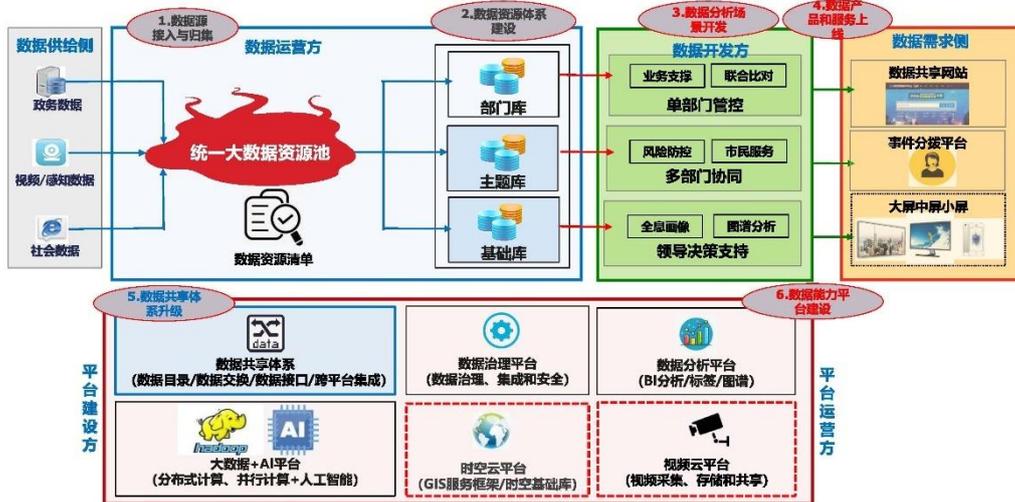
数据提供方如何提升数据质量



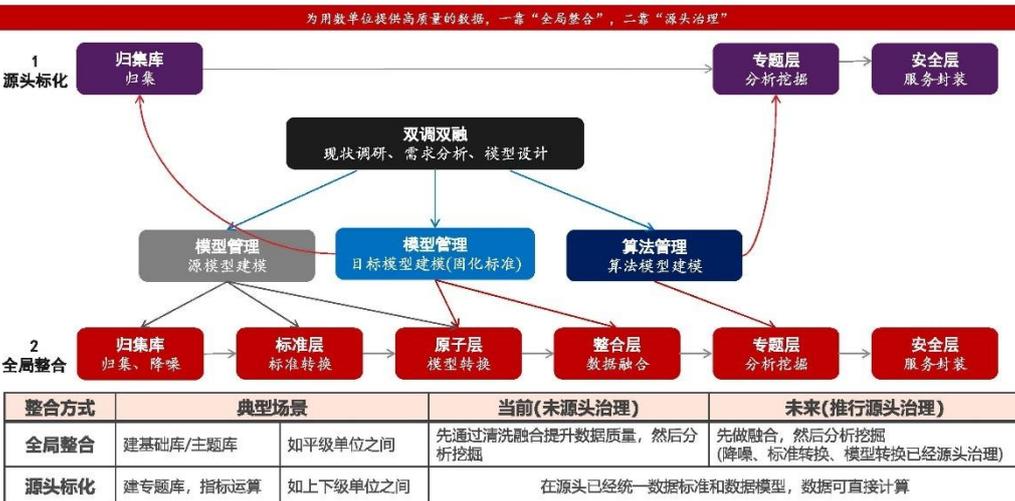
能力提升

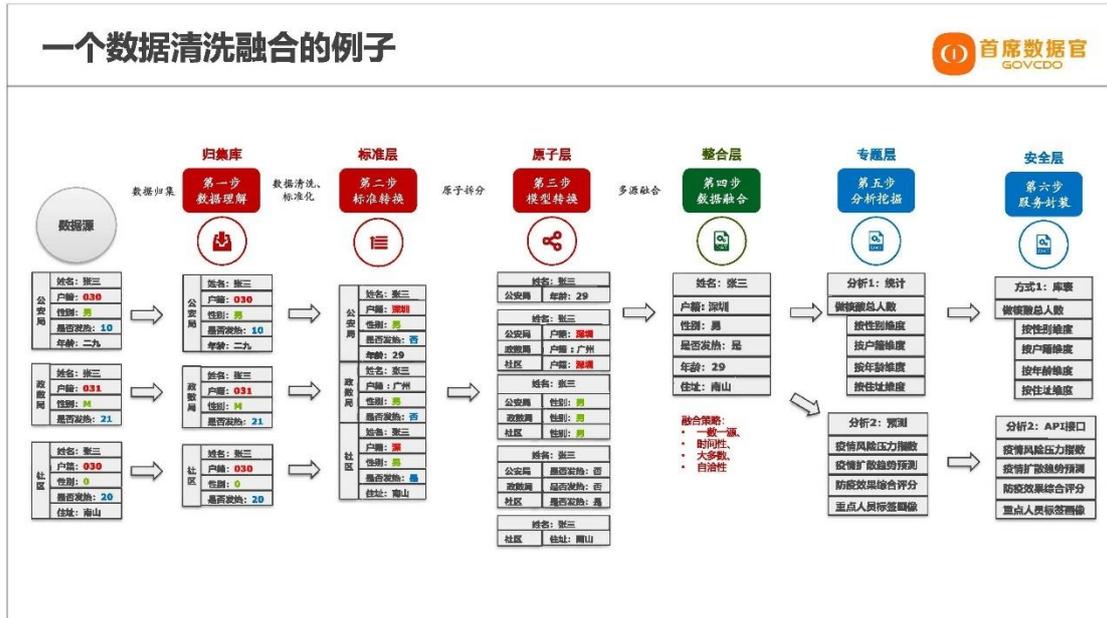
数据质量治理能力有四个方面

中心治理实现数据资源的统收统发、集中治理



中心化治理可减少重复投资、节约成本





源头治理有“系统新建”与“存续改造”两种情况



新建系统

系统建议：

- **模型驱动设计：**在系统设计阶段嵌入数据标准与主数据模型，通过业务建模（流程/产品/数据模型联动）统一业务语言
- **强制落标机制：**开发时自动校验数据标准合规性（如字段长度、格式、枚举值），拒绝非标数据入库
- **全链路治理工具集成：**采用一体化开发平台，将数据质量规则、元数据管理等功能嵌入开发测试流程

核心逻辑：

- **预防性治理：**通过前置数据模型设计（如五级建模体系），从源头规避数据孤岛与冗余
- **成本最优路径：**新建系统改造成本仅为存续系统的1/5（IBM研究），早期治理可节省后期70%整合成本

VS

存续系统

系统建议：

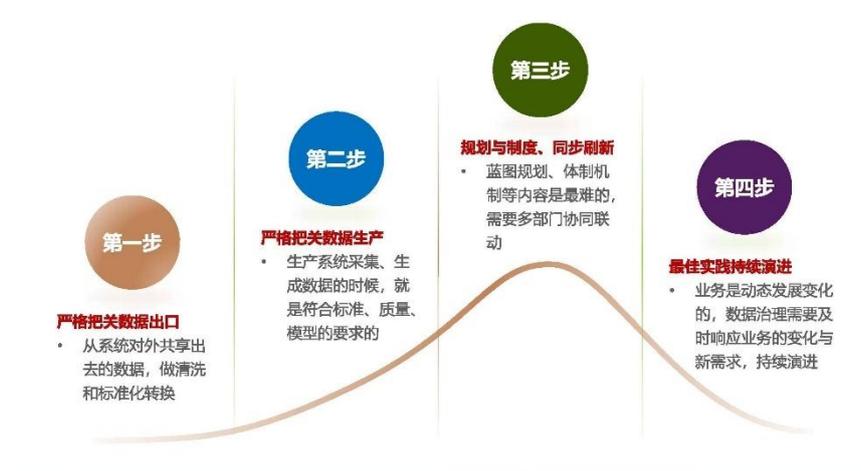
- **逆向工程治理：**通过数据资产盘点建立业务实体画像，反推源头系统改造优先级（如高频使用/高价值数据优先）
- **渐进式清洗：**建立增量数据质量阀（如ETL过程强制清洗），存量数据分批次修复（如每月修复10%异常数据）
- **架构解耦改造：**通过API网关实现数据服务化，逐步替换直连数据库的访问方式

核心逻辑：

- **风险可控改造：**某银行案例显示，通过API解耦可使系统改造宕机时间缩短83%
- **业务价值牵引：**优先治理支撑核心业务场景的数据（如某零售企业通过“实时战报大屏”倒逼供应链数据治理）

针对“存续系统”的数据源头治理的演进建议





第一步

严格把关数据出口

- 从系统对外共享出去的数据，做清洗和标准化转换

第二步

严格把关数据生产

- 生产系统采集、生成数据的时候，就是符合标准、质量、模型的要求的

第三步

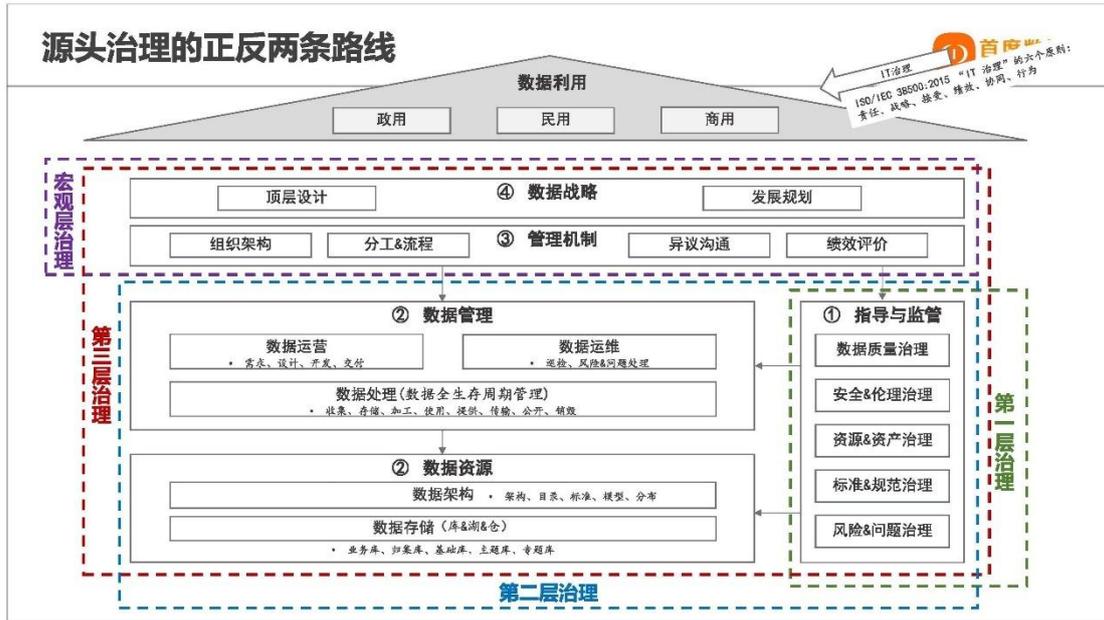
规划与制度、同步刷新

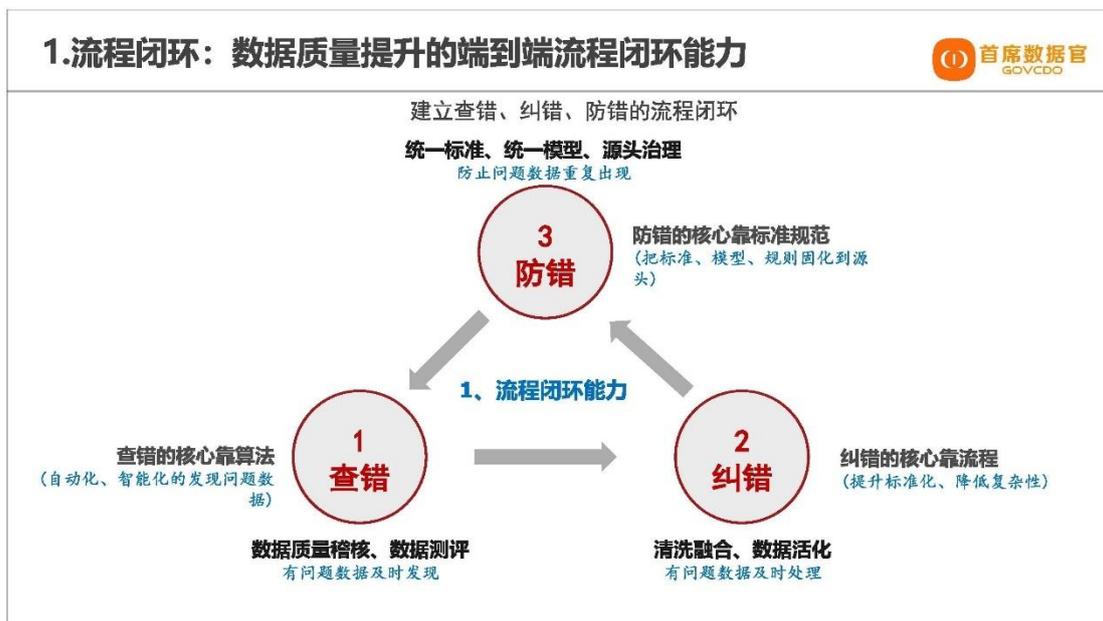
- 蓝图规划、体制机制等内容是最难的，需要多部门协同联动

第四步

最佳实践持续演进

- 业务是动态发展变化的，数据治理需要及时响应业务的变化与需求，持续演进





2.体制机制：数据质量管理能力的体制机制保障



参考深圳数据条例，结合首席数据官制度，从组织、制度、考核等方面构建能力



3.技术支撑：数据质量管理能力的技术支持保障



包括方法论、产品工具、服务运营等三个方面的能力

技术支撑能力



数据治理要怎么干

- 架构设计怎么干？
- 需求分析怎么干？
- 实施交付怎么干？



降低复杂性、提升标准化

- 如何衡量工具能力？
- 需要哪些平台工具？



满足需求、创造价值

- 需：数据供应链如何支撑好业务需求？
- 供：如何总结知识与经验，优化供给？
- 审：如何通过测评考核，推动数据大循环？

04 数据运营



数据运营的概念



从数据的 DIKW (认知深化) 与 DRAC (资产转化), 分别对应了数据运营的**对内数据服务运营**与**对外数据资产运营**



DIKW: 侧重数据价值的认知深化
(从原始数据到智慧决策)



DRAC: 侧重数据价值的资产转化
(从资源管理到资本增值)

数据服务运营就跟“牧羊”一样



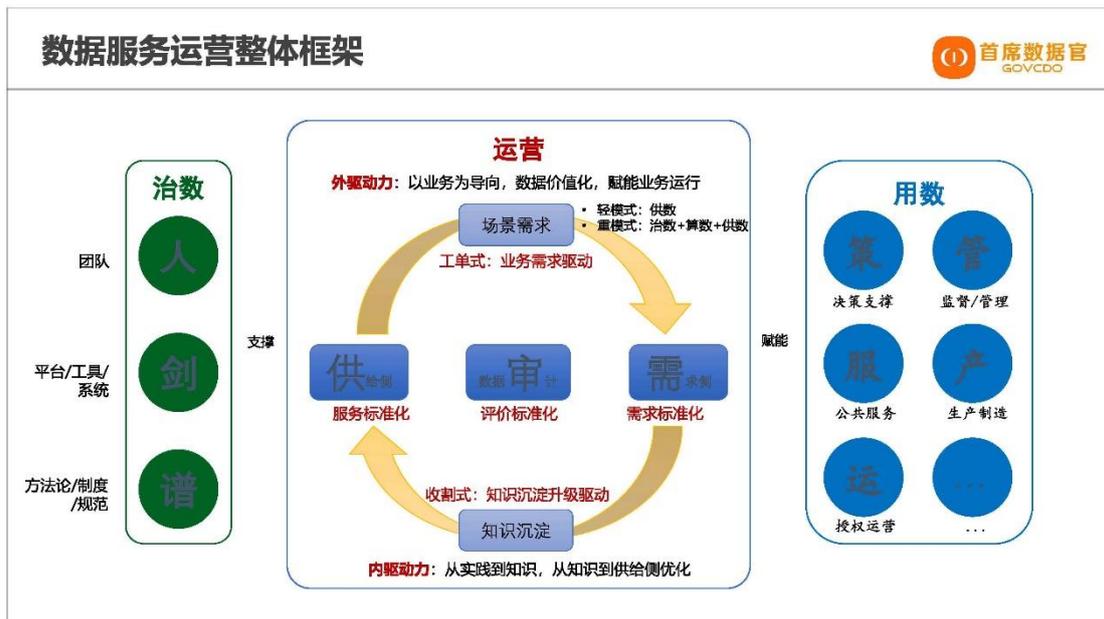
牧羊不是有了一群羊就行, 必须得放牧。羊儿得吃草、喝水、防疫疾病、预防丢失、.....

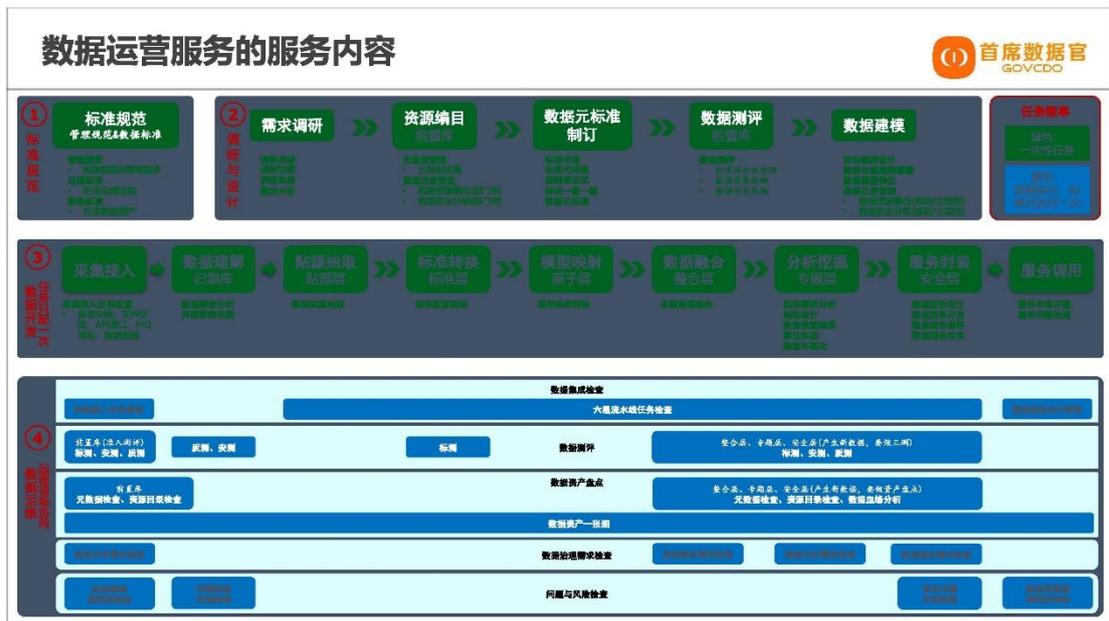
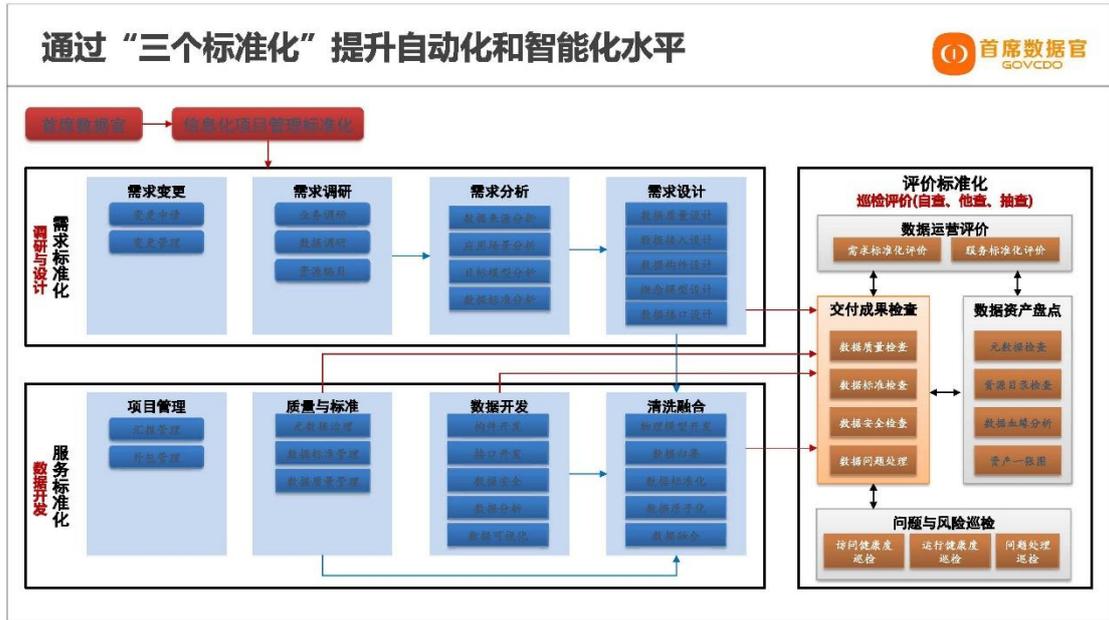


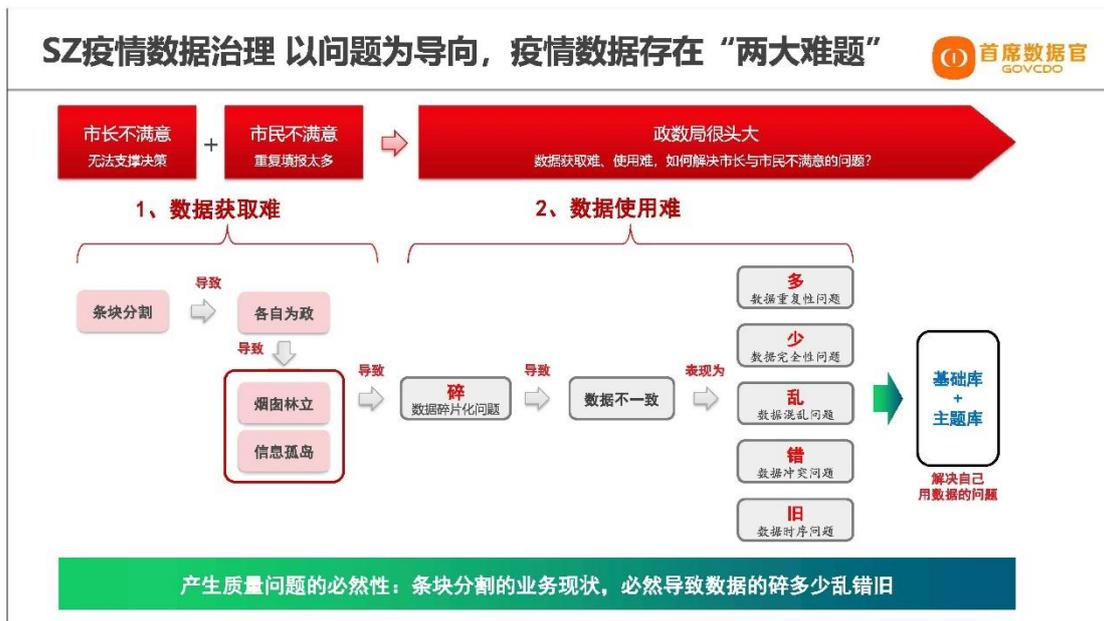
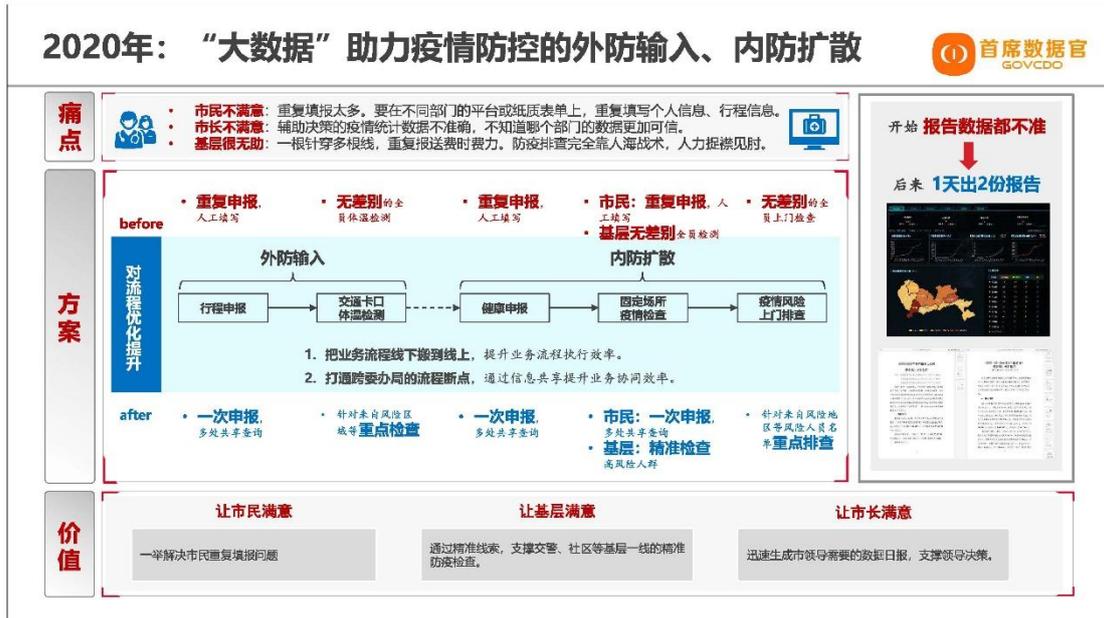
VS

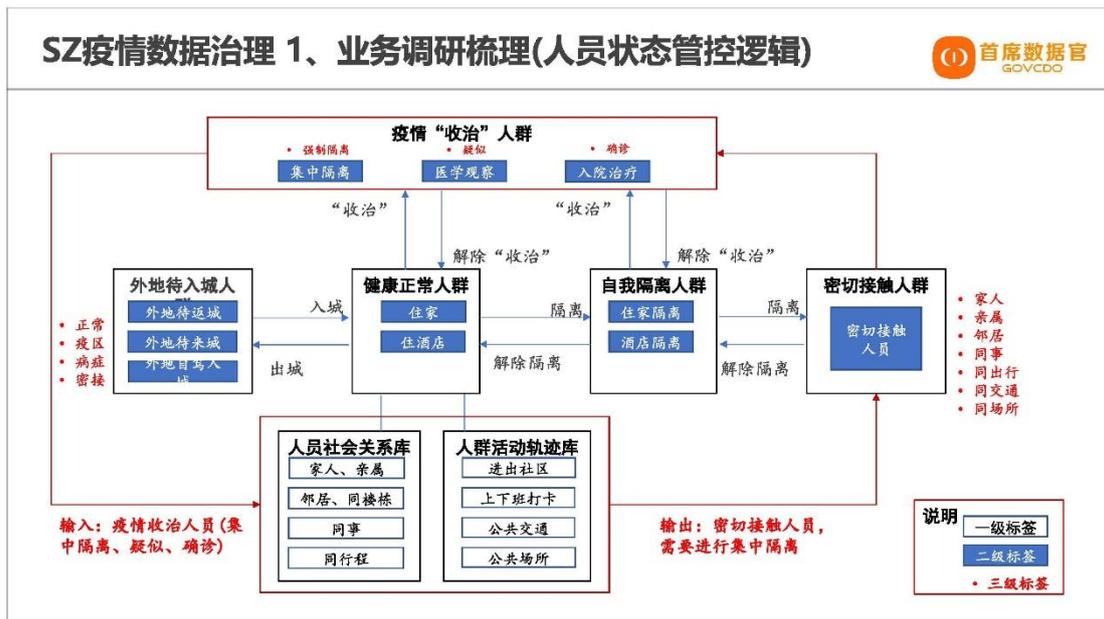
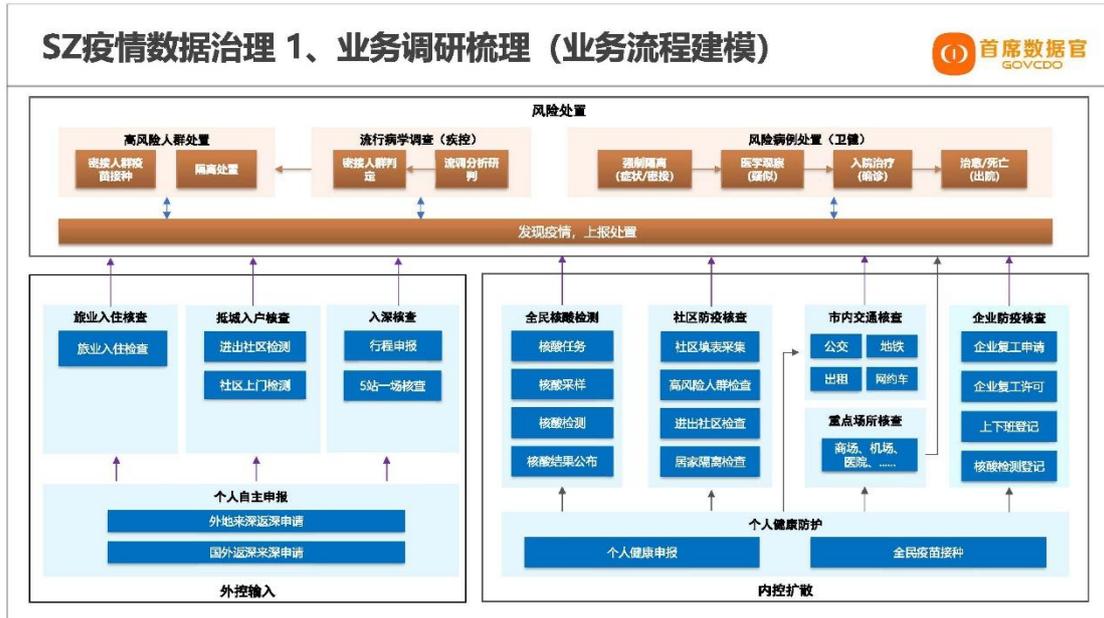


数字化不是建了一堆设备、工具就行, 必须得服务运营。数据要响应业务需求、要质量提升、要监管保护、.....









SZ疫情数据治理 2、数据调研梳理

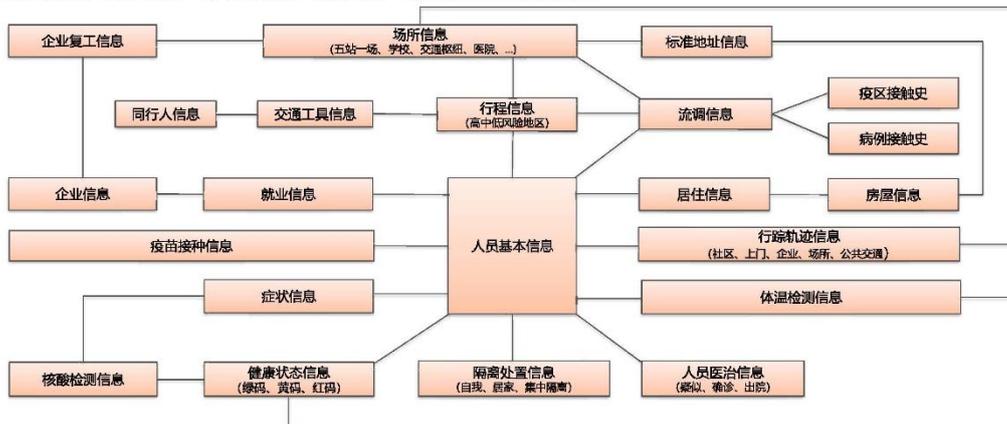


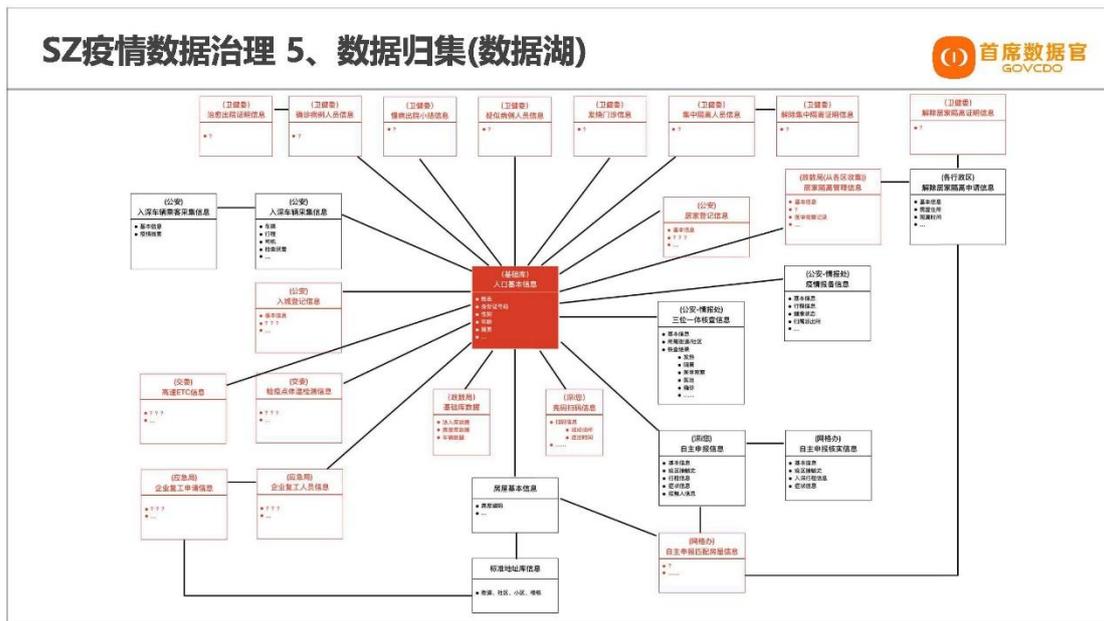
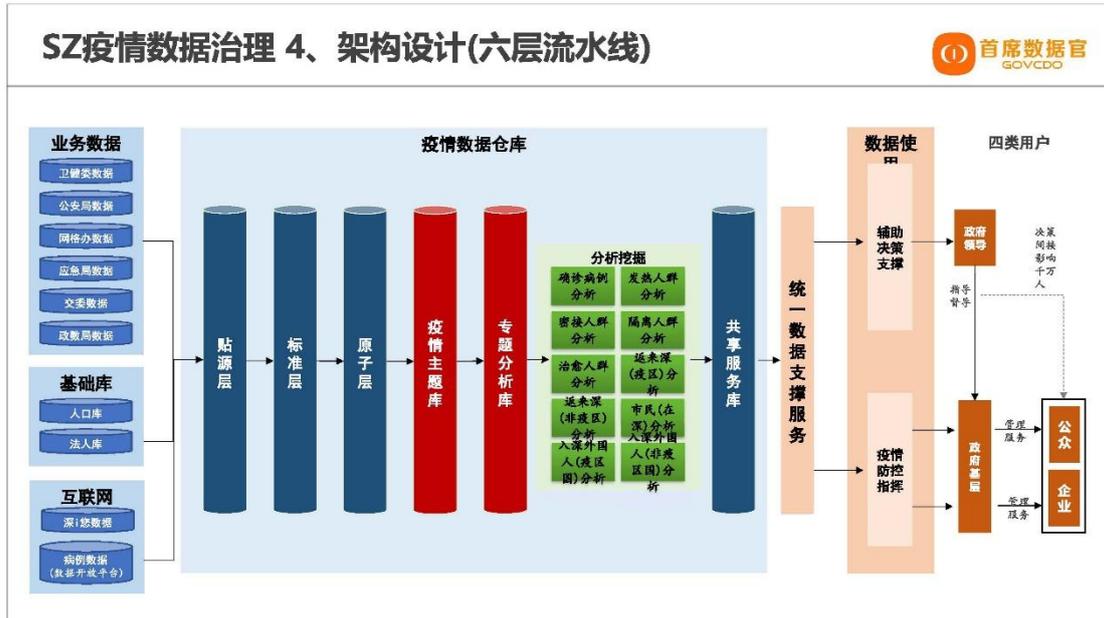
序号	部门	优先级	数据	接口表名	核心字段	是否已沟通	是否已承诺数据	是否已获取数据结构	是否已获取数据	数据量 (截止0217)	数据所在的网络环境
	数据管理部门								数据目录的维护		电子政务外网 政务网 专项 互联网接入区
1	卫健委	高	确诊病人入院/出院/转院信息 (第三人民医院) (卫健委公安, 公安给资源)		病别号, 姓名, 身份证号, 出院日期等	✓	✓	✓			
2		高	发热门诊信息		姓名, 身份证号, 门诊号, 就诊时间等	✓	✓				
3		高	集中隔离信息		姓名, 身份证号, 隔离点, 隔离原因等	✓	✓				
4		高	解除集中隔离证明信息(科主任让社工落实)		姓名, 身份证号, 集中隔离点, 解除隔离时间等	✓	✓				政务网
5	公安	高	解除居家隔离证明信息(科主任让社工落实)		姓名, 身份证号, 居家隔离点, 解除隔离时间等	✓	✓				政务网
6		高	确诊病例人员信息(卫健委公安, 公安给资源)		姓名, 身份证号, 确诊日期等	✓	✓				政务网
7		高	确诊病例人员信息(卫健委公安, 公安给资源)		姓名, 身份证号, 确诊日期等	✓	✓				政务网
8		高	入深车辆号牌信息(交警车牌照)与资源网的结构进行对比)	RKFR_GAJ_RSCD_INFO	车牌号, 司机, 行驶, 类型等	✓	✓			302,368	能不能共享交换可以拿到?
9	公安	高	入深车辆号牌信息(与资源网的结构进行对比)	RKFR_GAJ_RSCD_LIST	姓名, 身份证号, 行驶信息, 同行人员, 健康信息等	✓	✓			504,454	能不能共享交换可以拿到?
10		高	入深车辆号牌信息(与资源网的结构进行对比)	RKFR_GAJ_RSCD_LIST	姓名, 身份证号, 行驶信息, 同行人员, 健康信息等	✓	✓				
11		高	入深车辆号牌信息(与资源网的结构进行对比)	RKFR_GAJ_RSCD_LIST	姓名, 身份证号, 行驶信息, 同行人员, 健康信息等	✓	✓			3,250,352	
12		高	居家隔离信息(与资源网的结构进行对比)	RKFR_GAJ_GZJLJLJK	姓名, 身份证号, 住址, 健康状态等	✓	✓			2,273,471	
13	交警	高	疫情排查信息(与资源网的结构进行对比)	RKFR_GAJ_FYBQJLJK	姓名, 身份证号, 健康信息等	✓	✓			3,923,033	
14		高	三位一体排查信息	RKFR_FYBQJLJK	姓名, 身份证号, 健康信息等	✓	✓				
15		高	疫情排查信息(与资源网的结构进行对比)	RKFR_FYBQJLJK	姓名, 身份证号, 健康信息等	✓	✓				27 共享交换
16		高	健康码信息	RKFR_FYBQJLJK	姓名, 身份证号, 健康码, 健康码, 健康码, 健康码	✓	✓				
17	应急局	高	健康码ETC信息	RKFR_ETC_GS_DATA	车牌号, 通行时间等	✓	✓			6,876,701	共享交换
18		高	企业复工复产信息		姓名, 身份证号, 上下班打卡记录等	✓	✓				
19		高	企业复工复产信息		姓名, 身份证号, 上下班打卡记录等	✓	✓				
20	版数局	高	居家隔离数据(从各区文件收集, 从共享交换过来)(要社工派来)		身份证号, 身份证号, 健康信息等	✓	✓				
21		高	居家隔离数据(从各区文件收集, 从共享交换过来)(要社工派来)		身份证号, 身份证号, 健康信息等	✓	✓				
22	网信办	高	健康码个人信息(包含)		身份证号, 身份证号, 健康信息等	✓	✓				
23		高	健康码个人信息(包含)	RKFR_SWQB_BZDZKZD19	身份证号, 身份证号, 健康信息等	✓	✓			7,355,329	电子政务外网
24		高	健康码个人信息(包含)	RKFR_SWQB_ZZBZDZKZD19	身份证号, 身份证号, 健康信息等	✓	✓			188,232	
25	政法委	高	健康码个人信息(包含)	RKFR_SWQB_ZZBZDZKZD19	身份证号, 身份证号, 健康信息等	✓	✓			24,033	
26	基础库	高	健康码个人信息(包含)		身份证号, 身份证号, 健康信息等	✓	✓				政务网
27		高	健康码个人信息(包含)		身份证号, 身份证号, 健康信息等	✓	✓				政务网
28		高	健康码个人信息(包含)		身份证号, 身份证号, 健康信息等	✓	✓				政务网

SZ疫情数据治理 3、模型设计(疫情主题库LDM)



根据疫情防控的业务流程抽象出目标数据模型, 基于目标数据模型建设深圳疫情主题库, 将来自于跨部门、跨系统、跨层级的各类疫情数据, 清洗融合起来, 通过深圳疫情主题库的数据, 一方面支撑疫情防控的精准调度, 一方面支撑疫情防控的分析研判。





SZ疫情数据治理 7、数据分析挖掘



基于疫情防控的各细分场景，构建疫情态势监测与趋势研判的分析模型，支撑疫情防控的指挥调度与领导决策。



SZ疫情数据治理 8、数据服务应用



(1) 领导决策：疫情态势分析展示



(2) 数据报告：疫情信息及时上报



(4) 疫情管控支撑

支持自主申报的信息核验、亮码扫码的及时信息共享等场景



(3) 疫情线索：疫情线索及时下发

将自主申报等业务数据，与基础库数据进行融合碰撞，实现人房关联，并通过资源网实现数据下发至各区，支撑各区对疫情风险及时防控。





数据标准与数据安全

IITC工信人才岗位能力评价培训
数据产业高层次人才培训课程

首席数据官
GOVCDO

数据标准与数据安全

CDO机制如何把“战略”落到制度/机制/技术，
再用“安全”兜住底线，驱动数字化转型

讲师：何旭珩 数字理政 2025年10月

课程定位：推进从战略到落地的两条主线

首席数据官
GOVCDO

01 CDO四大作用

- 数据管理与合规的守护者
- 数据驱动和决策支持者
- 业务流程优化的推动者
- 创新与数字化转型的引领者

02 课程主线

- 数据标准：
统一口径 | 一致可比 | 共享流通
- 数据安全：
底线护栏 | 防泄止损 | 资产护航

03 课程目标

- 了解操作：将“标准评审”
“安全准入”无缝嵌入现有流程。
- 未来成效：标准一致、数据可比；
风险可控、守住底线。

课程前铺垫，这些缩写到底啥意思？



- CDO: 首席数据官。像“CFO管钱”，CDO“管数据这类资产”。
- ER图: 实体关系图。像“组织架构图”，但画的是数据对象怎么关联。
- DSMM/DMM/CMM: 能力成熟度模型。像“等级考试”，评你组织在数据上的级别。
- CIA: 保密性/完整性/可用性。信息安全的“三道护栏”。
- DLP: 数据防泄漏。相当于“带不走的U盘”和“发不出去的敏感表”。
- MFA: 多因子认证。像网银“短信+指纹”，更安全。
- RTO/RPO: 恢复时间/恢复点目标。系统出故障“多快恢复”和“最多丢多久的数据”。
- KPI/KRI: 业绩指标/风险指标。前者盯产出，后者盯风险。
- PII/敏感数据: 能识别个人的信息，如姓名、身份证、手机号等。
- UTM: 无人机交通管理；
- ETL: 抽取—转换—加载数据的流水线；
- MDM: 主数据管理，像“全公司统一通讯录”。讲师备注：讲法统一：“它就像……”，





数据标准对象视角：主/参考/交易/指标四象限



可量化收益：标准化后看得到的三类回报



02 标准体系是什么

标准“六件套”总览：统一语言的最小闭环

首席数据官 GOVCDO

 <p>① 命名规范：同名同义，避免歧义。 ② 编码规则：唯一可校验，便于追溯 ③ 数据元/属性：字段定义+取值域+校验规则。 ④ 指标口径：业务为锚、算法唯一、版本可追。 ⑤ 数据字典：主/参考、交易分域管理，权威来源。 ⑥ 交换/接口：最小字段集+白名单机制，互操作。</p> <p>六件套构成</p>	 <p>✓ 可追溯：术语统一→口径定版→字段映射→接口发布→版本留痕。 ✓ 可验证：每一步都有责任人、时间戳与变更记录。 ✓ 可闭环：争议回溯到卡片与版本，定位—修订—同步—复验。</p> <p>管理目标</p>	 <p>➢ 卡片化：先做“指标口径、字段最小集、接口清单”，版本可追。 ➢ 平台化：规则参数化进平台与流程（评审/准入/发布）。 ➢ 强化化：试运行达标后，统一切换为强制校验与拦截。</p> <p>工作方式</p>	 <p>● 对账口径：以《统一口径清单》为唯一依据，差异按清单逐项逐条定位。 ● 接口验收：最小字段集必须齐全有效；错误码体系完备且可追。 ● 版本审计：记录版本号、生效日与试运行期结果，变更有据可回溯。</p> <p>验收口径</p>
--	---	--	---

命名与编码：从“看得懂”到“管得住”



常见踩坑：一物多码、临时码常态化、含义码导致改版代价陡增。

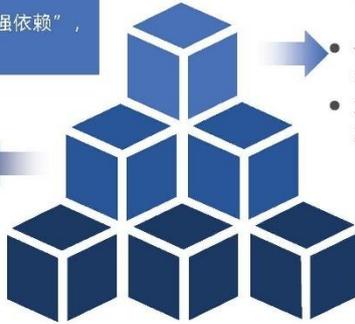
管理落点：命名与编码规则入库为“强依赖”，新增对象必须给出发码策略+冲突解决。

中英混用规则

- 何时用中文：制度、组织、业务实体与监管 名词（如“发票”“供应商”“增值税”）。
- 何时用英文：通用技术/协议/缩写与单位（API、SDK、HTTP、ID、URL、GB、ms）。
- 如何搭配：固定顺序+统一分隔（如 invoice_id / 发票ID 二选一且项目内唯一）。

大小写/下划线统一

- 定风格：明确采用何种风格。
- 定范围：代码、数据库、API、文件名一体化规范。
- 定约束：工具校验强制执行，禁止混用与随意缩写。



表/字段前后缀明确

- 层级前缀：如ods(贴源) | dwd(明细) | dws(汇总) | dim(维表)。
- 领域前缀（可选）：ec(电商) | fin(财务) | ops(运维)...
- 时间后缀：_h/_d/_w/_m/_q/_y 表示粒度；增量/全量：_inc/_full。

术语标准的重要性



案例：身份证号码多种命名

id_card_number
idno
identity_number

• 后果：跨系统数据 JOIN 失败，
• 需额外开发映射逻辑

术语管理生命周期：



01

业务术语标准的定义与作用

- 统一业务概念的语言与表达
- 避免同义不同名、歧义和口径冲突
- 为字段命名、数据元定义、指标计算提供统一基础

02

统一业务语言的重要性

- 让跨系统协作、数据对接更高效、更准确
- 降低数据集成成本和出错概率

03

术语标准与数据库字段命名的关系

- 术语标准 = 数据库字段命名规则的词根词典
- 保证全库字段命名一致性

数据元与指标口径：标准化解决争议

诊断方法：若同指标出现多口径，先查“是否存在”与“版本是否一致”。

落地提示：对外披露或经营看板上的指标，必须有口径与验证用例，且被纳入版本管理。

数据元 (7+2要素)

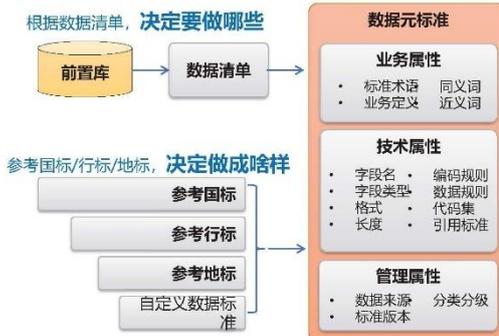
- 名称 | 定义 | 取值域 (枚举/范围/正则) | 单位 | 精度 (小数位/进位规则) | 空值 (可否为空/默认值) | 示例
- 来源系统 | 责任人 (Owner)



指标口径

- 指标定义 (业务含义与计量对象)
- 分子/分母 (含排除项)
- 口径范围 (主体/地区/渠道/是否含税等)
- 过滤条件 (状态、去重、异常剔除)
- 时间粒度 (h/d/w/m)
- 归属账期 (发生日/入账日/支付日等)
- 口径负责人 (Owner)
- 生效日 (版本号)
- 验证用例 (样例SQL/对账方法)

数据元结构与作用



1. 数据元的定义

- 数据元是字段的最小标准化单元
- 包含：名称、定义、数据类型、长度、单位、取值规则
- 是业务语言与数据库语言的桥梁

2. 数据元的构成

- 业务属性：术语名称、业务定义、适用场景
- 技术属性：数据类型、长度、单位、取值规则、编码格式
- 管理属性：版本号、状态、维护人、更新时间

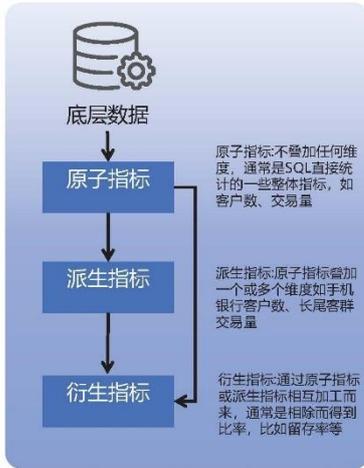
3. 数据元与字段的关系

- 字段是数据库的物理实现
- 数据元是字段的标准化设计规范
- 一个数据元可映射到多个系统的同义字段

4. 典型问题案例

- 字段长度不一致 (例：税务号15位 vs 20位)
- 系统对接时，超长数据被截断或接口报错

指标分类与作用



1. 指标标准化解决口径不一致问题

- 确保所有部门和系统使用同一逻辑计算指标
- 减少因定义差异导致的数据争议和混乱

2. 指标分层结构

- 原子指标：最基本、不可拆分的指标，如注册用户数
- 派生指标：基于原子指标通过简单计算得到，如活跃度=活跃用户数/注册用户数
- 复合指标：多个指标组合而成，用于综合评价

3. 标准化内容

- 指标名称
- 计算口径（定义清晰）
- 计算周期（如日、月、年）
- 计算逻辑（具体算法说明）

4. 案例说明

- 不同部门对“注册用户数”的定义不同，导致报表结果差异明显
- 标准化后，BI分析与报表数据统一，决策更准确

统一口径清单——样例



字段/指标名	业务定义（一句话）	计算口径（伪SQL/规则）	来源系统	更新频率	口径Owner	版本	生效日	验证用例（GIVEN/WHEN/THEN）
月活用户MAU	30天内登录过的独立用户数	THEN count(distinct user_id where login_dt >= today - 30)	用户中心	日	运营数据域	v1.2	2025-11-01	GIVEN新增访客 →WHEN未登录 →THEN不计入
对账成功率	对账无差异的批次占比	THEN ok_batches / total_batches	结算系统	日	财务数据域	v0.9	2025-10-01	GIVEN异常重跑 →THEN算一次批次

数据字典与交换接口：让系统“说同一种话”



安全前置：接口白名单、鉴权方式、访问留痕与节流。

同一种话

数据字典

01

- 主数据（客户/商品/供应商）：唯一主键 | 变更留痕 | 权威归口。
- 参考数据（地区/行业/币种）：统一来源 | 版本号 | 统一分发。
- 交易数据（订单/支付）：最小必填 | 一致口径 | 时序可追。

交换接口

02

- 最少字段：只传办业务必需，其余不传。
- 版本：新旧并行（如 90 天），提前公告，按计划切换。
- 不重复：同一请求多次也只处理一次，用唯一请求号判重。
- 分批/增量：可分批取；也可只取“最近有变化”的数据。
- 出错与重试：分成功/使用方问题/系统故障；给可读提示 + 追踪号；标明可重试情形与间隔。
- 质量校验：必填、选项、范围、格式、前后逻辑一致。

主数据与参考数据区分及标准化意义



主数据

- 描述企业核心业务对象的数据
- 高复用、高价值，支持跨系统、跨部门共享
- 需全系统唯一、权威
- 如：客户、供应商、员工、产品、物料等

参考数据

- 定义其他数据属性值范围的数据
- 枚举/代码类字段，稳定且有限
- 基于国际/国家/行业标准如：币种、单位、国家/地区、时间周期等

主数据	人员			
定义	人员主数据是指所有与企业签署了正式劳动合同的人员。人员主数据是从企业管理输出的人员实体的数字化描述。			
序号	属性名称	属性性质	类型	取值范围
1	人员编码	系统自动生成	字符串	系统自动生成 7 位流水码
2	姓名	必填项	字符串	需员工工姓名，同身份证上的名称一致，必须保证姓名输入准确
3	身份证号	必填项	字符串	位数为 15 位或 18 位的身份证号码（港澳台及外籍除外）
4	性别	必填项	枚举型	男、女
5	出生日期	必填项	日期型	须与身份证上的出生日期保持一致
6	民族	必填项	参照型	参照民族档案
7	电子邮件	必填项	字符串	不能为空，格式：zhangan@cn-mhg.com
8	办公电话		字符串	区号加 6-8 位电话号码（含号），中间以“-”连接，如 010-12345678 或 010-12345678-8888
9	手机	必填项	字符串	位数为 11 位（港澳台及国外除外），不限其他符号代替
10	学历	枚举型	枚举型	小学、初中、高中、大专、本科、硕士、博士
11	状态	必填项	枚举型	在职、离职
12	备注		字符串	

标准化目的

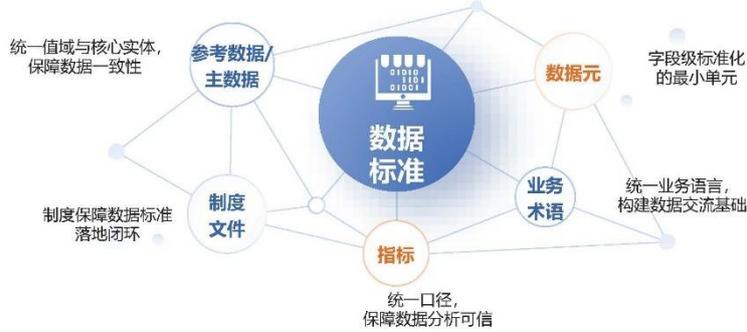
- 保证跨系统数据匹配
- 避免因值域不一致导致数据关联失败



案例

- 性别字段值“男/女” vs “M/F”导致统计结果差异
- 行政区划代码不统一，JOIN 查询结果缺失

数据标准体系小结



● 落地流程:



数据标准制度类型概览与保障机制



层次清晰、上下衔接的制度文件体系，是数据标准化“可执行、可监督、可持续”的关键保障。



保障数据标准的规范设计、评审、发布、维护流程
融入开发、运维、上线、变更各环节，实现标准持续落地
缺乏制度支撑时，标准难以执行，字段命名混乱，系统维护成本大幅增加

制度与数据库标准化执行的关系



制度文件示例



建立由“1+12+N”制度框架和“1+10+N”企业标准框架组成的“一套制度规范”体系在集团层面统一构建管理制度和标准规范，上下贯通、刚性执行，全面提升数据治理成效

制度的标准化要求

- 明确规范制度制定范围、内容和责任人
- 细化标准设计、评审、发布、维护等关键环节
- 保障制度与技术标准深度融合，便于执行
- 制度语言统一、结构清晰，便于理解应用

制度文件执行的闭环机制

- 制定 → 评审 → 入库 → 稽核 → 反馈
- 保障标准持续有效执行，形成闭环管理

制度落地关键成功因素

- 高层支持，确保制度强制执行
- 结合自动化工具，实现制度自动校验
- 明确责任分工，保障各环节闭环落实
- 持续监控和优化，形成制度执行常态化

CDO拍板 vs 团队落地



CDO拍板：

- 标准Owner名单与职责；
- 三闸口制度与通过门槛；
- 重大变更的审批人/并行期/回退阈值；
- KPI/KRI 目标值与奖惩挂钩；
- 安全SOP的升级路径与对外口径。

团队落地：

- 字段口径、样例SQL、接口最小字段；
- 数据质量规则清单与异常处理；
- 工具选型、表结构、作业编排；
- 日志路径、证据留痕与稽核脚本。

制造业（良率对账）

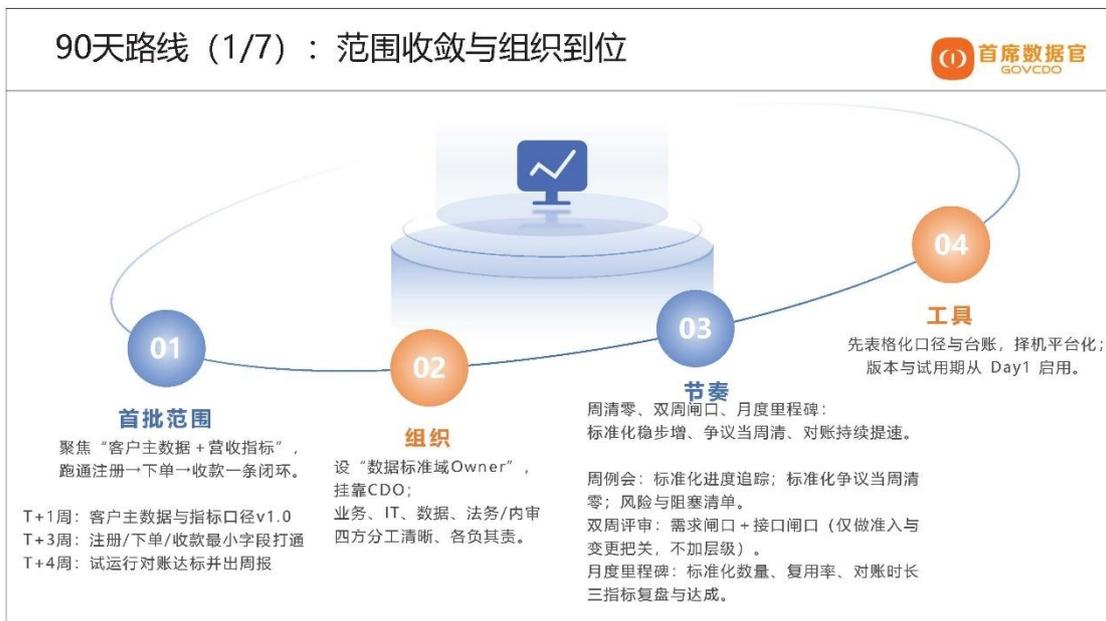
- 口径统一→复判率从[%]降到[%]；
- 每周人工对账工时从[h]降到[h]×[人]；
- 年节省=工时节省×人工成本×周数 + 库存周转改善收益。

政务/公共服务（多源对账）

- 多部门口径统一→跨部门对账周期从[天]→[天]；
- 年收益=办理时效提升×件量×价值系数 + 投诉率下降×罚金/舆情成本降低。 讲师备注：刻意展示“算式”，便于CFO复核。

03

90天路线：从0到1



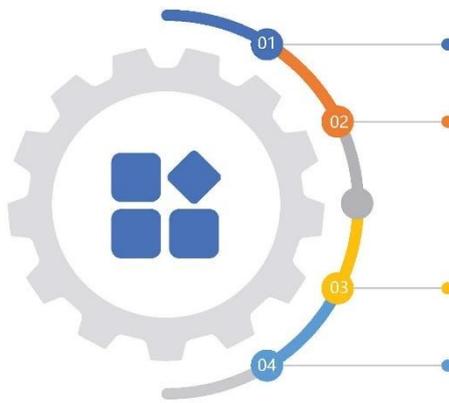
90天路线 (2/7) : 需求调研与痛点清单



90天路线 (3/7) : 标准卡片起草与快速评审



90天路线 (4/7) : 宣贯与试运行



宣贯对象

- 覆盖面：产品、研发、测试、运营、财务
- 形式：每日传达关键信息；FAQ沉淀高频问题
- 目标：统一口径 + 快速答疑 + 当日可执行

试行处理

- 先试行：选定“注册-下单-收款”闭环，上新口径与最小字段集。
- 保对照：老口径只读保留，面向报表/查询，不再写入。
- 看差异：对账看板每日比对（新vs旧），异常阈值触发复盘与修正。

及时反馈

- 先修正：当班发现-当班修-当班灰度验证。
- 经版本：版本微更 + 生效日 + 变更记录。
- 有护栏：试运行比例、回滚点、公告触达，审计留痕

成果产出

- 最终产出《试运行差异单》：记录试运行期新旧口径差异，为接口改造与数据质量治理持续限数。
- 定位差异原因，派单整改，支撑切换决策。

90天路线 (5/7) : 接口改造与准入闸口生效



改造范围

- 改造范围：最小字段集（必填/取值域）+ 字段/枚举映射（单位/币种/状态）。
- 完整性补齐：错误码体系、幂等机制、分页/增量规则。
- 不动项：业务逻辑与非关键字段不改；接口路径与鉴权不变（除非有合规风险）。

准入闸口

- 三处把关：
 - ①需求评审
 - ②接口评审
 - ③上线前检查
- 硬性红线：无标准不接入；无最小字段集不接入；未标版本与生效日不接入。
- 目标：源头控质、减少返工、保障稳定与安全。

伴随改动

- 鉴权与白名单：强制鉴权，仅白名单系统可调用，最小权限。
- 留痕与节流：全量id、审计日志留痕，按租户、IP、接口限速与 并发门限，防抖峰值。
- 对外台账：登记外部对接方、版本、生效日、用途，支持追溯与年度复核。

度量指标

- 评审一次通过率：反映方案成熟度与规范到位。
- 上线缺陷回滚率：反映改造质量与稳定性。
- 平均接入时长：反映对业务接入效率的影响

回退策略

- 只读窗口：上线改造期保留 只读查询，禁止写入，保障业务可查可核。
- 字段对照：提供新旧字段映射表（含取值域/单位/枚举），回退与修复均有据可循。
- 快速回切：预留回退脚本与切换开关，触发即生效，影响面最小化。

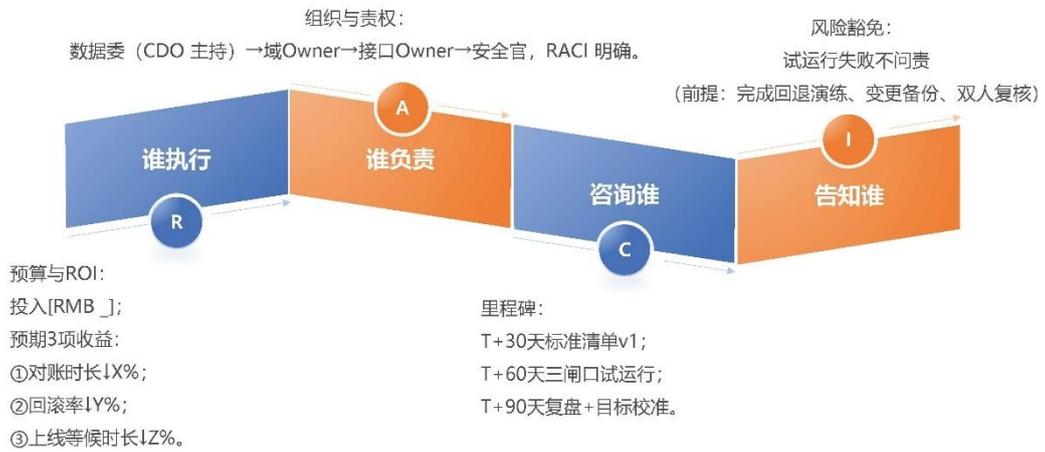
90天路线 (6/7) : 数据质量校验与验证用例闭环



90天路线 (7/7) : 复盘评估与扩域复制



90天路线：高管视角



04 数据标准管理落脚点

统一口径清单：对账唯一依据

首席数据官 GOVCDO

定义

- 内容：指标名、业务含义、计算口径（含分子/分母、过滤、账期）。
- 形态：逐项标准化内容 + 版本号/生效日，集中成清单。
- 作用：对账唯一依据、复盘唯一依据、争议回溯锚点。

必填字段

口径清单必填十项：
指标名、业务定义、分子/分母、维度与过滤、账期与粒度、生效日、口径负责人、版本号、验证用例和影响系统。

使用规则

- 所有看板/报表/对外披露/绩效考核一律引用“清单编号”；
- 新增/变更指标必须走评审 + 版本公告，禁止私改。

纠纷处理

- 标准台账证据化管理：每条记录配样例数据 + 日志位置。
- 支持随机抽查，可追可证。

标准台账：从字段到接口的可追溯

首席数据官 GOVCDO

三类台账

标准台账三类：字段台账（库/表/列）、指标台账（口径清单）和接口台账（最小字段集）——自底向上、贯通采用。

接口台账必填项

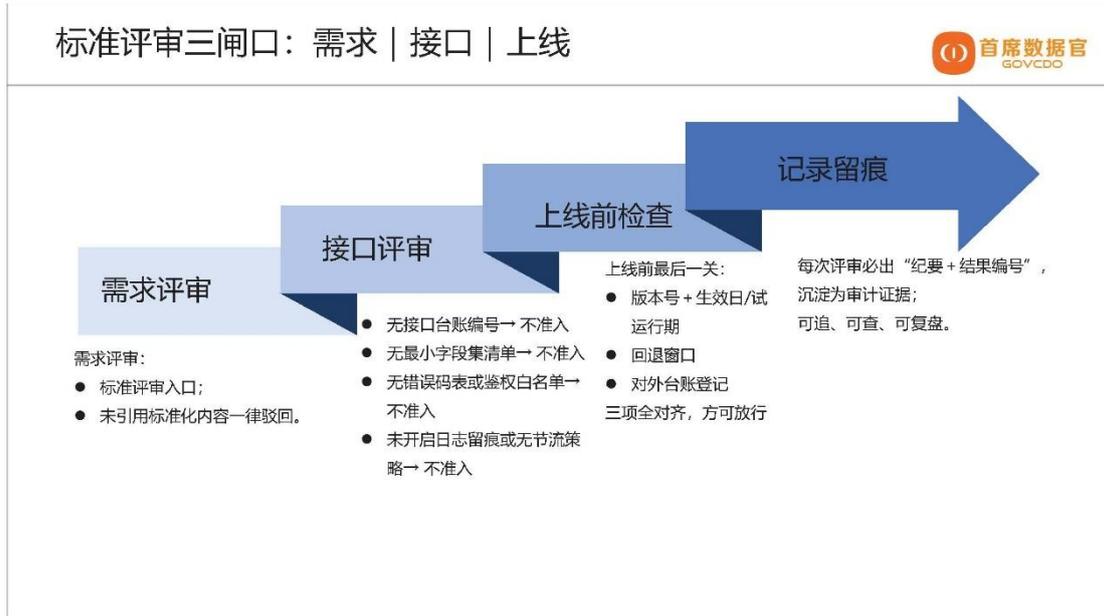
接口台账必填：
接口名、调用方、最小字段集、错误码、分批/增量、鉴权/白名单、兼容窗口、回退策略。

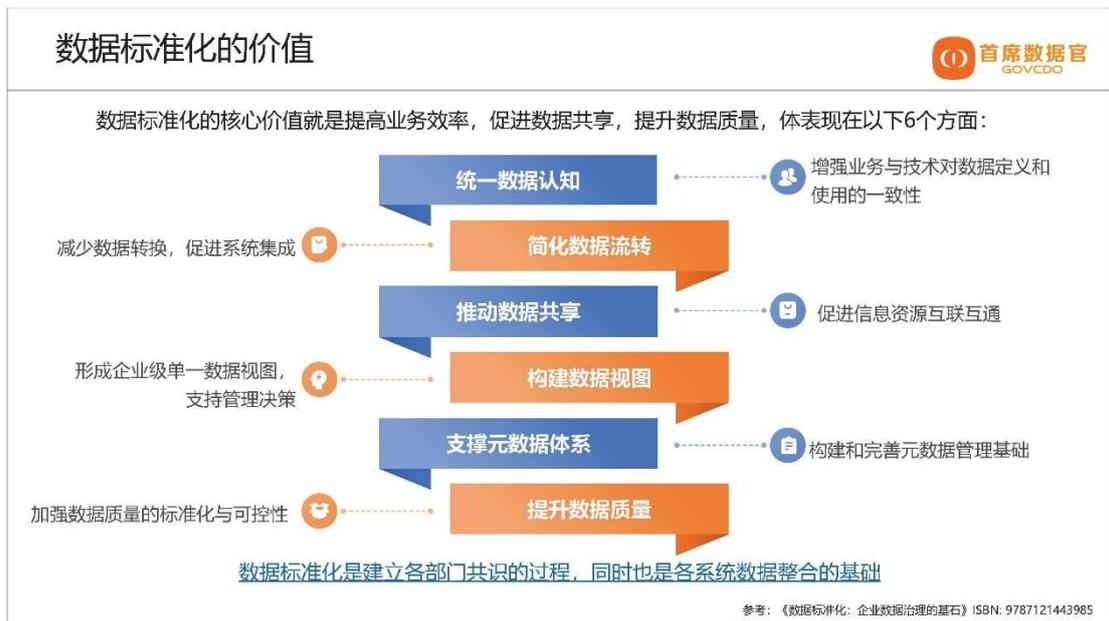
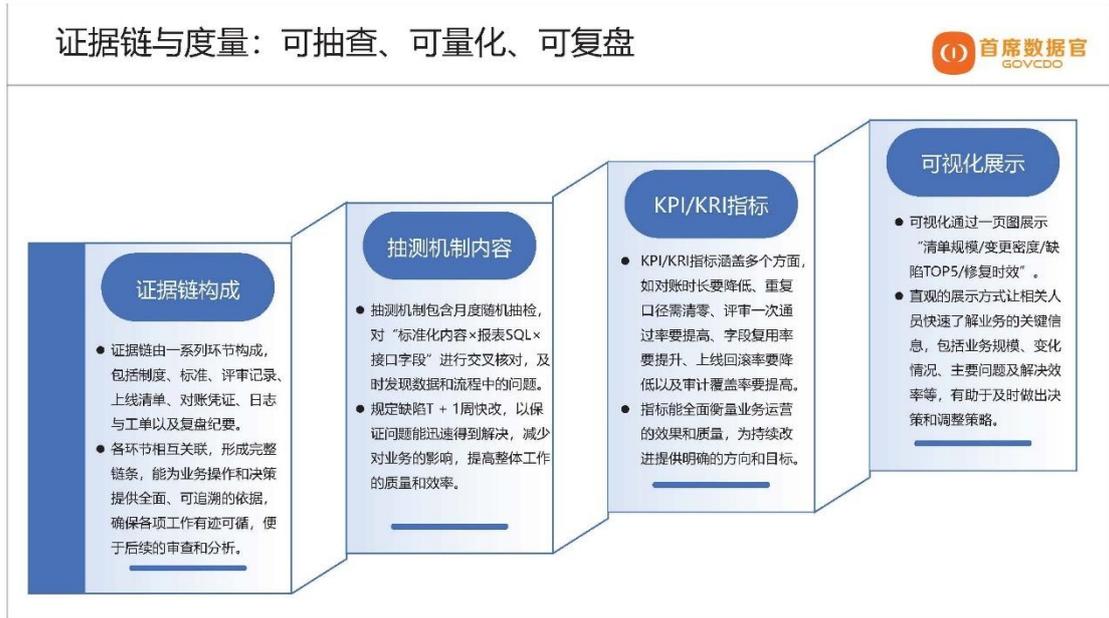
证据化要求

标准台账证据化管理：每条记录配样例数据 + 日志位置，并支持随机抽查，可追、可验、可复现。

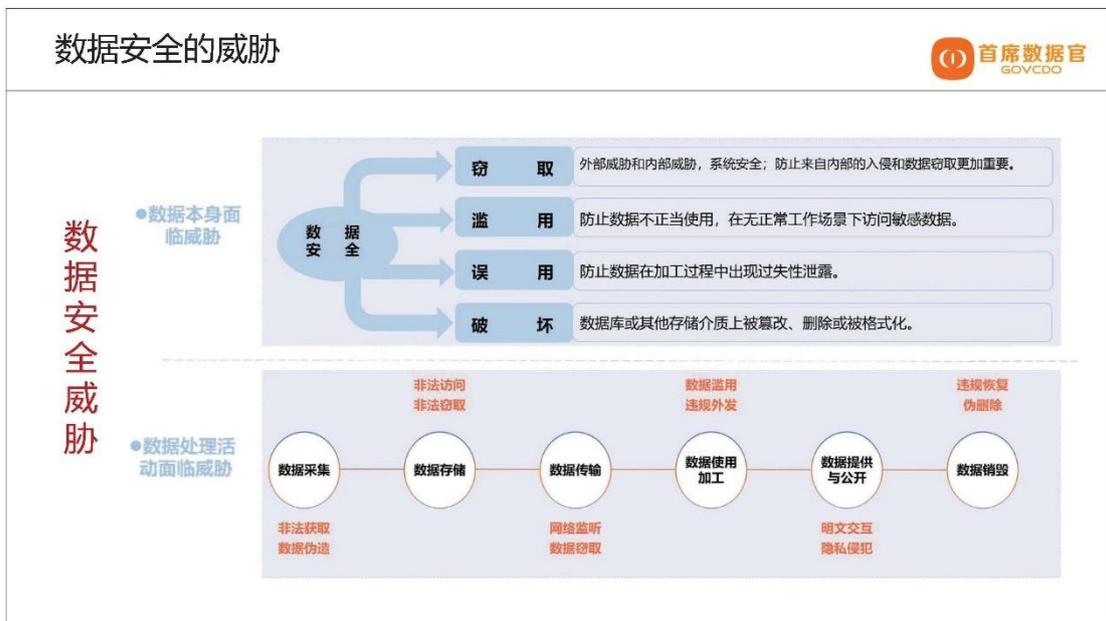
字段台账必填项

字段台账：定义清楚、来源可追、规则可检和责任可问——以必填项锁定一致性与可维护性。





05 数据安全总纲



个人隐私与伦理规定

合法性基础

- 隐私维度的合法性基础是保障个人隐私的根本。
- 在处理个人信息时，需确保有明确合法的依据，为后续操作奠定坚实基础，保证整个信息处理流程符合法律规定，避免因合法性缺失带来的法律风险，维护个人信息主体的合法权益。

告知同意

- 告知同意是尊重个人对自身信息控制权的重要体现。
- 在收集和使用个人信息前，要清晰明确地告知信息主体相关情况，包括用途、方式等，获得其明确同意，使信息主体在充分了解的基础上自主决定是否让渡信息使用权。

最小必要

- 最小必要原则要求在处理个人信息时，仅收集和使用完成目的所需的最少信息。
- 避免过度收集和使用信息，减少个人信息泄露风险，保障信息主体的隐私安全，确保信息处理活动既满足业务需求，又不过度侵犯个人隐私。

目的限定

- 目的限定明确了个人信息处理的范围和方向。
- 收集和使用信息必须围绕特定、明确的目的进行，不得超出该目的进行额外处理，防止信息被滥用，让信息处理活动始终处于可控状态，保护个人信息不被随意扩散和不当利用。

治理+技术双轮驱动

01

安全治理侧要点

治理侧五要素：

- ① 组织与职责
- ② 制度与流程
- ③ 角色与权限
- ④ 第三方管理
- ⑤ 对外提供合规条款

以制度固化安全与合规。

02

安全技术侧要点

技术侧七柱：

- ① 身份鉴别/多因子
- ② 访问控制
- ③ 加密
- ④ DLP/水印
- ⑤ 终端
- ⑥ 网络
- ⑦ 云
- ⑧ 日志与审计

从入口到留痕，全链路护航数据安全。

03

事件SOP节奏

事件SOP——5小时节奏

- 1h分级研判
- 2h遏制
- 3h通报法务
- 4h第三方处置
- 5h董监层与外部口径统一。

04

KRI挂经营指标

KRI与经营挂钩如：

- 重大事件=0
- 离职账号回收≥99.5%
- 高敏访问MFA≥98%
- 对外提供合规=100%。



落实法规政策要求



国家立法工作有序推进，数据安全领域的基础性法律体系逐步完善

以《网络安全法》《数据安全法》《个人信息保护法》和《网络数据安全管理条例》为核心的数据安全领域“3+1”基础性法律框架已形成，保证数据安全实践有法可依，将数据安全保护提升至国家层面的新高度



立足载体



立足本体



立足价值



综合实践

法律与强制性要求



第二百五十三条之一 【侵犯公民个人信息罪】违反国家有关规定，向他人出售或者提供公民个人信息，情节严重的，处三年以下有期徒刑或者拘役，并处或者单处罚金；情节特别严重的，处三年以上七年以下有期徒刑，并处罚金。
违反国家有关规定，将在履行职责或者提供服务过程中获得的公民个人信息，出售或者提供给他人的，依照前款的规定从重处罚。
窃取或者以其他方法非法获取公民个人信息的，依照第一款的规定处罚。
单位犯前三款罪的，对单位判处罚金，并对其直接负责的主管人员和其他直接责任人员，依照各该款的规定处罚。

第五条 非法获取、出售或者提供公民个人信息，具有下列情形之一的，应当认定为刑法第二百五十三条之一规定的“情节严重”：

- (一) 出售或者提供**行踪轨迹信息**，被他人用于犯罪的；
- (二) 知道或者应当知道他人利用公民个人信息实施犯罪，向其出售或者提供的；
- (三) 非法获取、出售或者提供**行踪轨迹信息、通信内容、征信信息、财产信息五十条以上的**；
- (四) 非法获取、出售或者提供**住宿信息、通信记录、健康生理信息、交易信息等其他可能影响人身、财产安全的公民个人信息五百条以上的**；
- (五) 非法获取、出售或者提供**第三项、第四项规定以外的公民个人信息五千条以上的**；
- (六) 数量未达到第三项至第五项规定标准，但是按相应比例合计达到有关数量标准的；
- (七) 违法所得五千元以上的；
- (八) **将在履行职责或者提供服务过程中获得的公民个人信息出售或者提供给他人，数量或者数额达到第三项至第七项规定标准一半以上的**；
- (九) 曾因侵犯公民个人信息受过刑事处罚或者二年内受过行政处罚，又非法获取、出售或者提供公民个人信息的；
- (十) 其他情节严重的情形。



合规怎么落地：条款—控制—证据—责任人



外部条款/要求	内部控制/制度	可验证证据 (样例)	责任人 (岗位)
《数据安全法》“最小必要”	《最小字段集与脱敏规范》	脱敏策略截图、字段白名单、稽核日志路径	数据标准Owner/安全官
《个保法》个人同意与告知	《隐私告知与同意管理办法》	告知模板、同意记录、拒绝记录	法务/产品Owner
ISO27001 访问控制	《账号权限分级与定期盘点》	权限台账、季度盘点记录	IT运维/安全官
等级保护测评要求	《变更双人复核与回退演练》	演练记录、回退脚本、录像	变更负责人/运维



数据分类分级规则——国家标准



- 《数据分类分级规则》作为全国网络安全标准化技术委员会更名后，发布的第一部以“数据安全”命名的国家标准，是指导各领域数据分类分级工作的基础性国标；
- 通过合并“数据分类分级指南”和“重要数据识别要求”两个标准，代替《信息安全技术 重要数据识别指南（征求意见稿）》，是贯彻国务院办公厅《扎实推动高水平对外开放更大力度吸引和利用外资行动方案》（国办发〔2024〕9号）提出的“科学界定重要数据的范围”的重要一环。



数据分类分级通用规则



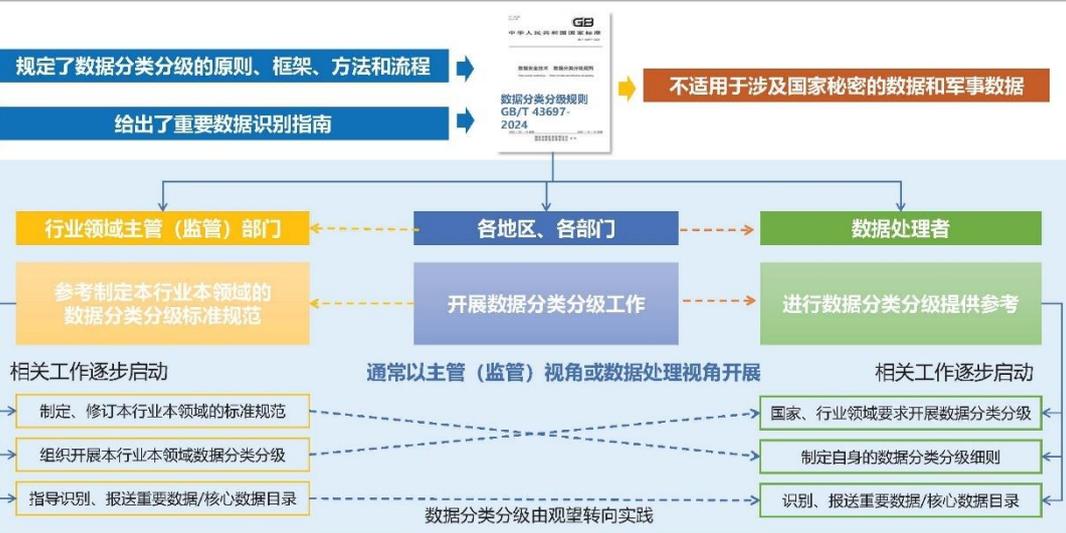
本标准给出数据分类分级基本原则、数据分类方法、数据分级框架和数据定级方法等，包括：



- 本标准在编制过程中遵循了问题导向原则、协调性原则旨在支撑《数据安全法》第二十一条提出的数据分类分级保护制度的贯彻落实。
- 解决由于缺乏国家统一的数据分类分级规则，导致相关国家数据安全制度、数据分类分级保护要求不易落地的问题。

适用范围 本文件规定了数据分类分级的原则、框架、方法和流程，给出了重要数据识别指南。
 本文件适用于行业领域主管（监管）部门参考制定本行业本领域的数据分类分级标准规范，也适用于各地区、各部门开展数据分类分级工作，同时为数据处理者进行数据分类分级提供参考。
 本文件不适用于涉及国家秘密的数据和军事数据。

适用范围



数据分类分级的五大基本原则



在遵循国家数据分类分级保护要求的基础上，按照数据所属行业领域进行分类分级管理，依据以下原则对数据进行分类分级。



数据分类方法



基本要求：数据分类可根据数据管理和使用需求，结合已有数据分类基础，灵活选择业务属性将数据细化分类。



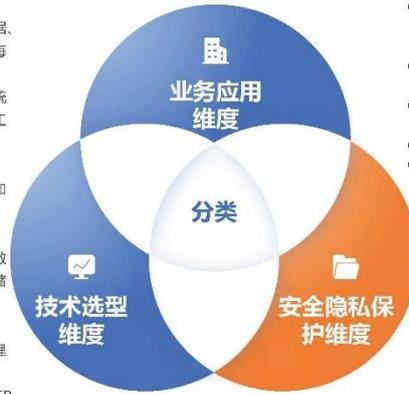
分类维度



根据国家标准GB/T 38667-2020《信息技术大数据数据分类指南》可知，可以从技术选型、业务应用和隐私安全保护三种视角对数据进行分类。

技术选型维度

- **按产生频率可划分为：**每年更新数据、每月更新数据、每周更新数据、每日更新数据、每小时更新数据、每分钟更新数据、每秒更新数据、无更新数据等。
- **按产生方式分类可划分为：**人工采集数据、信息系统产生数据、感知设备产生数据、原始数据、二次加工数据等。
- **按结构化特征分类可划分为：**结构化数据，如零售、财务、生物信息学、地理数据等；非结构化数据，如图像、视频、传感器数据、网页等；半结构化数据，如应用系统日志、电子邮件等。
- **按存储方式可划分为：**关系数据库存储数据、键值数据库存储数据、列式数据库存储数据、图数据库存储数据、文档数据库存储数据等。
- **按稀疏程度可划分为：**稠密数据和稀疏数据。
- **按处理时效性可划分为：**实时处理数据、准实时处理数据和批量处理数据。
- **按交换方式可划分为：**ETL方式、系统接口方式、FTP（文件传输协议）方式、移动介质复制方式等。



业务应用维度

- **按产生来源可划分为：**人为社交数据、电子商务平台交易数据、移动通信数据、物联网感知数据、系统运行日志数据等。
- **按业务归属可划分为：**生产类业务数据、管理类业务数据、经营分析类业务数据等。
- **按流通类型可划分为：**可直接交易数据、间接交易数据、不可交易数据等。
- **按行业领域分类可划分为：**类别见GB/T 4754-2017。
- **按数据质量可划分为：**高质量数据、普通质量数据、低质量数据等。

安全隐私保护维度

按数据安全隐私保护维度可划分为：高敏感数据、低敏感数据、不敏感数据等。

数据分级方法



数据分级是为了保护数据安全



级别确定规则



影响对象	影响程度		
	特别严重危害	严重危害	一般危害
国家安全	核心数据	核心数据	重要数据
经济运行	核心数据	重要数据	一般数据
社会秩序	核心数据	重要数据	一般数据
公共利益	核心数据	重要数据	一般数据
组织/个人权益	一般数据	一般数据	一般数据

满足以下任一条件的数据，识别为核心数据：

- 1) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对国家安全造成特别严重危害（如直接影响政治安全）或严重危害（如关系其他国家安全重点领域）；
- 2) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对经济运行造成特别严重危害（如关系国民经济命脉）；
- 3) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对社会秩序造成特别严重危害（如关系重要民生）；
- 4) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对公共利益造成特别严重危害（如关系重大公共利益）；
- 5) 对领域、群体、区域具有较高危害程度，直接影响政治安全的重要数据；
- 6) 达到较高精度、较大规模、较高重要性或深度，直接影响政治安全的重要数据；
- 7) 经有关部门评估确定的核心数据。

满足以下任一条件的数据，识别为重要数据：

- 1) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对国家安全造成一般危害；
- 2) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对经济运行造成严重危害；
- 3) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对社会秩序造成严重危害（如影响社会稳定）；
- 4) 数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对公共利益造成严重危害（如危害公共健康和公共安全）；
- 5) 数据直接关系国家安全、经济运行、社会稳定、公共健康和安全的特定领域、特定群体或特定区域；
- 6) 数据达到一定精度、规模、深度或重要性，直接影响国家安全、经济运行、社会稳定、公共健康和安全的；
- 7) 经行业领域主管（监管）部门评估确定的重要数据。

数据分类分级流程-数据处理者



在遵循国家和行业领域数据分类分级要求的基础上，开展数据分类分级



08 DSMM成熟度与评估

五级模型与现状定位

安全能力维度

组织建设 | 制度流程 | 技术工具 | 人员能力

数据安全过程维度

数据收集安全 | 数据传输安全 | 数据存储安全 | 数据处理安全 | 数据销毁安全

能力成熟度等级维度

5级: 持续优化
4级: 量化控制
3级: 充分定义
2级: 计划跟踪
1级: 非正式执行

DSMM认证企业所属行业分布

2021—2025年7月 DSMM 认证企业所属行业分布情况

DSMM认证机构所在区域分布

2021—2025年7月 DSMM 认证机构所在区域分布情况

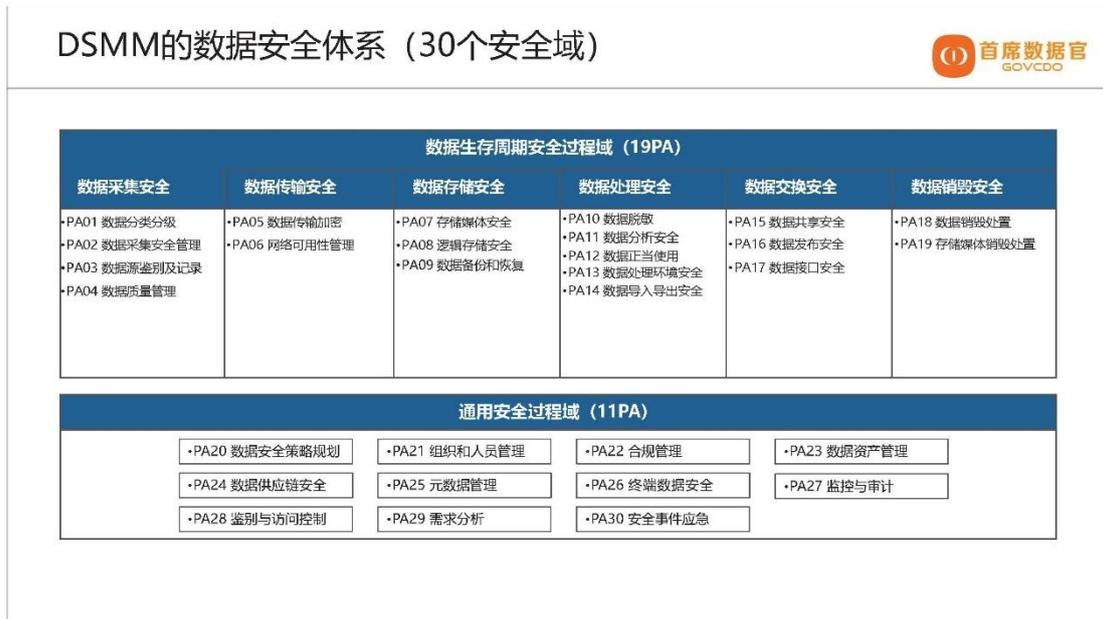
- 借鉴 DMM思想，基于CMM的框架，通过能力成熟度等级对组织在数据安全上的能力进行标准化的定义和区分。
- 针对 评估对象组织
- 围绕 数据生命周期
- 聚焦 组织在数据上的安全管理能力

DSMM认证企业所在区域分布情况

2021—2025年7月 DSMM 认证企业所在区域分布情况

DSMM等级认证分布图

2021—2025年度 DSMM 认证等级数量分布情况





治理框架与能力域组织制度角色



组织分工

- 在组织分工方面，CDO承担着统筹数据治理与合规接口的重要职责，确保数据的合理治理与合规应用。
- CISO/安全负责牵头技术与事件响应，保障数据安全。
- 法务/内审则专注于制定规则与审计工作，维护组织数据活动的规范性和合法性。



角色与权限

- 角色与权限管理遵循岗位×资源最小集原则，精准配置权限。
- 入转离实现闭环管理，做到“人走权收”，防止权限滥用。
- 对于高敏访问，采用多因子认证并留痕，进一步保障数据安全，确保只有授权人员在合规情况下访问敏感数据。



制度体系

- 制度体系构建层级清晰，以总纲政策为引领，明确整体方向；
- 职能域政策细化各领域要求；
- 操作手册提供具体操作指南；
- 应急预案应对突发情况。
- 版本可追溯，便于对制度的更新和管理进行跟踪，确保制度的有效性和适应性。

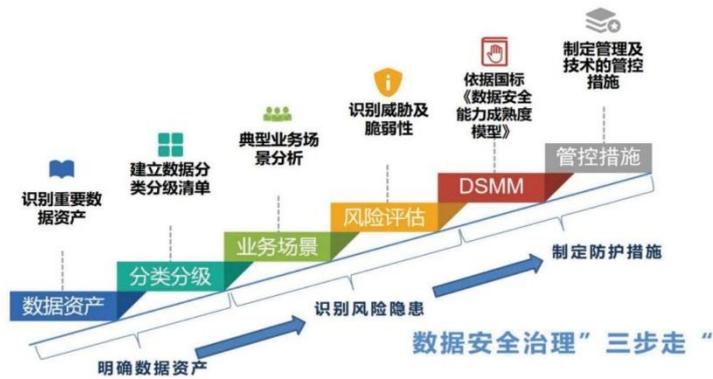


第三方治理

- 第三方治理涵盖准入评估，筛选合格的合作伙伴。
- 通过合同锚点明确用途限定和违约责任，约束第三方行为。
- 建立共享台账与抽测机制，并纳入季度盘点，及时发现和解决问题，保障与第三方合作过程中的数据安全和合规性。

如何推进数据安全治理

以DSMM为基准线，通过“目标设计—能力建设—试点运营—持续优化”的闭环流程，制定“明确数据资产—识别风险隐患—制定防护措施”的三步走数据安全治理路径，聚焦重要数据资产和典型业务场景，通过量化指标定义、技术体系支撑和安全运营优化三个方面最终达成了建设目标。



数据安全治理技术保障

01.

身份鉴别

身份鉴别是技术域的重要一环，它用于确认用户或系统的真实身份。通过有效的身份鉴别机制，可以防止非法用户进入系统，保护系统和数据的安全。例如，使用密码、指纹识别、面部识别等多种方式进行身份验证，提高鉴别准确性和安全性。

02.

访问控制

访问控制决定了用户对系统资源的访问权限。它可以根据用户的身份、角色和权限，限制对敏感信息和关键功能的访问。通过精细的访问控制策略，确保只有授权人员能够访问特定资源，降低数据泄露和系统被攻击的风险。

03.

加密/密钥

加密/密钥技术用于保护数据的机密性和完整性。通过加密算法对数据进行加密，只有持有正确密钥的用户才能解密和访问数据。合理管理密钥的生成、存储和分发，是保障加密效果的关键，有效防止数据在传输和存储过程中被窃取或篡改。

04.

DLP/水印

DLP（数据丢失防护）/水印技术可防止敏感数据的非法泄露。DLP通过监控和控制数据的流动，阻止敏感信息的不当外传。水印则是在数据中添加特定标识，便于追踪数据来源和使用情况，增强数据的安全性和可追溯性。

CIA与“可证明合规”



证据链构建

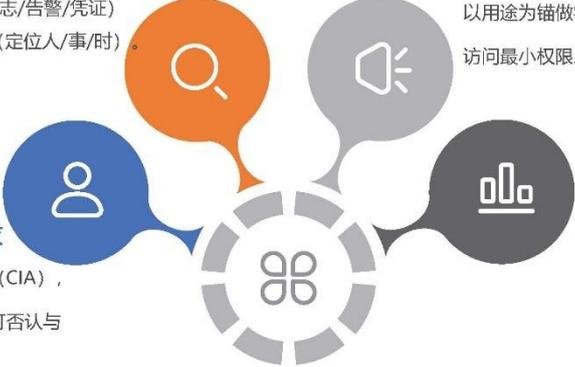
制度（政策/流程）→ 实施（日志/告警/凭证）
→ 复核（审计/抽测）→ 追责（定位人/事/时）。

最小必要原则

以用途为锚做字段最小化、去标识化/脱敏、
访问最小权限。

目标要求

- 保密性、完整性、可用性（CIA），
- 外加真实性、可问责、不可否认与
- 审计留痕≥6个月。



风险前移策略

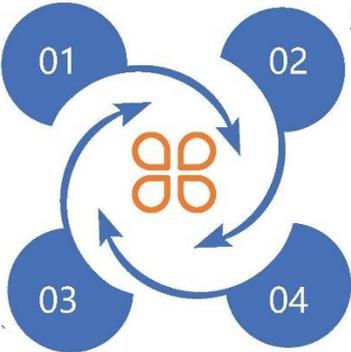
把“分级标签/接口准入/对外台账”前置到立项与采购环节。

安全与业务连续性



合规性与业务连续性目标

- 合规 = 依法有据：制度/流程/留痕/审计全到位，外部检查可证。
- 连续 = 不停摆：极端场景下关键业务可持续（RTO/RPO达标、可回切）。
- 共振 = 同频落地：任何变更先过合规，再做连续性演练与灰度切换。



备份与恢复

- 定期备份：覆盖关键数据+核心系统，多地多副本、加密存放。
- 快速恢复：明确 RTO/RPO，预置一键恢复/只读切换。
- 常态演练：按季验证恢复链路，问题当班修复

双活与演练

- 可用性：两活同城/异城，单侧故障不掉单；健康探测+自动切换。
- 一致性：明确定义一致性级别（强一致/最终一致）、冲突与幂等策略。
- 可演练：按季“有脚本可回放”，覆盖断电/链路/数据污染等场景。

RTO/RPO指标落地

- RTO = 多快把业务拉起来；
- RPO = 最多丢多少数据。
- 落地 = 给目标、配手段、能演练、可验收。



总结

首席数据官 GOVCDO

- 1.把“指标和接口”说清楚、写清楚、审清楚;
- 2.把“变更与安全”阈值化、SOP化、可回溯;
- 3.把“价值”算成钱、写进KPI、对齐预算。



企业数字化转型路径

IITC工信人才岗位能力评价培训
数据产业高层次专业人才培养课程

首席数据官
GOVCDO

企业数字化转型路径

讲师：陈立 数字理政 2025年10月

目/录 CONTENTS

01	02	03	04	05
背景	转型思路	数据资产	AI创新	总结



数字化转型的十大趋势一



来源于：清华大学企业数字化转型研究报告 2024

数字化转型的十大趋势二



来源于：清华大学 企业数字化转型研究报告 2024

国家对企业数字化转型推动



国家数字化转型政策布局

国家围绕数字经济与实体经济融合，出台系统性数字化转型政策，覆盖制造业、服务业等多领域，构建“顶层设计+专项支持”的政策框架。《制造业企业数字化转型实施指南》2024年工信、国资委、工商联联合印发（规划、落地、评估、迭代）

专精特新企业转型目标

聚焦专精特新企业，通过技术改造、研发投入补贴等政策，推动其成为细分领域数字化标杆，目标培育万家数字化示范单位。《中小企业数字化赋能专项行动方案（2025—2027年）》2027年，专精特新中小企业实现数字化改造应改尽改，全国规上工业中小企业关键工序数控化率达到75%，中小企业上云率超过40%

中小企业转型指南

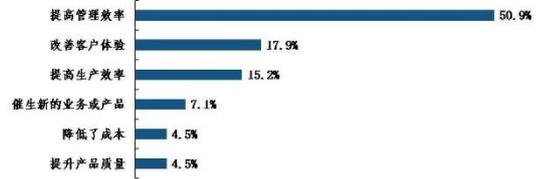
《中小企业数字化转型指南》：2022年，工业和信息化部办公厅印发，从五个方面为中小企业提供了明确的转型路径，解决了企业“不会转”的难题，引导中小企业根据自身实际情况，选择合适的数字化转型策略和方法，逐步提升数字化水平。



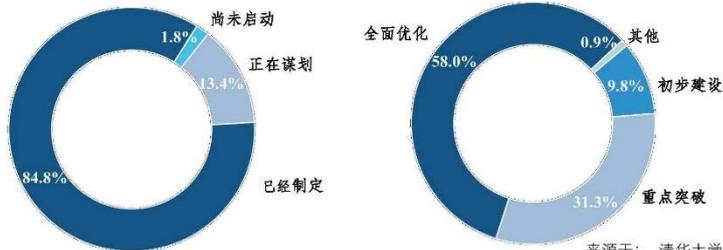
企业落实数字化转型的现状 一



增效降本是企业数字化转型主要动因之一



数字化战略指引企业全面优化数字化转型



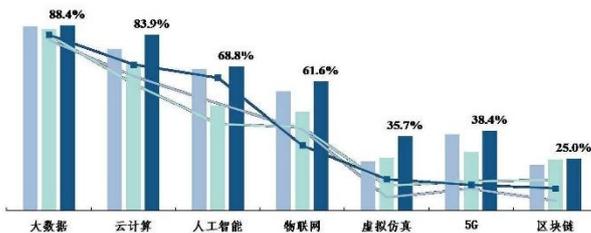
来源于：清华大学 企业数字化转型研究报告 2024

国家对企业数字化转型推动 二



企业管理高层主导数字化转型

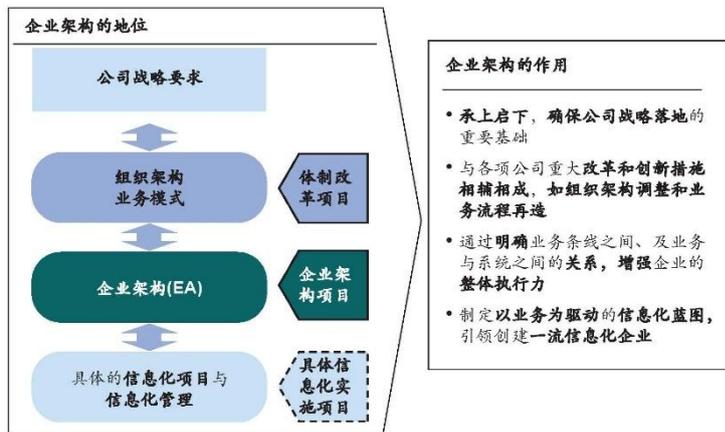
数字化转型促进技术及管理知识的积淀

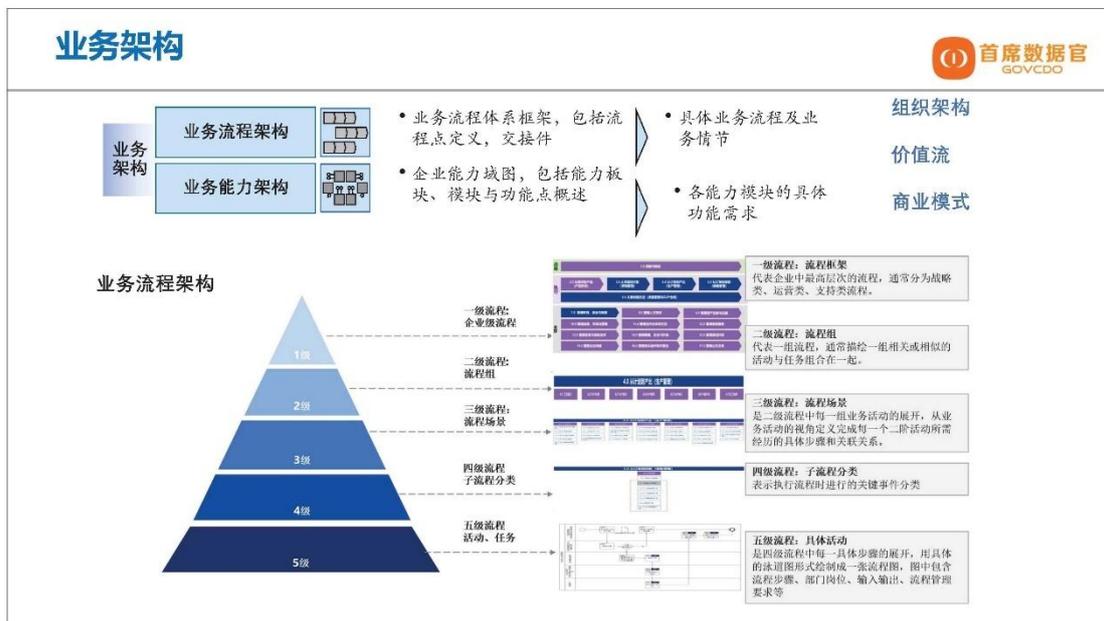
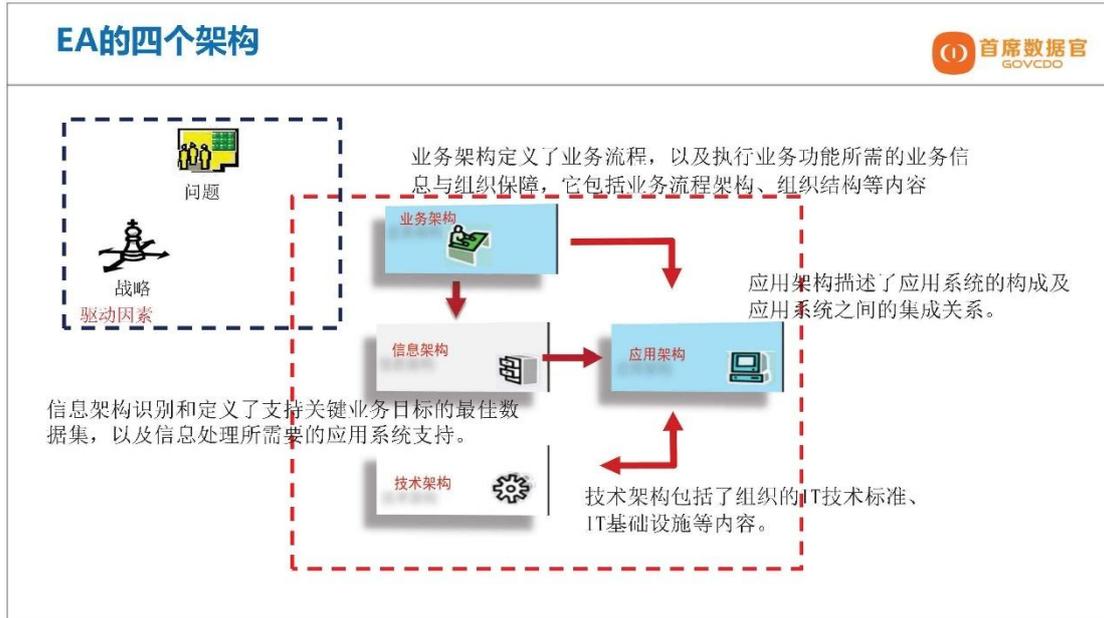


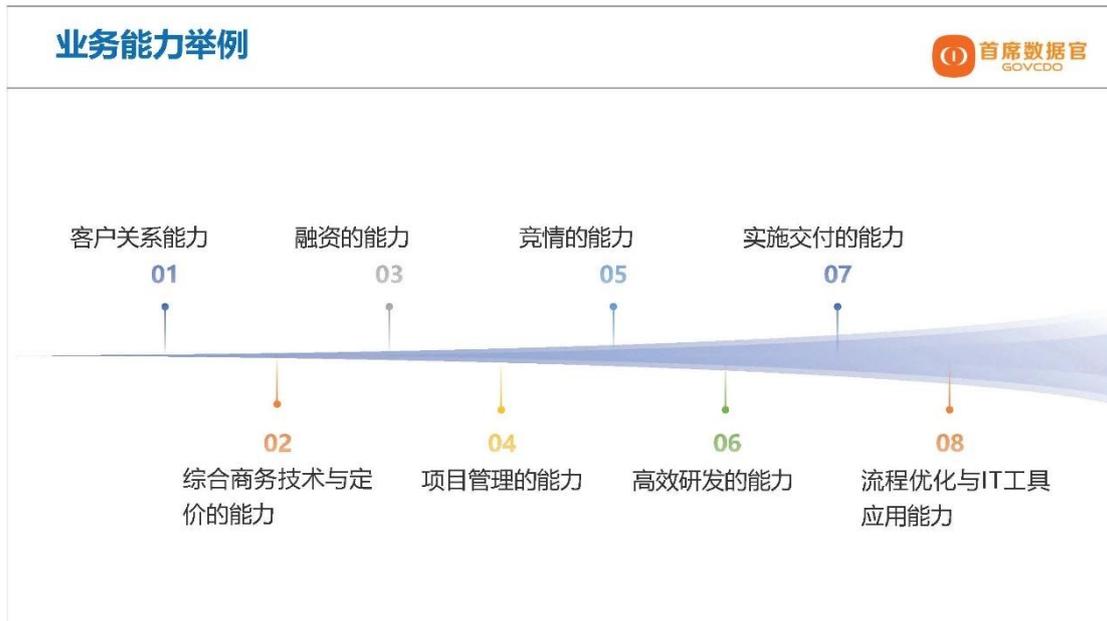
来源于：清华大学 企业数字化转型研究报告 2024

02 转型思路

企业架构是增强公司执行力，打造一流信息化企业的基石







案例一：流程和IT能力的建设基础——三大管理平台



明确IT定位



案例一：IT的组织架构



IT的资源集中管理

■ 2000年对分散的IT资源进行整合，成立了IT部门，负责IT的统一规划和建设，避免了部门各自为政、重复投入的IT建设局面

■ 2006年成立了机构及流程优化部，各单位成立流程工作组，形成了统一管理的流程运作体系，打破了流程建设方面的行政壁垒

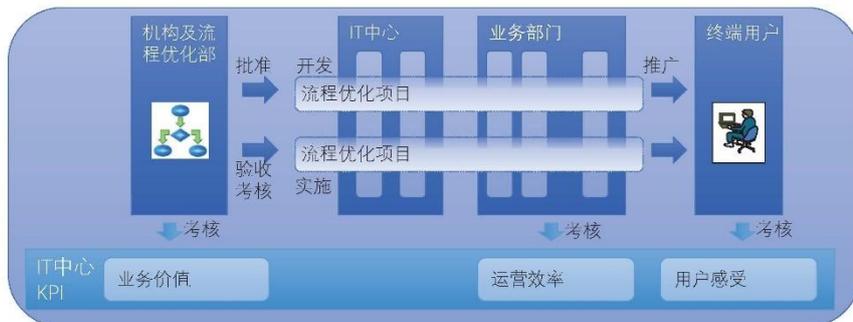
案例一：全球IT运维管理架构



总部IT、区域IT及海外机构本地IT三层管理模式，明确了各自职责，有效推动各项工作开展，有利于全局IT的布局，解决了各机构各自为政局面



案例一：IT中心的考核及与业务部门的横向关系



- 项目以端到端流程打通和效率提升为目标，从公司层面统一管理及优先级排序，矩阵运作
- 考核采用平衡记分卡方式，确保IT投入效益的最大化，也促进了IT运作效率和服务质量的提升

案例一：管理好系统架构



信息管理平台 “6 + 1” 的核心业务：以ERP为核心按照业务领域打造6大外围IT系统

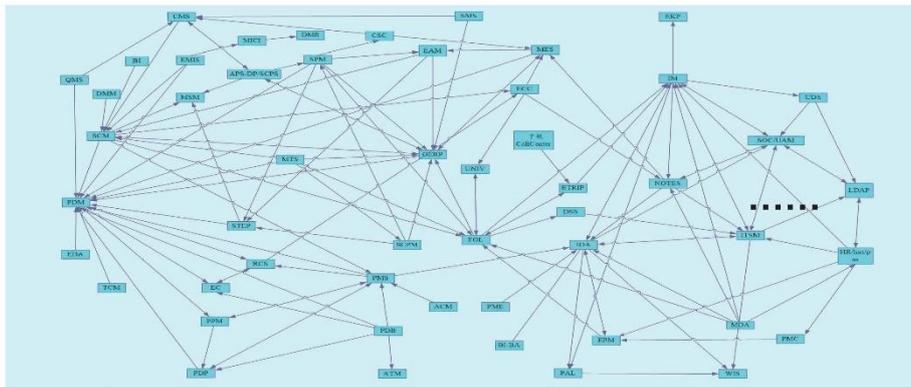
- 确立核心系统，明确规划方向
- 集中零散数据，促进系统融合
- 减少系统数量，提升用户体验



案例一：平台为核心，关注主数据管理，消除系统间数据的壁垒



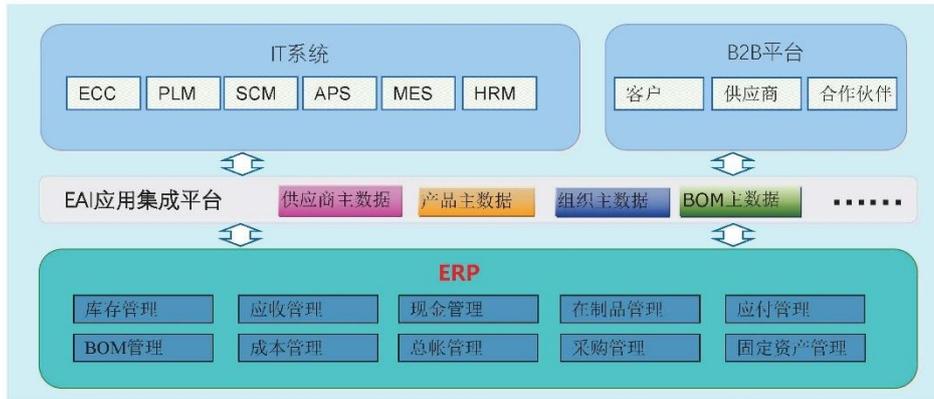
实施EAI平台以前，几十个系统蜘蛛网般的系统关系



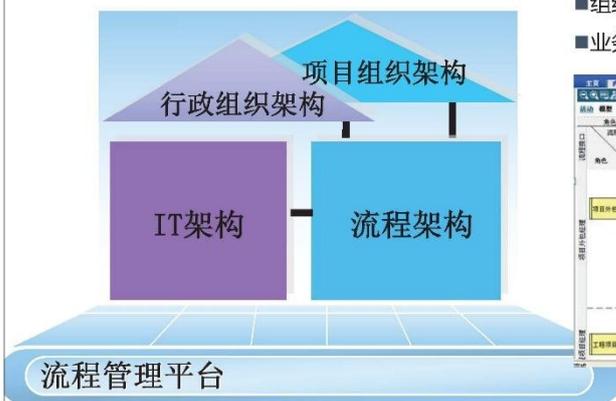
案例一：平台为核心，关注主数据管理，消除系统间数据的壁垒



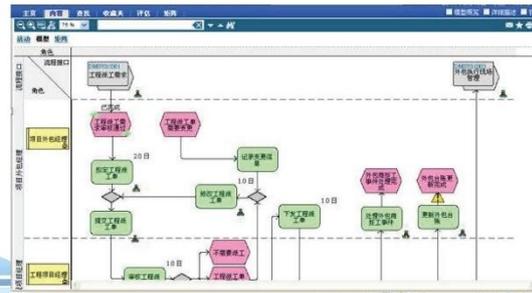
实施EAI平台，统一主数据管理以后，整合规范了各系统关系，优化了600多系统间接口



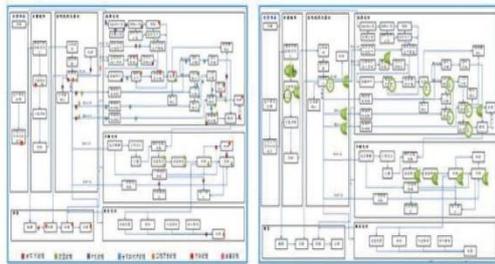
案例一：流程管理平台 建立流程、组织和IT的关联关系



- 组织架构和流程的管理可视化
- 业务流程和IT系统建立对应关系



案例一：端到端业务流程的IT系统贯通

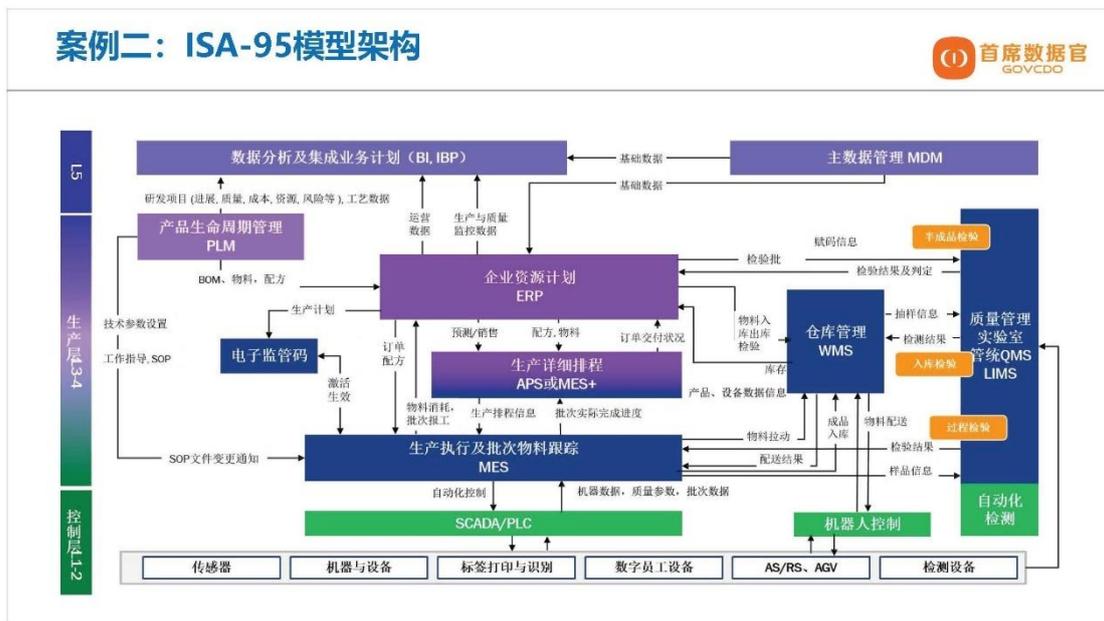
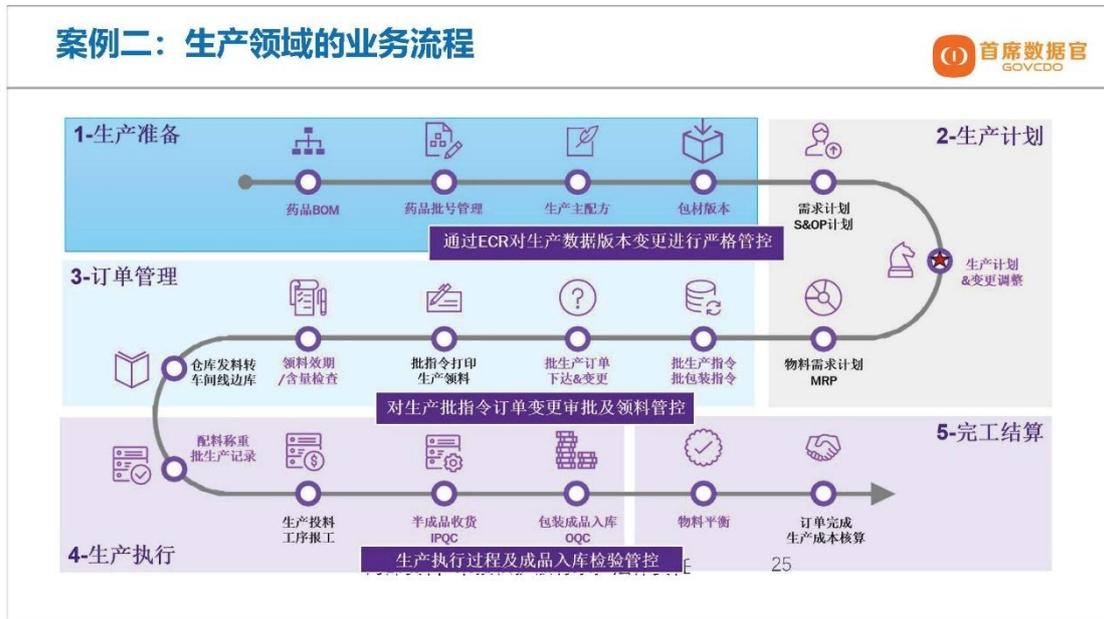


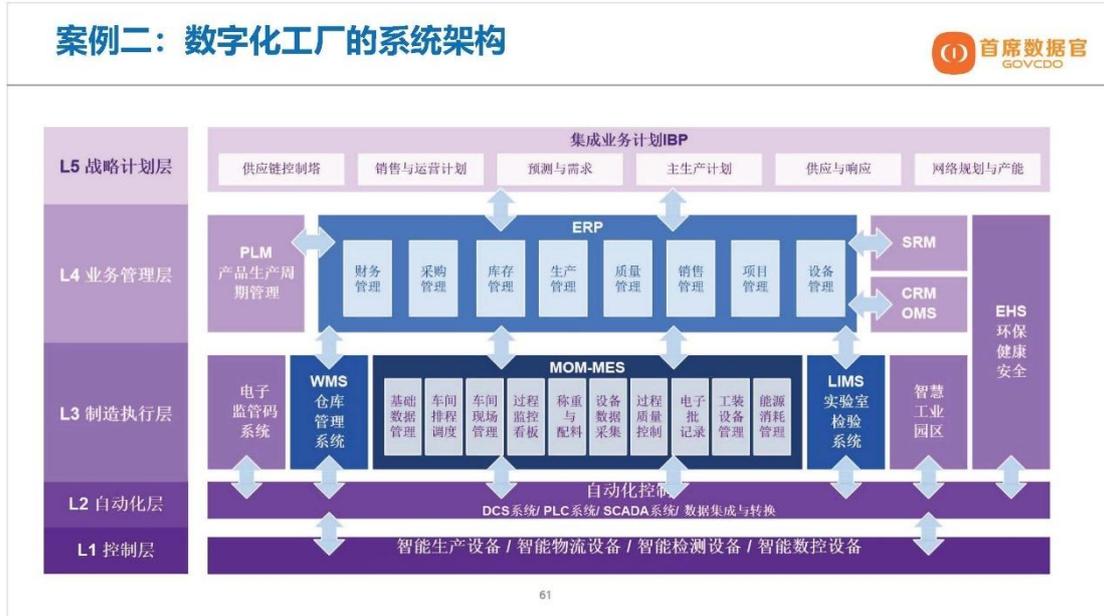
端到端业务流程和IT系统数据的梳理及打通，减少了手工录入和传递，实现了流程可视，流程执行效能显著提升：

- 供应链全程可视，提升了交付能力，欧美片区运输周期降低20%至8天，海外库房帐实准确率提升至99.9%。
- 高端运营商的服务执行过程可视，服务请求处理能力提升15%，主动服务计划能力提升50%
- 研发从客户需求到产品发布贯通，客户需求及时处理率提升到93%，产品研发周期降低15%
- 财务结帐周期缩短43%至4天，集团报表编制周期缩短80%至3天

案例二：制造业的业务架构







公司对数据能力提出全面更高的要求



数据安全规范

《中华人民共和国数据安全法》已于2021年9月1日正式施行，标志着我国数据安全保护进入有专门法律可依的新阶段。此后，配套法规体系不断完善，2024年国家网信办发布了《数据安全管理办法（征求意见稿）》，进一步优化了数据分类分级、风险评估预警、安全审查等制度的具体要求，为企业合规提供了更清晰的指引。

数据收集

收集使用个人信息需遵循的原则要求：

- 告知并公开收集使用规则
- 取得并保留同意
- 目的、范围、方式和期限一致
- 不得因未提供个人信息而拒绝提供服务

数据处理使用

数据处理使用环节需遵循的原则要求：

- 加强对个人信息和重要数据保护
- 注明来源及个人信息来源
- 合理、必要、充分、必要原则
- 推送广告信息或营销信息方式标注“定向”、“定向”等字样

数据安全监督管理

国家网信办及行业主管部门监督检查内容广泛，除个人信息保护外，还包括数据安全、数据跨境流动、数据出境安全评估、数据分类分级、风险评估预警、安全审查等制度。企业应主动接受监督检查，并依法履行义务。

法律责任

法律责任更加严厉，根据《数据安全法》和《个人信息保护法》，最高可处以营业额1%的罚款，并可吊销营业执照、停业整顿、吊销业务许可等处罚。对直接负责的主管人员和其他直接责任人员处以罚款、拘留等处罚。情节严重的，依法追究刑事责任。

跨境数据规范

在《网络安全法》、《数据安全法》基础上，《个人信息保护法》设立了个人信息出境的基本规则。国家网信办相继出台了《数据出境安全评估办法》（2022年9月1日生效）和《个人信息出境标准合同办法》（2023年6月1日生效），为不同场景下的数据出境提供了具体路径（安全评估、标准合同、保护认证）。

关键点	关键信息基础设施运营者	数据出境对象明晰
	跨境数据安全评估	重要数据境内储存

上市公司数据规范

2024年证监会修订《上市公司信息披露管理办法》，要求上市公司在招股书和年报中必须披露数据安全风险评估措施、重大数据事件及合规情况。截至2025年，证监会将数据安全纳入ESG（环境、社会、治理）评价体系，违规可能导致融资限制。

关键点	系统内控环境评估	运营数据的披露质量
	企业基础数据质量	异常交易数据分析

数据违规不良影响

监管措施

- 证券公司出具监管函
- 证券公司出具警示函
- 对公司及主管给予警告

行政处罚

- 公司及主管承担罚款
- 违法所得全部予以没收
- 责令停产停业

纪律处分

- 公司及高管面临公开谴责
- 行业内面临通报批评

立案调查

- 上市公司被迫强制退市
- 公司董监高面临市场禁入

行政处罚实例：2023年，一家知名科技公司因未履行数据保护义务，非法处理个人信息，被国家网信处以罚款8000万元，直接责任人被罚款100万元。2024，一家全国性银行因数据泄露事件，违反《数据安全法》，被罚款合计2000万元，并暂停相关业务3个月。

数据价值凸显成为企业新的收入增长极



发展阶段	2012年 探索期	2015年 启动期	2020年 发展期	2025年 爆发期
使用特征	看数/知数	统一用数	数据驱动	数智资产
业务赋能	<ul style="list-style-type: none"> 数据作为载体呈现业务现状 业务模式常态化，数据形态相对固化 	<ul style="list-style-type: none"> 业务百花齐放，各种数据平台支持数字化转型 启动统一数据平台建设，解决数据孤岛问题 	<ul style="list-style-type: none"> 数据中台支撑业务可持续发展 数据中台建设 数据治理 	<ul style="list-style-type: none"> 收入与模式创新：数据资产化运营成为企业新的收入增长极 运营与决策革新：“数智一体化”开始深入业务核心
产业特征	<ul style="list-style-type: none"> 传统信息系统产品占主流 数据产业初步萌芽，大数据产品服务相继上线 <p style="text-align: center;">互联网、金融等行业率先落地</p>	<ul style="list-style-type: none"> 大数据概念普及 数字化转型启动 头部公司开始购买数据产业 <p style="text-align: center;">技术不成熟，商业模式不清晰的数据厂商逐渐退场</p>	<ul style="list-style-type: none"> 新商业模式 新产品 新服务 <p style="text-align: center;">数据产业细分市场初现</p>	<ul style="list-style-type: none"> 市场与要素成熟：数据被正式确立为核心生产要素 生态与主体演变：产业进入“数智生态”时代 <p style="text-align: center;">大厂负责“数字基建”，垂类领域百花齐放</p>

数据为企业数字化转型保驾护航

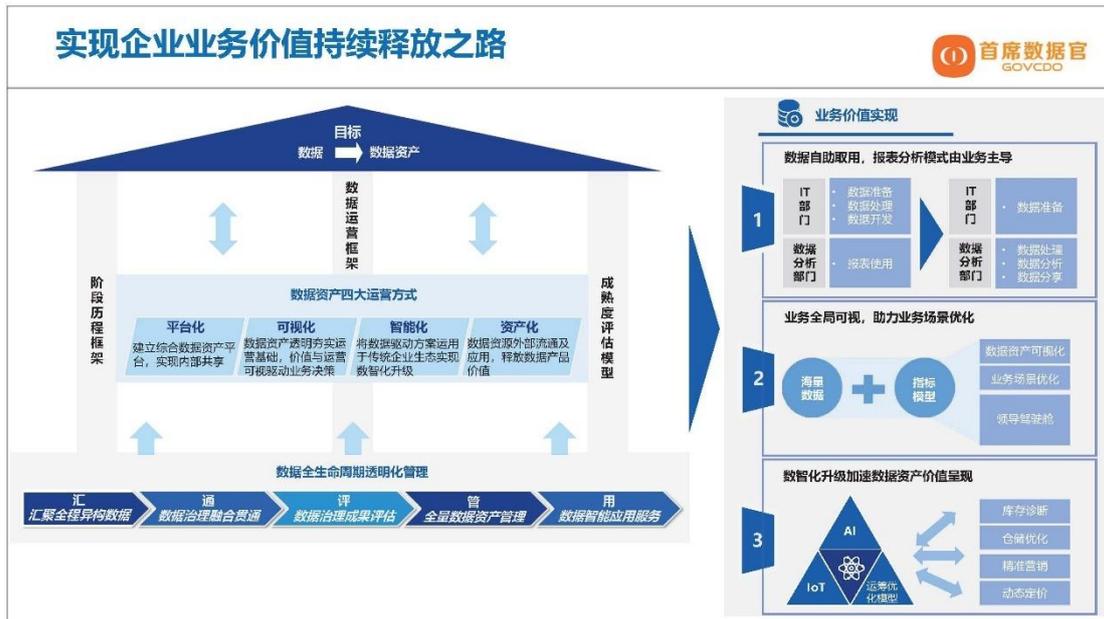


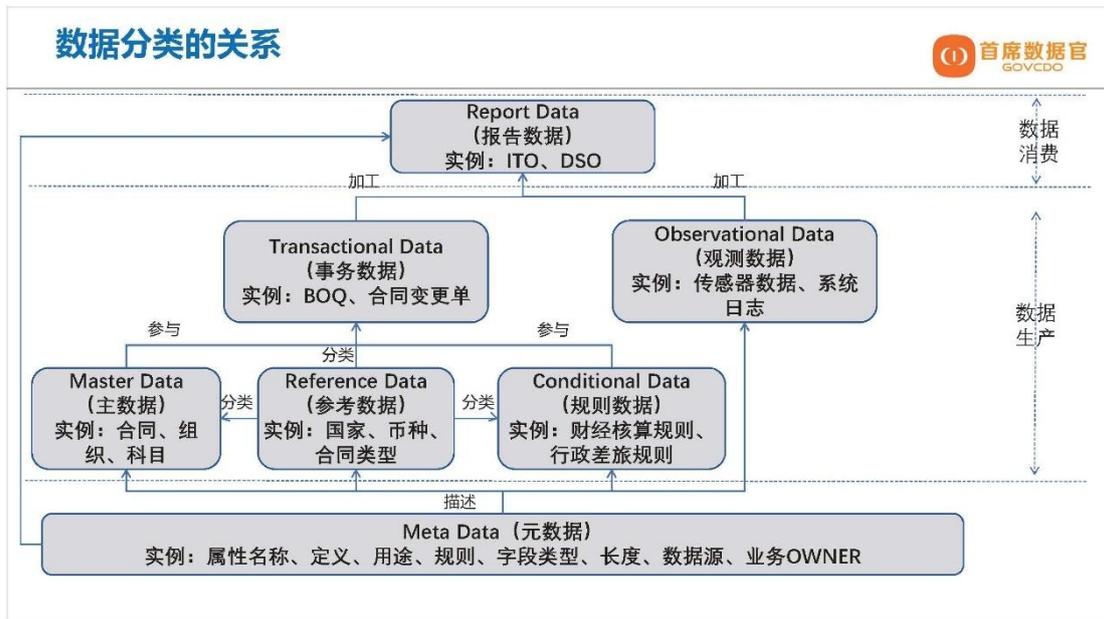
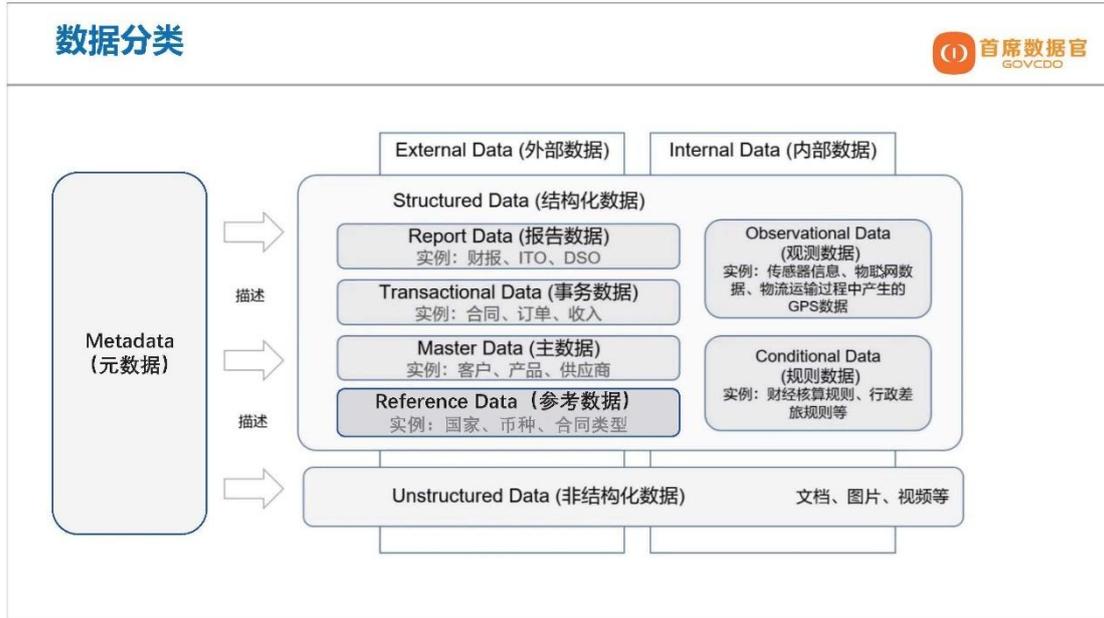
在智能化应用期，数据在商机洞察、降本增效、风险管控与辅助决策等方面为企业数字化转型保驾护航

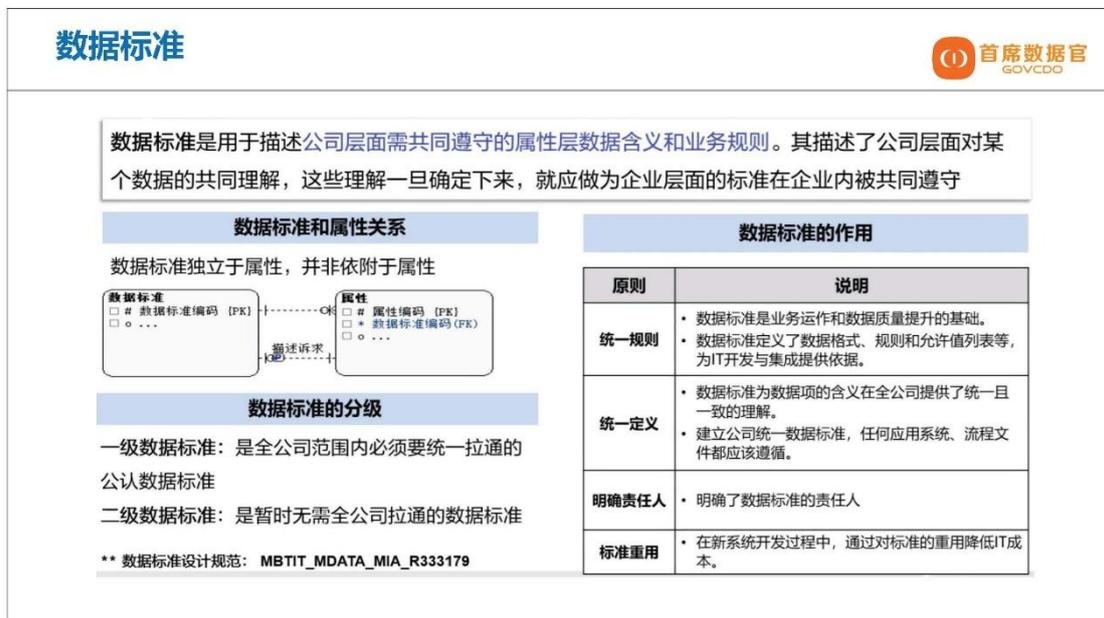
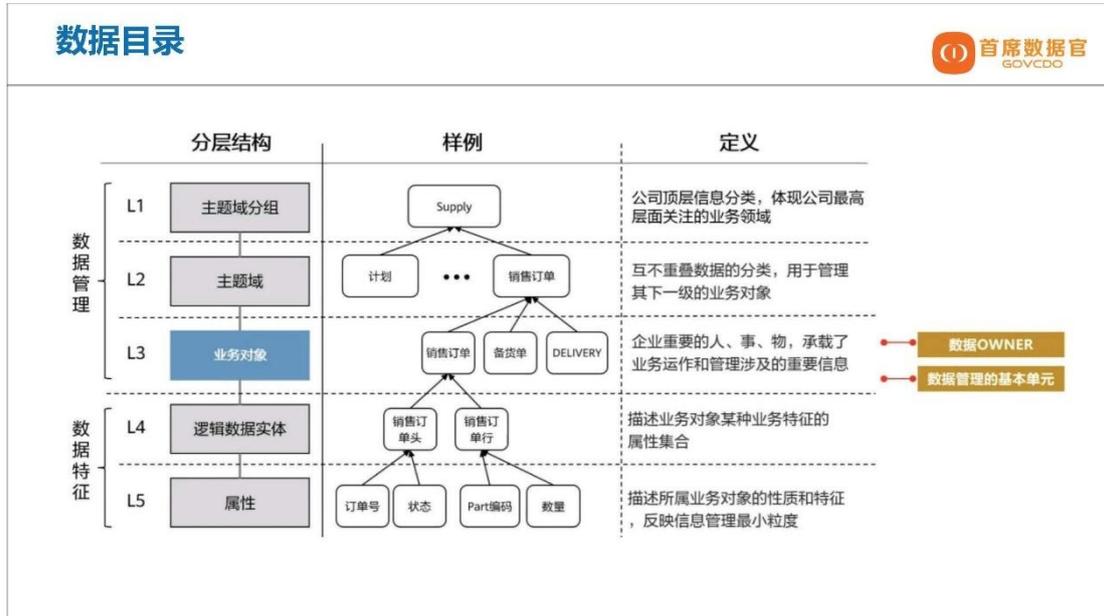


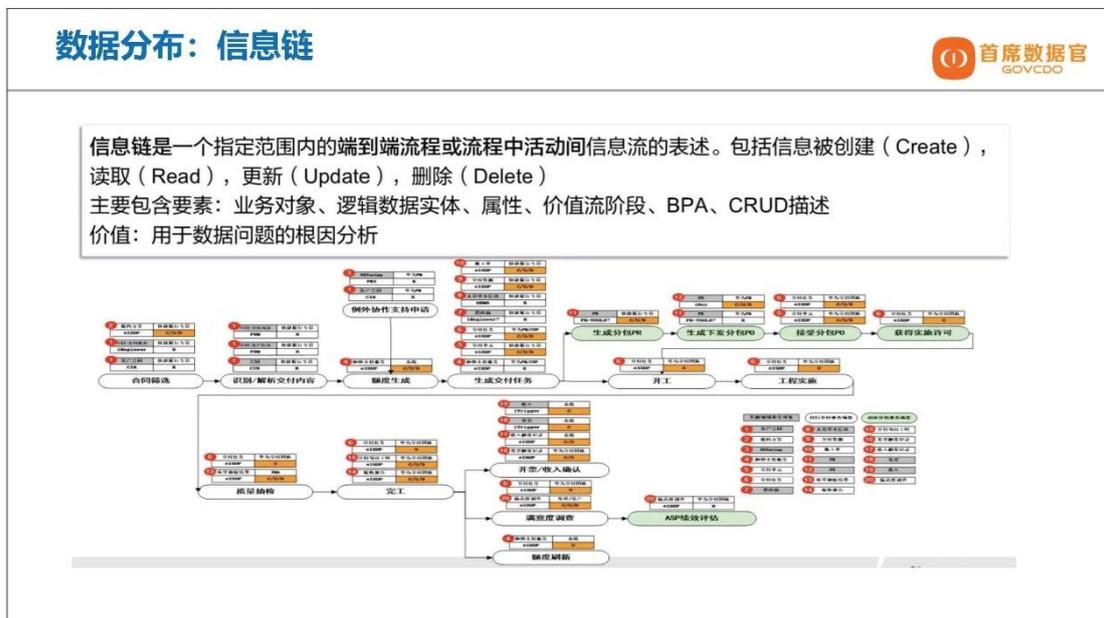
现存问题导致无法满足业务需求、呈现数据价值











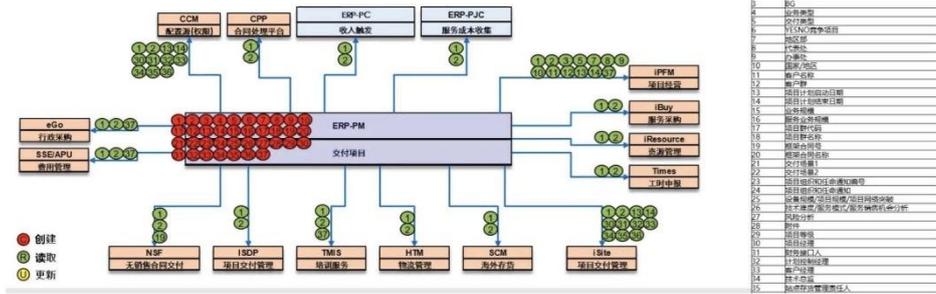
数据分布：数据流



数据流是用于描述某一数据在应用系统中的如何被创建 (Create)，读取 (Read)，更新 (Update)，删除 (Delete)

主要包含要素：业务对象、逻辑数据实体、属性、应用系统、CRUD描述

价值：用于数据质量的根因分析，指导系统的集成。



数据分布：数据源



数据源 (Data Source) 是指业务上首次正式发布某项数据的应用系统，经过数据管理专业组织认证，作为唯一数据源头被周边系统调用。

数据源的管理原则：

- 所有关键数据必须认证数据源。关键数据是指影响公司经营、运营报告的数据，通过数据字典在公司范围内统一发布。
- 数据管理专业组织为关键数据指定源头，数据源必须遵从信息架构和标准，经IA SAG(信息架构专家委员会)认证后成为数据源。
- 所有关键数据仅能在数据源录入、修改，全流程共享，其它调用系统不能修改。下游环节发现的数据源质量问题，应当在数据源进行修正。
- 所有系统必须从数据源或数据源镜像获取关键数据。
- 数据Owner确保数据源的数据质量，对不符合数据质量标准的数据源，必须限期整改。

数据资产的三基础与四步走



“在数字经济时代背景下，企业通过数据资产运营实现数据价值的重要性凸显。数据资产如何运营？其核心在于向下扎根与向上生长。

向下扎根，意味着围绕数据建立良好的**组织与意识、流程与规范、平台与工具**，以在公司内部培育数据资产运营土壤。

向上生长，意味着在深深扎根数据文化土壤的基础上，通过开展**数据资产的盘点、加工、共享与评估**，将核心业务数据资产“紧握手中”，实现数据资产稳健运营。

二者缺一不可，只有做到这两点，才能助力企业数据互联互通，释放独有价值。

三基础之组织与意识



建立完善组织架构体系，培养数据资产运营文化

组织与意识需要企业管理层的引领，构建基于数据驱动的组织模式，实现组织内人、物、知识等资源弹性供给和单元的动态协作是构建数字企业的基础，也是组织适应、利用、驾驭不确定环境的利器，数据资产运营组织架构的调整帮助企业打造一体化柔性运营管控能力，针对公司需求作出更敏捷的响应。

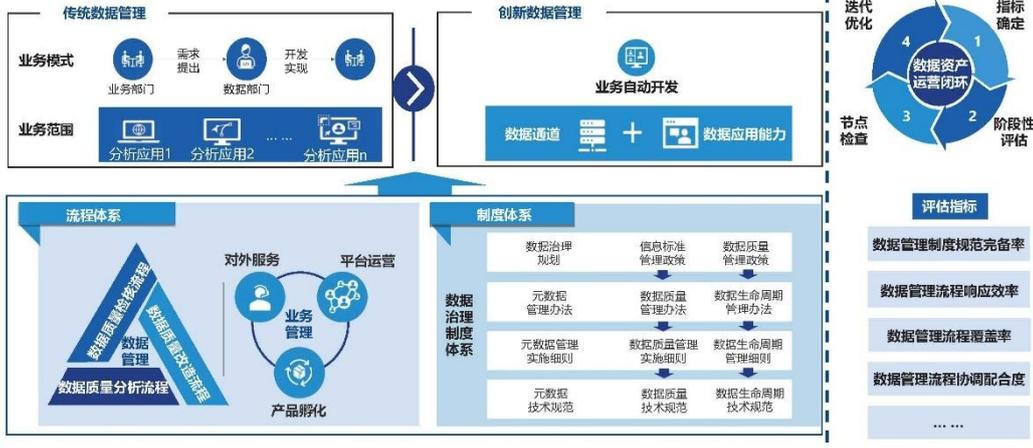


三基础之流程与规范



数据管理规范化，确保规章制度有效实施，奠定创新数据管理基础

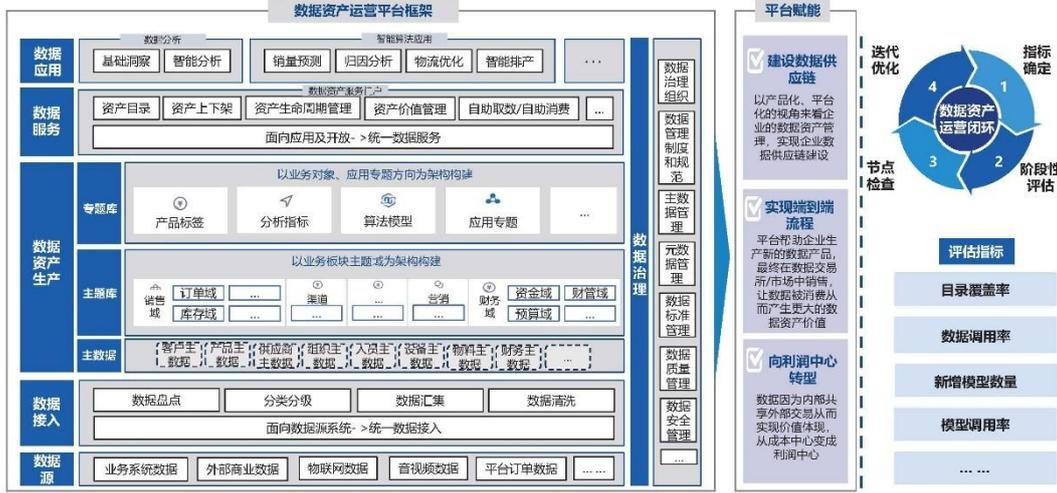
无规矩不成方圆，数据驱动型企业在创新拼搏的同时，也需对数据全生命周期进行有机管理。规范化的数据管理的流程与制度能够让数据的交互、整合与使用更加流畅，从而大大减少数据上的问题与冲突，助力公司的数据管理由传统模式健康稳定地转向创新模式。



三基础之平台与工具



平台与工具意味着生产力，是开展数据资产管理不可或缺的底层基石。通过一体化的系统框架体系，集中治理数据问题，集中进行数据监控运维与服务运营，不仅将传统数据管理工具各个组件进行了整合，更是将其进行打通与融合，实现数据在平台上的有效运转。

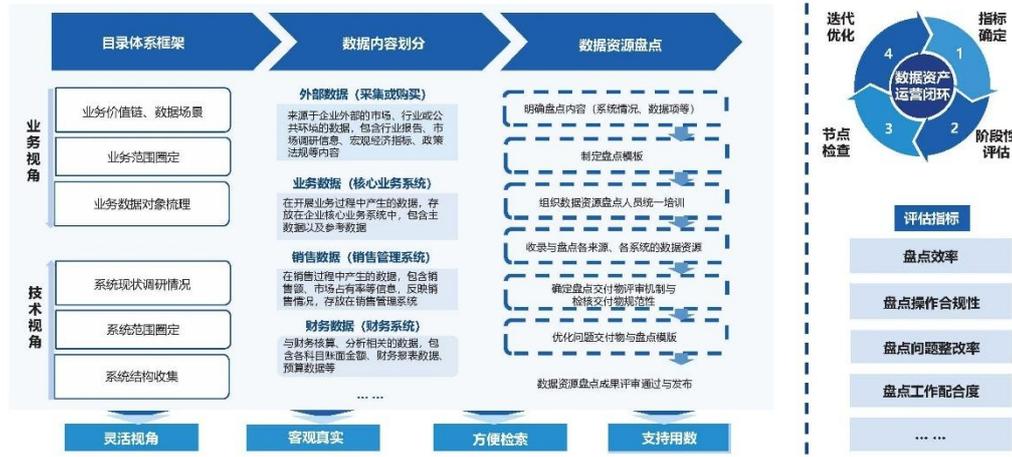


四步走之数据资源盘点



双视角厘清数据资产，绘制企业级数据资产地图

数据资源盘点是数据资产运营的先行任务，旨在解决数据资源“有什么”的问题。从业务视角与技术视角出发，形成企业级数据资产框架和数据资源目录，支持建立全面覆盖的企业级数据资源地图，为数据资产“用什么”以及“如何用”奠定基础。

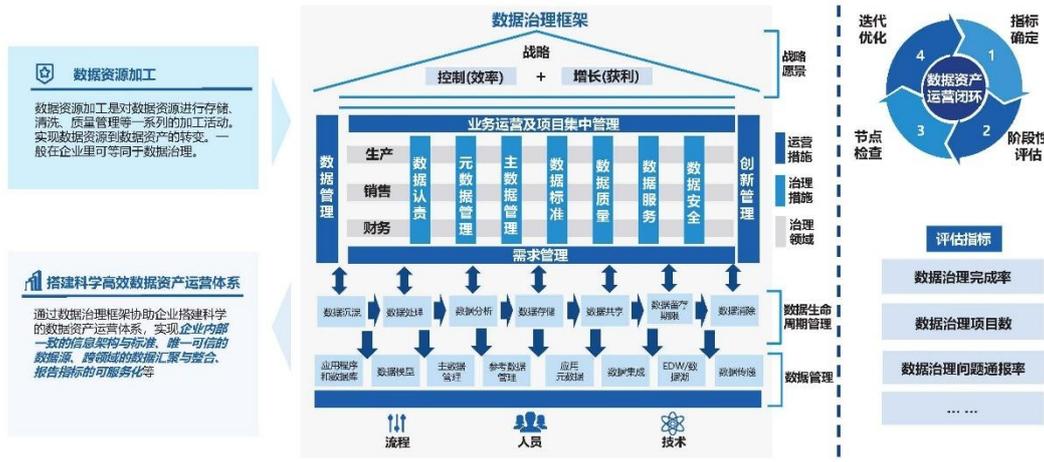


四步走之数据治理



全方位构建数据治理完整链条，奠定高质量数据基石

数据资源加工不仅仅是一套工具组合的产品级解决方案，更是从决策层到技术层，从管理制度到工具支撑，自上而下贯穿整个组织架构的完整链条，以期通过持续的评估、指导和监督，确保数据资源变成有价值的数字资产并常有成效且高效的利用，促进组织协作和结构化决策，为企业创造价值。



四步走之数据价值挖掘



全面释放数据价值

数据价值挖掘是盘活企业数据资产的有效手段。企业通过开展数据资产的内部循环与外部流通，运用数据分析与挖掘获取新的信息，在企业内部形成数据流转与共享，在企业外部通过数据产品上架提供数据交易的价值，为企业谋取创新型的收益，实现数据的增值。



四步走之数据资产评估



多维度评估企业数据资产

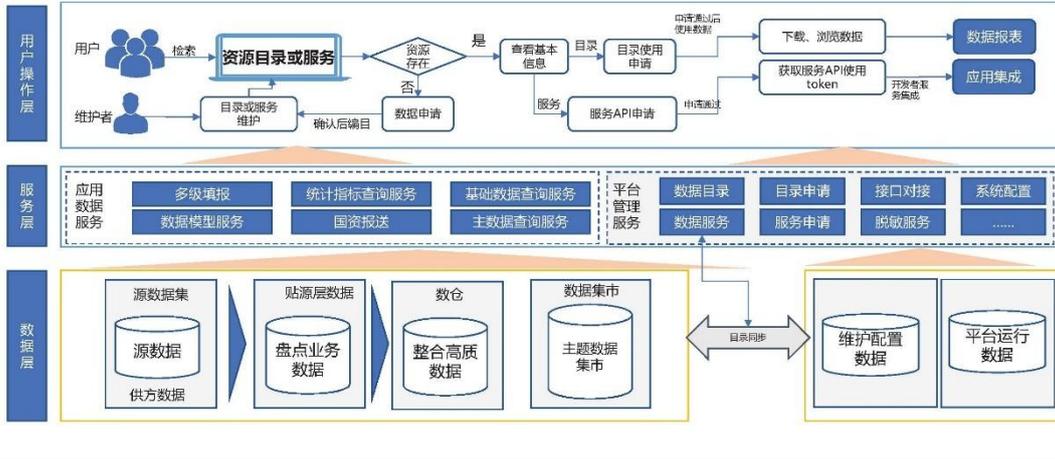
数据资产评估是推动数据要素市场化配置的关键基础。它不仅为**内部共享提供价值锚点**——通过量化数据质量与应用成熟度，明确哪些数据应优先治理、可供复用、具备业务赋能潜力；更支撑**外部流通的合规与定价**——基于多维评估模型，将高价值数据资源转化为可交易、可授权的标准产品，为数据资产入表、市场交易与创新应用提供决策依据，最终实现数据从成本到资本的价值跃迁。



应用场景：数据资产自助式消费



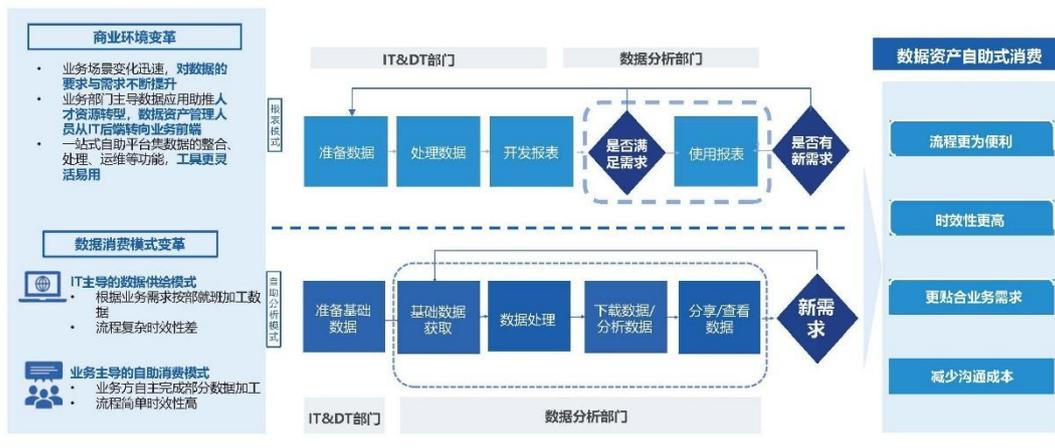
以数据目录为核心，通过对资源目录的管控、流程、配置、发布及便捷的检索服务，实现数据资源“可查”、“可管”、“可用”



应用场景：业务主导的自助分析



数据资产自助式消费最典型的场景是从“IT主导的报表模式”向“业务主导的自助分析模式”转变





国家要求-加快推进人工智能与产业的深度融合

首席数据官
GOVCD0

党和国家高度重视人工智能建设工作，人工智能是推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源。要持续推进“人工智能+”行动，将数字技术与制造优势、市场优势更好结合起来，支持大模型广泛应用。25年2月19日国资委发布了“AI+”专项行动要点，涵盖应用领航、数据脉动、模型跃升、算力基座、攻关强芯五大核心方向。



习近平总书记对人工智能发展的重要指示

- 人工智能是引领新一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的“头雁”效应，“加快发展新一代人工智能是事关我国能否抓住新一轮科技革命和产业变革机遇的战略问题。”
- 我国数据资源丰富，产业体系完备，应用场景广阔，市场空间巨大。要推动人工智能科技创新与产业创新深度融合，构建企业主导的产学研用协同创新体系，助力传统产业改造升级，开辟战略性新兴产业和未来产业发展新赛道。

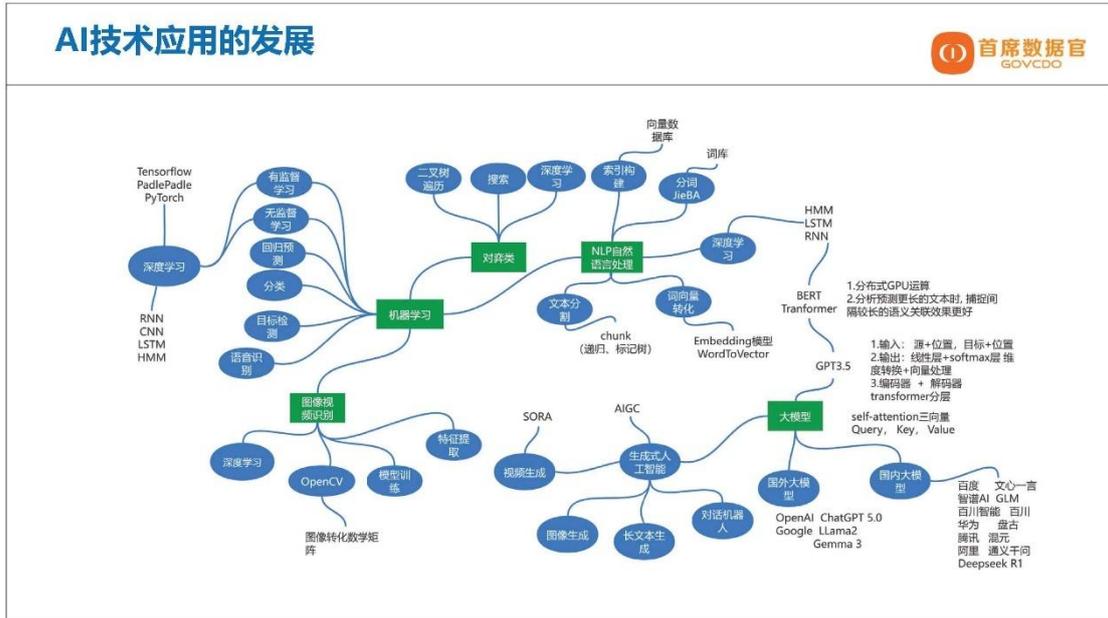
2025.3.5	2025.4.25	2025.4.26	2025.4.29
2025年政府工作报告	中共中央政治局会议	中共中央政治局第二十次集体学习	习总书记在上海考察时的重要讲话
<ul style="list-style-type: none"> 持续推进“人工智能+”行动，将数字技术与制造优势、市场优势更好结合起来，支持大模型广泛应用。 优化全国算力资源布局，打造具有国际竞争力的数字产业集群。 	<ul style="list-style-type: none"> 培育壮大新质生产力，打造一批新兴支柱产业。持续用力推进关键核心技术攻关。 加快实施“人工智能+”行动。大力推进重点产业提质升级，坚持标准引领，规范竞争秩序。 	<ul style="list-style-type: none"> 人工智能作为引领新一轮科技革命和产业变革的战略性技术，深刻改变人类生产生活方式。 人工智能领域要占领先机、赢得优势，必须在基础理论、方法、工具等方面取得突破。 	<ul style="list-style-type: none"> 人工智能技术加速迭代，正迎来爆发式发展，人工智能是年轻的事业，也是年轻人的事业。 我国数据资源丰富，产业体系完备，市场空间巨大，发展人工智能前景广阔，努力开发更多安全可靠的优质产品。

国家要求-深化部署推动央企“AI+”专项行动

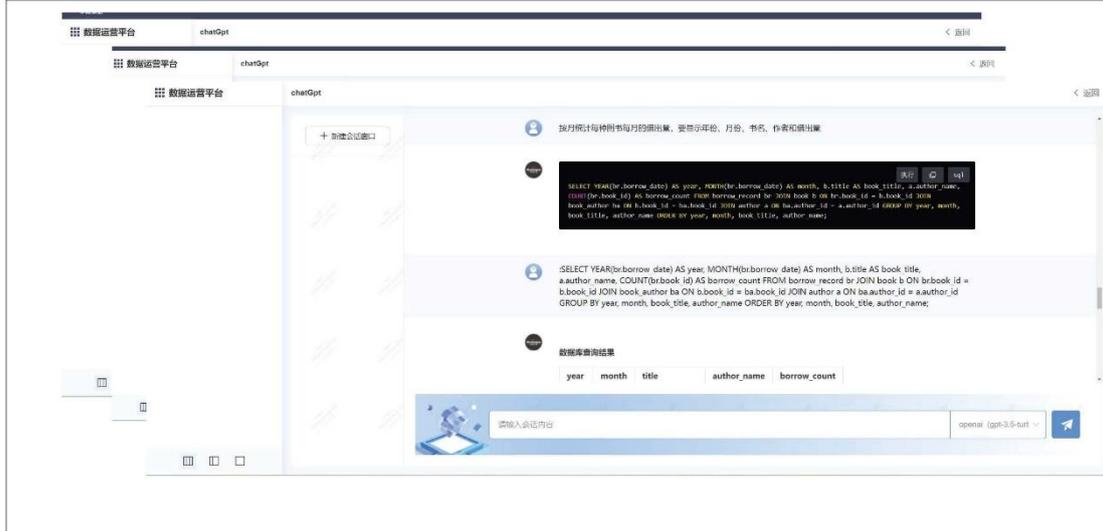
首席数据官
GOVCD0

国资委等国家部委多次要求，中央企业作为发展新质生产力的主力军，要求进一步增强责任感、使命感和紧迫感，加快推动人工智能产业高质量发展，实现更多标志性成果和突破性进展，全力塑造产业新优势、培育发展新动能。2月19日国资委发布了“AI+”专项行动要点，涵盖应用领航、数据脉动、模型跃升、算力基座、攻关强芯五大核心方向。

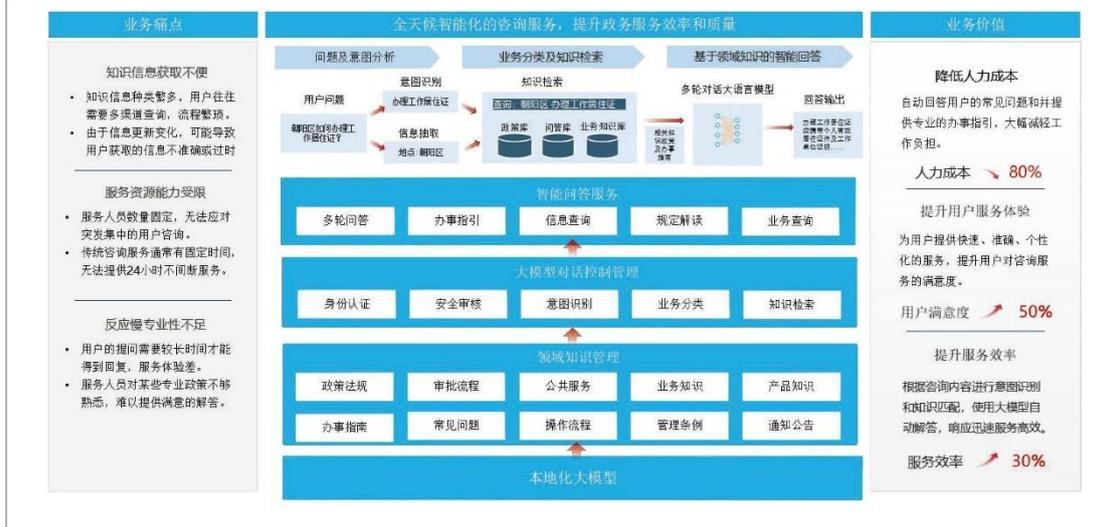
<p>2025年1月5日，中央企业人工智能特训班结业式</p> <p>紧跟人工智能最新发展，主动拥抱AI浪潮，强化人工智能思维，把握人工智能发展规律，努力成为人工智能发展的探索者、组织者、带动者。要加快学习应用人工智能，深扎人工智能“根技术”，深入实施“AI+”行动，强化人工智能赋能千行百业、驱动科技创新，让人工智能成为产业升级跃迁的强劲引擎。</p>	<p>国务院国有资产监督管理委员会</p>
<p>2025年2月19日，央企人工智能“AI+”专项行动深化部署会</p> <ul style="list-style-type: none"> 发布了“AI+”专项行动要点，涵盖应用领航、数据脉动、模型跃升、算力基座、攻关强芯五大核心方向。会议还发布了央企场景库和科技人才培养计划。 开展战略性高价值场景建设专项工作，包括垂直大模型开发、产业共同体建设以及十家央企签署任务书，聚焦关键领域。 加快推动AI高质量发展，聚焦科技攻坚，突破关键核心技术，夯实技术底座，破解数据难题，构建高质量数据资源体系。 强化要素支撑，设立创新基金，加大资金投入并优化结构，优化人才引进机制，打造高水平人工智能人才队伍。 	
<p>2025年3月28日，国资委召开国有企业经济运行座谈会</p> <p>面提升科技创新能力，推动重点产业提质升级，深化“AI+”专项行动</p> <p>聚焦新质生产力培育、产业链强基补短、基础设施建设、能源资源保障等，谋划实施一批强牵引、利长远的重大项目，有力支撑扩大内需。</p>	



智能问数：基于nl2sql的智能数据查询与数据分析



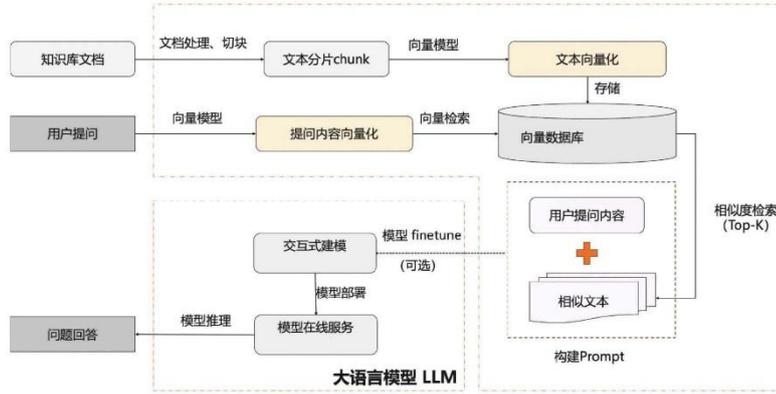
智能问答：基于领域知识和大语言模型，提供智能化问答



智能问答：RAG构建知识库



系统化建设企业知识库，实现企业数据检索，结合大模型实现检索验证和交互生成。其中涉及核心技术包括 chunk算法，Embedding向量化模型，提示词模板



智能问答：基于RAG和大语言模型，提供智能化问答



1. 多轮次知识问答
2. 按关键字极速检索文件名及文件内容
3. 单文档问答
4. 单文档摘要生成
5. 单文档脑图生产

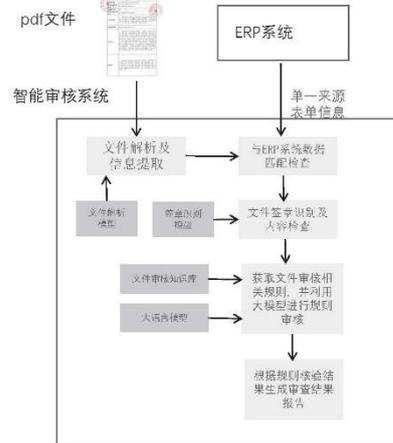
智能理解：基于规则引擎和大模型，实现对数据的自动审核



构建一个智能化、高效的在线申报审批数据审核应用。通过集成规则引擎、语义理解技术和RAG 大模型，实现对申报数据的自动化审核，提升审批效率，降低人工成本，增强数据准确性和合规性。

单一来源文件审核流程及技术实现

- 文件解析及匹配阶段：**对单一来源申报支撑文件，利用 NLP 技术对文件内容进行解析和关键信息提取，将提取的信息与 ERP 系统中的采购申请数据进行匹配。
- 签章识别审核阶段：**对文件进行图像识别，检查是否存在公章，若有则提取公章单位名称信息，记录签章审核结果。
- 规则审核阶段：**将文件解析提取结果、与文件审核知识库中相关规则，一同输入 RAG 大模型进行审核。RAG 大模型输出审核结果，判断文件是否合规。
- 结果输出阶段：**若判断合规存疑，生成详细的审核报告。报告中明确指出合规存疑的具体情况，与规则不符的内容等，并附上相关参考依据和建议，以清晰易懂的格式反馈给审批人员。



智能理解：基于自然语言理解提供政策服务



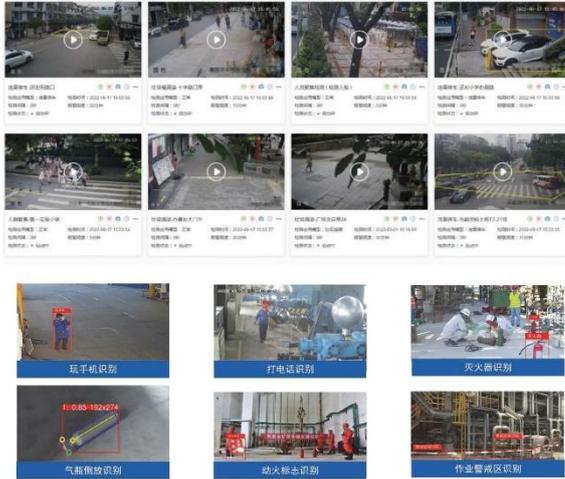
深圳市某区在2023年底创新推出的“政策AI计算器”是国内首个利用“大数据+AI算法”打造的智能化政策服务平台，通过数字化手段高质量归集全区政策补贴，构建“免申即享”“企业数据导办”“补贴事项导办”和“政策文件导办”四个功能模块，并配套“办件查询”和“人工客服”两个附属功能，实现政策“拆解、匹配、直申”全流程服务，解决企业政策申报过程中的信息不对称、流程繁琐等痛点。



视频分析：基于AI实现自动化监测



城市治理场景 + 园区企业安全生产场景

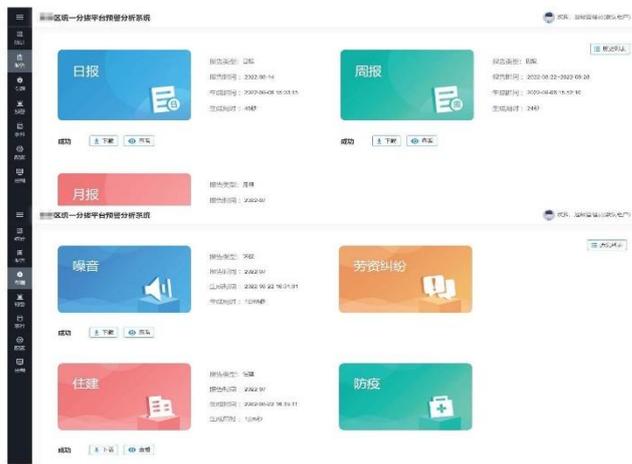


- **全天候监测** 兼容已有的前端摄像头和系统，对视频画面进行不间断分析判断，实现全天候自动化视频分析识别
- **响应及时** 非结构化数据利用率高，能快速处理海量前端数据，及时做出事中判断和预警，避免漏报
- **精确预警** 可根据实际需求定制开发高精度算法，利用强大的图像处理能力，准确识别异常状况，避免误报
- **智能管理** 支持自动报警、自定义分析区域、异常事件、视频自动存储等功能，管理更智能

智能生成：基于规则引擎和大模型，实现报告的自动生成



运用AI大模型、结构化与非结构化数据的融合治理与分析等技术，对各类数据做深度分析。结合AIGC模型及模板技术，自动生成日/周/月定例报告及专题报告，全面快速地提高基层工作人员工作效率

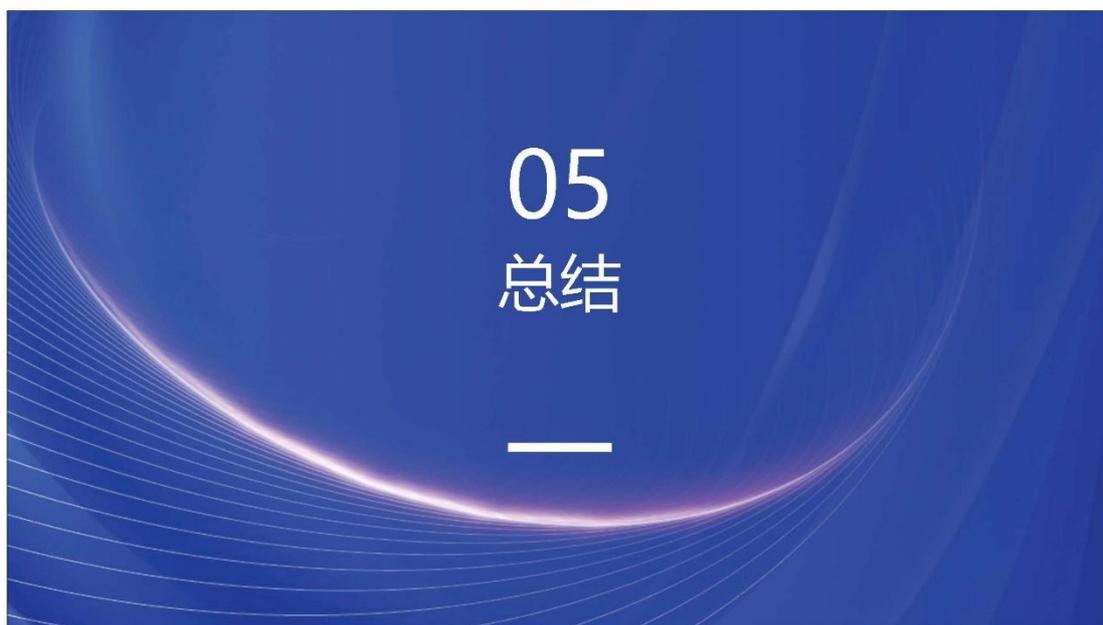
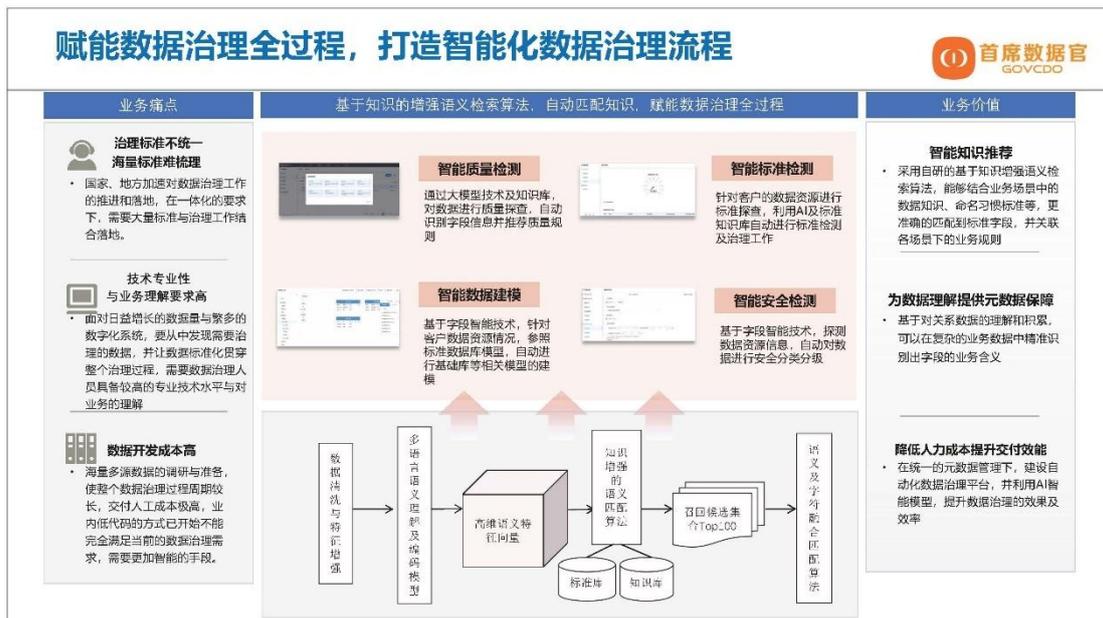


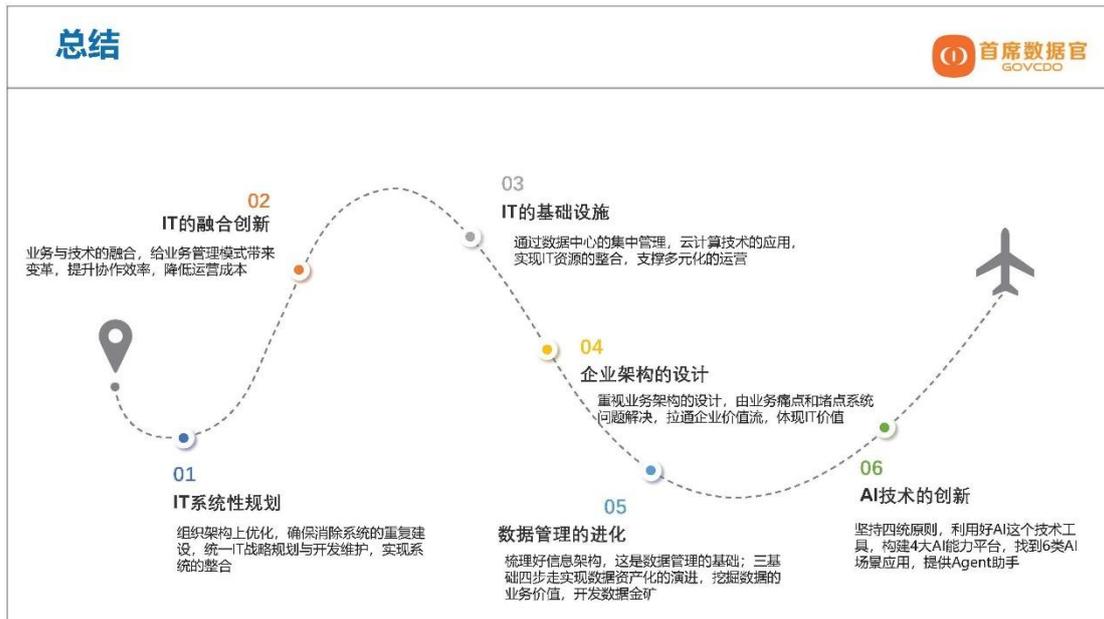
数字治理建设领导小组办公室

数字治理平台运行周报
(2022.7.11-2022.7.17)

一、整体情况
2022年7月11日至7月17日，各渠道共接收有效事件16761件，较上周环比上升9.87%，已结案13560件，结案率80.9%，办理中3056件，超期未办结145件。







数据战略与 CDO 职责

IITC工信人才岗位能力评价培训
数据产业高层次专业人才培养课程

 首席数据官
GOVCDO

数据战略与CDO职责

从成本中心到价值引擎

讲师：马欢 数字理政 2025年10月

 首席数据官
GOVCDO

CONTENTS 目录

01	破题：我们时代的变革——数据是新的石油，但如何炼油？	02	解密数据战略：它不仅仅是一份文档，而是一套行动体系
03	聚焦CDO：数据价值的“总建筑师”与“首席破壁人”	04	总结与行动指南

01

时代的变革——数据是新的石油，但如何炼油？

开场与困境共识



在座各位企业是否面临：

- 数据使用难、决策靠经验；
- 部门数据口径不一，数据分析时数据质量差
- 不知道数据如何变现
-

结论



这些数据管理问题并非个别现象
而是企业从“信息化”到“数字化”转型中普遍会经历的阵痛。

时代的机遇与挑战



机遇

详细描述：当下数据是新时代的“石油”，具有巨大价值，能够驱动企业增长、优化运营模式并促进创新，为企业发展带来了前所未有的机遇。



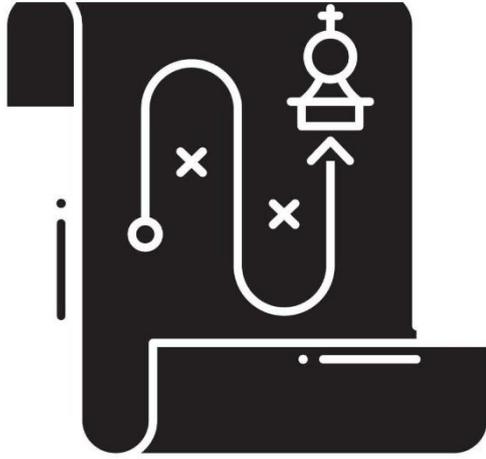
挑战

- 有数据，无资产，数据未被有效管理；
- 有技术，无体系，技术工具缺乏治理体系；
- 有尝试，无战略等问题，零散项目无法形成规模效益；



核心观点

要从“数据原油”中提炼出价值，需要一套系统的“炼油蓝图和总工程师”，也就是数据战略与CDO首席数据官。



接下来分享



- “是什么”：解密数据战略；
- “谁来做”：聚焦CDO职责；
- “怎么做”：落地框架；

STRATEGY TACTICS

02

解密数据战略：它不仅仅是一份文档，而是一套行动体系

数据战略的定义



战略定义

战略是指筹划和指导战争全局的方略。

数据战略是“将数据资产的价值最大化，以支撑企业业务战略和目标实现”的系统性规划和行动纲领。

强调要点

- 数据战略并非单纯的IT计划，而是业务战略在数据领域的具体体现。
- 它着重于将数据资产价值最大化，以此来支撑企业业务战略和目标的达成。



知名企业的数字战略



01

美的

美的的数据战略以“全面数字化、全面智能化”为指引，核心是“T+3”以销定产模式。通过整合终端销售、库存和供应链数据，将传统的压货式生产转变为以市场需求驱动的柔性生产周期。在制造端，利用工业互联网平台收集生产线数据，实现可视化管理和效率优化。在消费端，分析智能家电数据以理解用户习惯，指导产品创新和售后服务。数据成为打通用户、营销、研发与制造的生命线

02

海螺

海螺水泥的数据战略聚焦于“生产过程的智能化与绿色化”。作为传统高耗能行业，海螺通过建立全流程的数字化监控系统，实时采集生产线上的温度、压力、能耗和排放等关键数据。利用这些数据构建AI模型，优化窑炉等核心设备的运行参数，实现“智能制造”，在稳定提升水泥质量的同时，显著降低了煤耗和电耗，减少了碳排放，将数据转化为实实在在的节能效益和环保竞争力

03

华为

华为的数据战略核心是“将数据作为战略资产”，并推行“数据底座”战略。通过建立统一、清洁、透明的数据体系，打破各业务领域（如运营商、企业、消费者）的数据隔阂。该战略旨在支撑智能化的研发、生产、销售和服务，同时确保全球业务的合规与安全。华为也强调“数据入湖”，将原始数据汇聚起来，为后续的数据分析和AI应用提供高质量“原料”，驱动整个公司的数字化转型和高效运营

知名企业的数字战略



01 亚马逊

亚马逊的数据战略核心是“数据飞轮”：利用数据形成自我强化的增长闭环。通过收集海量用户行为、交易和供应链数据，驱动精准的个性化推荐、优化库存物流、并赋能AWS云服务。其战略精髓在于将数据深度嵌入每个业务环节，通过A/B测试和数据洞察持续优化用户体验与运营效率，最终反哺更多数据产生，推动飞轮加速旋转。

02 特斯拉

特斯拉的数据战略是“以现实世界为实验室”。通过其全球数百万辆行驶中的车辆，持续收集真实的驾驶环境数据（包括视频片段）。这些海量数据是其训练和验证自动驾驶AI模型的核心优势，形成了一个不断进化的“数据网络效应”：车辆越多，数据越丰富，AI模型越智能，产品吸引力就越强，从而吸引更多用户，产生更多数据。

03 西门子

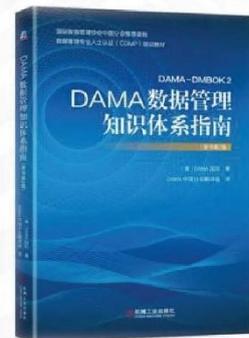
西门子的数据战略聚焦于“工业数据价值化”。通过其全面的“工业物联网”套件和XaaS（一切即服务）战略，收集全球范围内部署在工厂、楼宇和基础设施中的数百万台智能设备的实时运行数据。利用数字孪生技术，这些数据被用于在虚拟世界中模拟、预测和优化物理实体的性能，从而实现预测性维护、能效管理和产品即服务等创新商业模式，引领制造业的数字化转型。

DAMA-数据战略文件应该包括的内容



数据管理战略的内容：

- 令人信服的数据管理愿景
- 用于数据管理的业务案例概述，以及选定的示例
- 指导原则、价值观和管理观点
- 数据管理的使命和长期方向目标
- 数据管理成功的建议措施
- 符合SMART原则的短期数据管理计划目标
- 对数据管理角色和组织的描述，对其职责和决策权的总结
- 数据管理程序组件和初始化任务
- 具体明确范围的优先工作计划
- 一份包含项目和行动任务的实施路线图草案



数据战略的四大核心支柱



<p>01 组织与治理（车身与底盘）</p> <p>组织与治理确保数据战略稳定与可控，涵盖数据权责，明确谁对数据负责，还有决策体系、政策流程以及数据治理办公室等内容，如同车身与底盘为跑车提供基础支撑。</p>	<p>02 价值与赋能（引擎与燃料）</p> <p>价值与赋能驱动业务前进，包含数据应用场景如数据分析、BI、AI、数据产品，以及如何度量数据价值（Data ROI），就像引擎与燃料为跑车提供动力。</p>
<p>03 技术与平台（车轮与传动系统）</p> <p>技术与平台实现数据战略高效运转，包括数据架构、数据平台，以及工具选型，如集成、存储、处理、分析等方面，类似车轮与传动系统保障跑车的运行。</p>	<p>04 文化与素养（驾驶员与交通规则）</p> <p>文化与素养决定数据战略能走多远，内容有数据民主化、数据素养培训，建立数据驱动的决策文化，好比驾驶员和交通规则影响跑车行驶。</p>

DCMM 能力域框架



数据能力成熟度评估模型（Data Capability Maturity Assessment Model, DCMM）：针对一个组织数据管理、应用能力的评估框架，通过数据能力成熟度模型，组织可以清楚的定义数据当前所处的发展阶段以及和未来发展方向。DCMM包括8个能力域，28个能力项，对于每一个能力项进行能力等级判定。

<p>1 数据战略</p> <ul style="list-style-type: none"> • 数据战略规划 • 数据战略实施 • 数据战略评估 	<p>2 数据治理</p> <ul style="list-style-type: none"> • 数据治理组织 • 数据制度建设 • 数据治理沟通 	<p>3 数据架构</p> <ul style="list-style-type: none"> • 数据模型 • 数据分布 • 数据集成与共享 • 元数据管理 	<p>4 数据应用</p> <ul style="list-style-type: none"> • 数据分析 • 数据开发共享 • 数据服务
<p>5 数据安全</p> <ul style="list-style-type: none"> • 数据安全策略 • 数据安全治理 • 数据安全审计 	<p>6 数据质量</p> <ul style="list-style-type: none"> • 数据质量需求 • 数据质量检查 • 数据质量分析 • 数据质量提升 	<p>7 数据标准</p> <ul style="list-style-type: none"> • 业务术语 • 参考数据和主数据 • 数据元 • 指标数据 	<p>8 数据生存周期</p> <ul style="list-style-type: none"> • 数据需求 • 数据开发和设计 • 数据运维 • 数据退役

一份蹩脚的数据战略



- 愿景：“数据驱动”、“赋能”、“价值最大化”都是正确但无意义的空话，没有定义什么是“数据驱动”，也没有说明对业务的具体价值是什么。
- 目标：目标无法衡量。如何算“提升”？如何算“挖掘”？“建立基础设施”具体指什么？没有量化指标，最终无法评估成败
- 关键举措：只扔工具，不管后续。买了BI系统，但数据从哪里来？质量如何保证？谁去用？如何用？“鼓励”和“定期”是管理上的空头支票，缺乏强制性和具体计划
- 数据管理：完全没有执行细节。如何保障数据质量？数据清洗的流程和责任人是谁？安全策略和合规的具体要求是什么？一句话带过，意味着这些问题在战略层面被完全忽视
- 成功标准：“使用率”和“满意度”是极其主观的指标。只要有人打开了BI系统，就算“使用”了吗？业务部门可能因为数据太烂而不满意，但这并不妨碍他们“希望”用好数据，所以满意度调查可能依然很高。这些标准无法反映数据是否真正产生了业务影响

XX公司数据战略 (2024-2026)

- 1. 愿景**
成为数据驱动型的一流企业，通过数据赋能业务创新，实现可持续发展。
- 2. 目标**
 - 全面深化数据应用
 - 挖掘数据潜在价值
 - 构建统一数据平台
- 3. 关键举措**
 - 引入大数据分析工具
 - 推广数据决策文化
 - 开展全员数据素养培训
- 4. 数据管理**
确保数据准确性与安全性，完善数据管理机制。
- 5. 成功标准**
 - 数据应用范围显著扩大
 - 数据支持业务成效提升

一个常见的误区与纠正

误区内容

常见误区是认为“我们先建一个大数据平台，就是我们的数据战略”，这体现了典型的“技术先行”思维，没有全面考虑数据战略的系统性。

纠正逻辑

正确的逻辑是先明确业务目标，再确定数据应用场景，接着分析治理需求，最后进行技术选型。技术在数据战略中是赋能者，而非领导者。



如何建设数据战略



商业画布

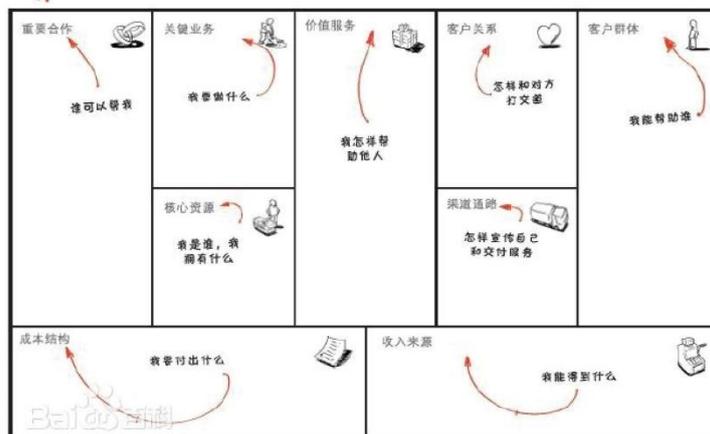
商业画布是由瑞士管理学家亚历山大·奥斯特瓦德与伊夫在《商业模式新生代》中提出的战略工具，通过九宫格结构将商业模式拆解为9个标准化模块。该工具帮助创业者以可视化方式审视商业逻辑，降低试错成本，精准定位目标客户群体，其核心价值体现在模块间动态关联的灵活调整特性，可应用于创业初期框架搭建、现有企业战略转型及个人职业规划等多场景

数据治理画布

利用商业画布的启发，通过一系列的画布，在不同的沟通层次上来构建和展示数据战略。



商业画布的概述



数据战略的分解



Copyright © 2023 María Guadalupe López Flores, Servicios de Estrategia y Gestión de Datos Aplicada, S.C., segda.com.mx

数据战略 - 业务一致性战略



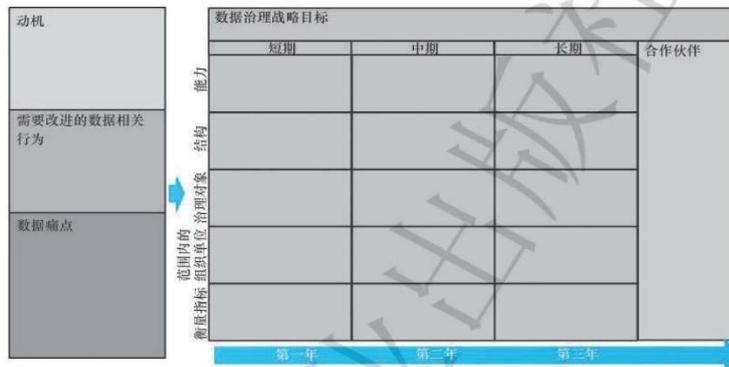
Copyright © 2023 María Guadalupe López Flores, Servicios de Estrategia y Gestión de Datos Aplicada, S.C., segda.com.mx

数据战略 - 数据管理战略



Copyright©2023 Maria Guadalupe López Flores., Servicios de Estrategia y Gestión de Datos Aplicada,S.C., segda.com.mx

数据战略 - 数据治理战略

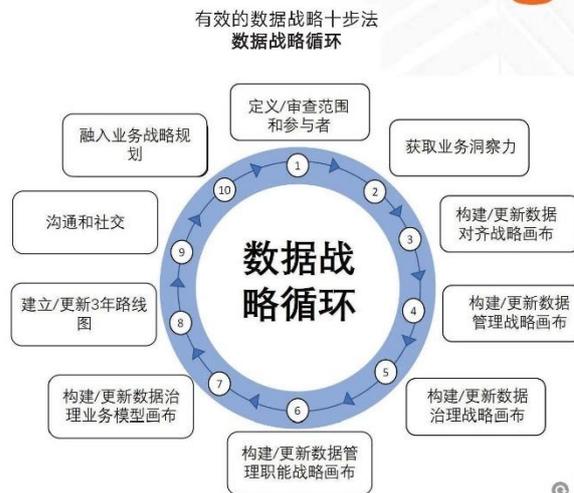


Copyright © 2023 Maria Guadalupe López Flores., Servicios de Estrategia y Gestión de Datos Aplicada, s.c, segda.com.mx

数据战略 – 数据管理各职能战略



构建数据战略十步法



03

聚焦CDO：数据价值的“总建筑师”与“首席破壁人”



CDO的角色演变：从“数据警察”到“价值伙伴”



过去（数据警察）

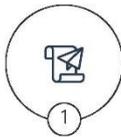
过去CDO侧重合规、风险控制、制定规则，但这种角色定位使其易被业务部门视为“障碍”，未能充分发挥数据价值，更多是在规范数据使用。



现在与未来（价值伙伴 & 产品经理）

现在与未来，CDO核心职责是催化数据价值，像产品经理一样经营“数据”产品，为内部“客户”即业务部门赋能，以更好地推动企业数据业务发展；以及进行数据资产化的探索，让数据为企业创造更大的价值，甚至带来企业业务的转型。

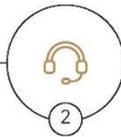
CDO的五大核心职责



1

定战略与路线图

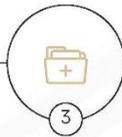
CDO需承接业务战略，制定并主导执行数据战略。例如结合企业整体业务方向，规划出数据战略蓝图和执行路径，确保数据工作与业务目标紧密结合。



2

建体系与保质量

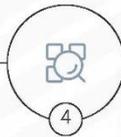
CDO要建立企业级数据治理框架，提升数据可信度。如明确数据权责、决策体系等，保障数据在整个企业内的准确、可靠，避免数据混乱和错误。



3

促赋能与创价值

CDO需与业务部门合作，孵化数据应用项目，并量化其业务价值。通过挖掘数据潜力，为业务提供支持，如利用数据分析优化业务流程，提升业务效益。



4

管资产与控风险

CDO要对数据资产进行盘点、估值，并确保数据安全与合规。准确掌握企业数据资产状况，防止数据泄露和违规使用，保障数据资产的安全。



5

塑文化与提素养

CDO要在企业内部推广数据文化，提升全民数据技能。营造重视数据的氛围，让员工具备运用数据进行决策和工作的能力，推动企业数字化转型。

(1) 定战略与路线图 - 制定与推动数据战略



战略规划

制定与公司整体业务战略紧密对齐的数据战略，明确数据投资的优先顺序和未来愿景，为企业数据应用指引方向。

价值赋能

识别数据直接赋能业务的途径，驱动收入增长、优化运营效率及创新商业模式（如数据变现）。

路线图管理

主导数据战略实施路线图，确保各项举措分阶段、可衡量地落地，保障战略有效执行。

案例：Netflix的业务战略是用户增长和内容成功



将用户数据融入内容生命周期

Netflix的CDO把用户数据深度融入内容生命周期，从投资阶段就通过分析用户数据评估潜在受众规模，为内容投资决策提供依据，让数据在内容的全流程发挥作用。



提升用户粘性和满意度

利用数据构建强大的推荐算法，为用户提供极度个性化的内容展示，显著提升了用户粘性和满意度，降低了取消订阅率，有助于实现用户增长的业务核心。



指导内容创作与营销

在内容创作阶段，数据指导优化剧情走向、选择演员；营销阶段，为不同圈层用户生成不同的预告片 and 海报，助力内容成功，推动业务发展。

结论：Netflix的CDO将数据战略锚定在公司核心业务上，使数据成为颠覆传统娱乐行业的“燃料”

数据战略锚定核心业务

Netflix的CDO没有仅仅管理数据，而是将数据战略直接锚定在公司的核心业务，即内容与增长上，让数据在各业务环节发挥关键作用。

数据成为颠覆行业“燃料”

通过将数据战略与核心业务紧密结合，数据在投资决策、用户体验、内容创作营销等方面发挥作用，成为Netflix颠覆传统娱乐行业的重要“燃料”。



(2) 建体系与保质量 - 建立与运营数据治理体系



政策与标准

建立企业级数据治理框架，涵盖数据标准、质量规则、安全政策、隐私合规和伦理指南，确保数据“可用、可靠、可控”。



组织与流程

设立数据治理委员会，明确数据所有者、管理者和使用者职责，建立配套流程保障治理工作有序开展。



生命周期管理

监督数据从创建、存储、整合到归档/销毁的全生命周期管理，保障数据完整性、安全性和可用性。

案例：某银行推进“客户360度视图”项目时数据分散问题

首席数据官
GOVCDO



数据分散难题

- 一家大型银行在推进“客户360度视图”项目时，发现客户数据分散在信用卡、储蓄、贷款等多个部门，同一客户的姓名、联系方式不一致，导致数据无法整合，影响项目推进。
- CDO推动成立数据治理委员会，指定各业务部门负责人为“数据所有者”，让其对各自领域的的数据质量负责，通过这样的流程保障数据在各业务环节的质量与合规。
- CDO牵头制定全行统一的客户主数据标准，明确客户ID、姓名、地址等关键信息的格式和规范，为数据的高质量和合规性奠定基础。

结论：银行通过有效治理使数据从“负债”变成“资产”

首席数据官
GOVCDO



实现精准营销

通过有效的治理，银行成功构建“客户统一视图”，能够基于准确的数据对客户进行精准营销，挖掘客户潜在需求，提高营销效果，让数据发挥出更大价值。

促进交叉销售

银行利用有效治理后的数据，实现对客户的全面了解，进而开展交叉销售，向客户推荐更适合的金融产品和服务，增加业务收入，使数据从“负债”变为“资产”。

加强风险管理

有效治理后的数据为银行的全面风险管理提供了有力支持，银行可以根据准确的数据评估客户风险，提前采取措施防范风险，保障业务的稳健发展，体现了数据的重要价值。

(3) 促赋能与创价值 - 领导数据架构与平台建设



技术规划

主导未来数据技术栈的选型与规划（如数据湖仓、数据管道、BI工具），为数据应用打造强大“底座”。



平台建设

推动构建统一、可扩展、安全的数据平台，打破数据孤岛，实现数据互联互通。



赋能创新

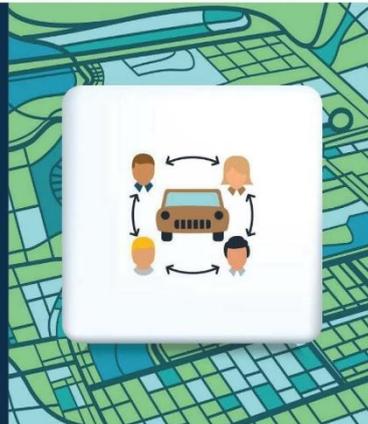
为业务部门和数据科学家提供易用的数据工具和环境，促进数据分析和AI应用，激发创新活力。

案例：Uber早期数据处理效率低下问题



海量数据处理难题

- Uber的业务会产生海量的实时位置、行程、支付等数据，早期，数据工程师和科学家们需要花费大量时间在不同系统中寻找和准备数据，导致处理效率十分低下。
- Uber的CDO与CTO协作，主导建设了“Uber数据仓库”的数据平台，为内部用户提供了强大的数据支持，提升了数据处理和利用的效率。该平台提供了标准化的数据接入、转换和查询工具，使得业务分析师和数据科学家可以自助地访问高质量、可信的数据，无需关心底层基础设施的复杂性，提高了工作的自主性。
- 这个强大的数据平台极大地加速了动态定价算法、路线优化、欺诈检测等各类数据产品的开发和迭代速度，推动了业务创新发展。



结论：CDO构建数据基础设施解放生产力，聚焦价值创造

首席数据官
GOVCDO



解放生产力

CDO通过构建一流的“数据基础设施”，就像为工程师提供了自来水厂，让他们无需自挖水井，从而解放了生产力，使员工能将更多精力投入到有价值的工作中。

聚焦价值创造

有了完善的数据基础设施，员工可以摆脱繁琐的数据准备工作，聚焦于价值创造，推动公司业务的持续发展 and 创新。



(4) 管资产与控风险 - 确保数据合规、风险与伦理

首席数据官
GOVCDO

合规监管

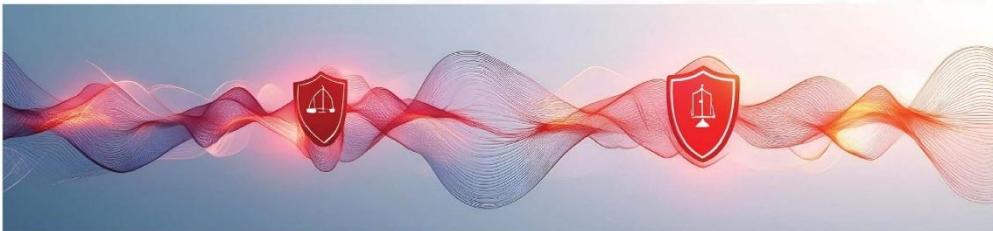
确保企业数据实践符合国内外的GDPR、《数据安全法》《个人信息保护法》等法规，保障企业运营在法律框架内。

风险管理

识别和评估数据安全、隐私泄露等风险，建立应对机制降低风险损失，保障数据安全稳定。

数据伦理

制定负责任的数据使用原则，防止数据滥用和算法偏见，维护企业声誉和社会责任。



案例：某中国电商公司计划拓展欧盟市场



合规评估

在该中国电商公司计划拓展欧盟市场时，CDO主导对公司数据全流程进行GDPR合规差距分析，查看是否符合欧盟关于用户同意、被遗忘权等严格规定，为公司在欧盟的业务开展筑牢合规基础。

制定策略

基于合规评估结果，CDO制定并实施一系列合规策略，如修改全球网站隐私政策，建立处理用户数据删除请求流程，对涉及欧盟用户数据的员工进行专项培训，确保公司数据处理符合欧盟法规。

伦理考量

在为欧盟用户进行个性化推荐时，CDO牵头制定伦理准则，防止公司利用数据对用户进行过度操纵，避免出现“暗黑模式”或价格歧视等争议性问题，维护公司良好形象。

结论：CDO避免公司因合规问题受损，为业务出海保驾护航



避免合规风险

CDO通过对公司数据实践进行全面的合规评估和制定有效策略，避免了公司因不符合欧盟法规而面临巨额罚款和声誉受损的风险，保障了业务在欧盟市场的平稳开展。



保障业务出海

CDO不仅确保公司数据实践符合法规，还在伦理层面进行考量，为公司业务出海提供了全方位的保障，使公司能够顺利拓展欧盟市场，实现业务的国际化发展。

(5) 塑文化与提素养 - 培育数据文化与提升数据素养



文化推广

在全公司倡导“数据驱动决策”的文化，取代凭经验决策习惯，推动企业向数据驱动型转变。



人才培养

设计并实施全员数据素养提升计划，为不同岗位员工提供针对性数据技能培训，提升数据处理与运用能力。



沟通协作

作为业务与技术之间的桥梁，确保双方对数据的理解和目标一致，促进数据应用顺利推进。

有家公司搞了一场内部培训

奖学金发了 150 万

咋回事儿?

一起来看

1、哪家公司这么牛？

听说是三一集团



6天
培训时长
脱产集中学

600+
参训人员
全体干部学

12个
全国会场
云端连线学

150万
现金奖励
高额悬赏学

2、什么内容这么牛？

培训主题是 DAMA



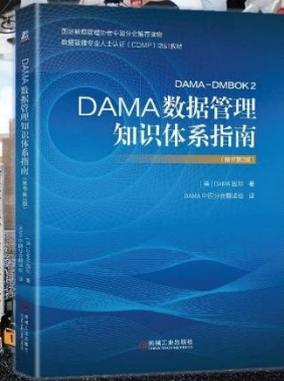
DAMA 英文全称 **Data Management**，是国际数据管理协会（Data Management Association）的昵称。国际协会也叫DAMA国际（DAMA International），是一个全球性数据管理和业务专业志愿人士组成的非营利协会，致力于数据管理的研究和实践。

DAMA国际通过对业界数据管理最佳实践的分析总结，出版了《**DAMA-DMBOK2数据管理知识体系指南（第2版）**》，DAMA中国组织专家翻译并于2020年出版了中文版。

3、培训教材长啥样？

教材是 DMBOK2

《DAMA数据管理知识体系指南》是目前市场上唯一一部综合了数据管理方方面面的权威性著作，是**数据治理的标准工具书**。
三一集团为了**全面提升数据管理和治理能力**，选它作为培训教材，豪爽地发放了**150万元奖学金**！！



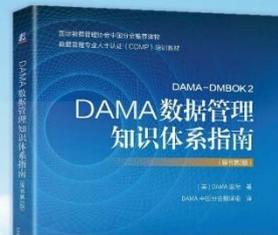
4、现在学它有啥用？

看看老大怎么说（续）

创建一流企业、造就一流人才、做出一流贡献。在三一数智化转型的路上，这场学习，或将成为一个伟大的起点。

学得好的同学，上台领钱！

——梁稳根（三一集团董事长）



对CDO角色的期待



治理与价值并重

作为“治理官”管控数据风险，建立治理体系确保数据“可用、可靠、可控”；作为“价值官”创造收益，驱动数据战略实现业务价值。



技术与业务融合

掌握数据技术，领导架构与平台建设；更懂业务需求，确保数据战略与业务战略紧密对齐，驱动业务发展。



规则与赋能平衡

制定数据治理规则与标准，设立治理委员会；同时为业务提供易用工具与环境，培育数据文化，赋能创新。



数据资产化

将数据从待管理的成本转变为核心战略资产，通过战略规划与价值实现，驱动企业增长和创新。

卓越CDO的“三栖”能力模型



业务专家

卓越CDO必须懂业务，能翻译业务需求为数据方案。深入了解企业业务流程和需求，将业务问题转化为数据可解决的问题，为业务提供针对性的数据支持。



技术通才

卓越CDO要了解数据技术趋势与架构，能与IT团队高效沟通。掌握各类数据技术，与IT人员共同规划和实施数据项目，确保技术方案符合业务需求。



管理领袖

卓越CDO需具备强大的影响力、沟通能力和变革管理能力，能打破部门墙。协调各部门之间的数据使用和合作，推动企业数据战略的顺利实施。

首席数据官CDO管理手册



第一部分：从职能角度介绍如何打造一个合格的数据办公室，包括现状调研、数据供应链、数据愿景、数据使命和数据战略、数据治理、数据质量、主数据、数据的组织角色、以及数据办公室的职责等传统数据管理的专业内容；

第二部分：从心理学角度介绍首席数据官应该具备和了解的一些软技能，包括应对 CDO 的挑战、利益相关者分析、治理心理学，以及 CDO 的常见误区等；

第三部分：介绍数据管理实践的一些新兴话题，如数据伦理、外部数据、数据分析、危机中的数据管理、并购中的数据，以及数据创新等内容。

04

总结于行动指南



核心观点回顾



- 01 数据战略是一套行动体系，而非单点技术**
数据战略是将数据资产价值最大化，支撑企业业务战略和目标实现的系统性规划和行动纲领，它是业务战略在数据领域的延伸，并非单点技术，有四大核心支柱。
- 02 CDO是价值创造者，而非仅仅作为规则守护者**
CDO的角色已从过去侧重合规、风险控制的“数据警察”，演变为如今催化数据价值的“价值伙伴”，核心职责是为业务部门赋能，创造价值。
- 03 成功 = 战略的清晰度 × 组织的执行力**
在数据战略实施中，战略的清晰度与组织的执行力共同决定了最终的成功与否，只有清晰的战略和强大的执行力结合，才能实现数据从成本中心到价值引擎的转变。

给企业的三条行动建议



-  **从最痛的业务场景出发，定义一个“速赢”项目，证明数据价值**
企业可从自身最棘手的业务场景入手，定义一个能快速取得成果的“速赢”项目，通过该项目展现数据的价值，为后续的数据战略实施奠定基础 and 信心。
-  **尽快启动数据战略的规划，哪怕从1.0简单版本开始**
企业应尽快开启数据战略的规划工作，即便最初版本较为简单，也能为后续发展指明方向，逐步完善数据战略体系，避免陷入“有尝试，无战略”的困境。
-  **尽早任命或培养一位具备“三栖”能力的CDO或数据负责人**
企业要尽早任命或培养一位集业务专家、技术通才、管理领袖能力于一身的CDO或数据负责人，其能有效推动数据战略的制定与执行，打破部门壁垒。

结束语与展望



企业需成为数据驱动组织，否则将被淘汰

在数字化时代，数据的重要性日益凸显，企业若不能成为数据驱动的组织，将难以适应时代的发展，最终可能被市场淘汰，所以要重视数据战略和CDO的作用。

绘制“炼油蓝图”，找到“总工程师”

希望企业通过本次分享，能更好地规划数据战略，绘制属于自己的“炼油蓝图”，并找到像CDO这样的“总工程师”，将数据转化为驱动企业发展的价值引擎。



05 互动问答



自问自答



如果公司暂时没有设置CDO的打算，数据工作该如何启动？

若公司暂不打算设置CDO，可从最痛的业务场景出发，定义“速赢”项目证明数据价值。同时，尽快启动数据战略规划，哪怕是简单版本，先让数据工作运转起来。



如何衡量数据战略的ROI（投资回报率）？

衡量数据战略的ROI可从数据应用场景入手，关注数据分析、BI、AI等带来的业务效益。通过量化数据产品的业务价值，如成本降低、收入增加等，来评估数据战略的投资回报率。



THE END
谢谢

数据资产（产业）全过程

IITC工信人才岗位能力评价培训
数据产业高层次专业人才培养课程

首席数据官

数据资产（产业）全过程

数据资产（产业）全过程管理，**重点围绕**数据资源治理、高质量数据集、数据资产入表、台账编制、登记、运营授权、收益分配、交易流通等环节开展；**难点在于**深度认知数据资产属性特点和价值运营规律；高起点谋划、高质量落实，**以点带面**推动数据资产管理提质增效，释放数据价值。

讲师：王毅 数字理政研究院IITC工信人才首席数据官讲师

2025年10月·湖北·武汉

目/录 CONTENTS

01	02	03
政策背景	技术思路	实践案例

01 政策背景

<h2>宏观政策概念和指导思想</h2>	<p>首席数据官</p>
	<p>中央两办、财政部、国家数据局 自2022年底“数据20条”出台以来，我国数据资源治理、数据资产入表及公共数据开发利用初显成效，数据成为继土地、劳动力、资本、技术之后第五大新型生产要素，推动经济社会创新发展，是新质生产力的关键要素。</p> <p>政府工作报告 2025年两会期间，政府工作报告强调完善数据基础制度，加强数据资产开发利用，规划2025年初至2026年底开展数据资产全过程管理试点。</p> <p>全过程试点范围 财政部联合国数局牵头，综合选取7个中央部委、6家央企及16省市财政部门，重点探索数据资产管理模式，完善相关制度、标准及运行机制。</p> <p>省级政策落地 截至2025年7月，已有7省2市财政部门印发政策文件推进数据资产全过程管理，强化公共数据开发利用，结合地方实际创新试点，形成可复制推广的经验。</p>

数据资产化项目概念

首席数据官

数据资产化项目

专为数据资源密集型机构定制，涵盖盘点、梳理、目录建设、入表、价值评估及开发利用等全流程技术运营服务。



项目目标

通过系统化的技术运营服务，推动数据资产化进程，有效释放数据价值，为机构量身打造新质生产力，提升核心竞争力。



受益对象

适合政府、企业、金融、电信、能源等千行百业广泛领域，助其高效管理数据资源，实现数据资产化，提升运营效率，增强市场竞争力。



项目价值

数据资产化项目不仅促进数据资源有效利用，释放数据价值，还推动产业升级与经济发展，为实施机构带来显著的经济效益与社会价值。



6部委（无财政）数据产业高质量发展意见

首席数据官

数据产业定义

支撑新质生产力，数据产品或服务开发所形成的新兴产业，包括数据采集汇聚、计算存储、流通交易、开发利用、安全治理和数据基础设施建设6个方向；到2029年复合增长率超过15%，数据产品和服务供给能力大幅提升。

产业规划布局

- 1、结构：核心技术领域突破，数据科技高水平自立自强；产业关键环节培育新业态、协同发展、生态能力；
- 2、产业链：大力培育创新性中小数据企业，支持专业化、精细化发展，打造数据产业创新联合体，生态体系；
- 3、区域聚集：区域资源禀赋，用数、用云、用电、用地便利政策，发挥地方特色；
- 4、重点方向：数据采集、数据存储、数据治理、数据分析、数据交易、数据应用、数据安全、数据基础设施；

多元经营主体

- 1、数据资源企业：支持开发利用，培育行业性数据资源企业，按市场化方式“授权使用数据、共同分享收益”；
- 2、数据技术企业：支持面向数据产业的技术创新，原创性引领数据科技创新发展；
- 3、数据服务企业：面向流通交易专业服务，业务咨询、交易撮合、合规评估、金融服务，新模式新业态；
- 4、数据应用企业：面向“智改数转”、全域数字化，创新应用模式，发挥数据价值，满足行业需求，赋能产业发展；
- 5、数据安全企业：面向数据大范围、高速度、高流通量，研发智能化数据安全产品；
- 6、数据基础设施企业：聚焦一体化算力、数据空间、低代码平台，培育竞争力的基础设施企业；

数据技术创新

- 1、关键核心技术：云边端计算能力、可信存储技术，数据采集、治理、应用智能技术，标注、合成、区块链、隐私计算；
- 2、创新支撑能力：数据技术科学实验室、企业技术中心、创新联合体、开源治理生态、社区等。

开发利用水平

- 1、扩大数据资源供给：公共数据供给、企业数据供给，城市数据空间、行业数据空间，跨行业领域共享、个人数据开发利用；
- 2、大力推动应用创新：政府、行业龙头开放场景，工业、农业、商贸、金融领域“数据要素×”，城市全域数字化转型促进数据产业、企业、场景在城市聚集，企业人工智能创新应用、开发高质量数据集；

数据流通交易

- 1、合规流通交易：交易方式多元化互通互认，贴近市场需求开发数据产品和服务合法合规流通，跨境流通国际合作；
- 2、第三方服务：数据经纪、数据咨询服务、合规认证、安全审计、质量评价、资产评估、争议仲裁、风险评估、教育培训；价值评估服务体系、数据资源入表、数据保险、数据信托等金融产品；

基础设施安全

- 1、推动数据基础设施互联互通，打造全国一体化算力体系，创新数据安全产品服务，加强动态安全保障，研究制定《国家数据标准体系建设指南》，培育专业投资机构，重点培养数据采集、治理、分析及合规建设方面专业人才。

数据产业技术标准：高质量数据集与可信数据空间

首席数据官



可信空间
三大能力

- 可信管控:** 构建接入认证体系 (身份权责安全)；数据资源使用合规体系 (合约技术管控能力)；合约履约存证体系 (全程溯源、公平竞争)。
- 资源交互:** 数据标识技术 (资源封装、目录维护、产品互认)；空间统一目录标识、身份认证、接口要求, 实现各类数据空间互联互通。
- 价值共创:** 开发环境 (数据产品和服务开发条件)；共治运营规则 (价值评估模型、收益分配)；发展生态 (开发、经纪、审计、合规)。

财政部--数据资产全过程管理试点

首席数据官



数据资产全过程部分省市试点进展

首席数据官

<p>01 贵州试点</p> <p>贵州省，重点探索数据资产全链条管理体系建设，覆盖9省级部门及重点国企，打造“一目录、一证书、一平台”，并探索“数据+行业”融合创新，推动数据资产全面管理。</p>	<p>02 河北试点</p> <p>河北省，数据资产试点范围覆盖3家省直单位与4个地市，要求试点单位将结合数据资源目录，梳理数据资源，探索编制台账和登记入账，健全数据资产授权运营管理机制。</p>		
<p>03 大连试点</p> <p>大连市，数据资产试点，明确要求编制数据资源目录，登记公共数据资源，清查盘点数据资产，建立台账，登记数据资产，推动国企数据入账入表，健全价值评估体系。</p>	<p>04 浙江试点</p> <p>浙江省，数据资产试点，鼓励金融、交通等多行业探索数据开发利用新模式，鼓励市场力量挖掘数据商业价值，并推动数据知识产权登记、挂牌和交易，以促进数据资产发展。</p>		

数据资产与高质量数据集关系总结

首席数据官

NOTES

数据资产是指能够为企业或组织带来经济价值的、可被管理和利用的数据资源。

高质量数据集是数据资产的核心组成，原始数据通过清洗、标准化等处理，形成高质量数据集，**高质量数据集通过权属确认、估值、入表等流程，成为可计量的数据资产**，最终推动数据要素价值释放。

高质量数据（如完整、准确、时效性强）能提升数据资产的可用性和可信度；**低质量数据**（如缺失、错误或冗余数据）会降低资产价值，甚至导致决策风险。**数据质量需从准确性、完整性、一致性、及时性、唯一性等维度量化评价**，并通过系统化工具实现高效管理。通常采用**数据质量评测系统**，经过“**确认数据-匹配规则-执行评价-生成报告**”等流程环节实现海量数据（如千万级记录）的人工评测难题。系统内置的评价模型和算法可针对不同数据类型（如结构化、非结构化）动态适配标准。例如金融机构通过部署数据质量管理体系，将**客户信用数据的完整性从70%提升至98%**，显著提高了数据资产抵押贷款的审批通过率。

02 技术思路

数据资产战略意义：数据价值释放与转型升级

首席数据官

- 数据资产化是数据要素释放价值的过程，也是企业数字化转型的重要组成部分
- 深入挖掘企业自身的数据资产价值，对提升企业核心竞争力和推动企业可持续发展具有重要意义



创新盈利

01



真实透明

02



增信融资

03



优化结构

04



转型升级

05

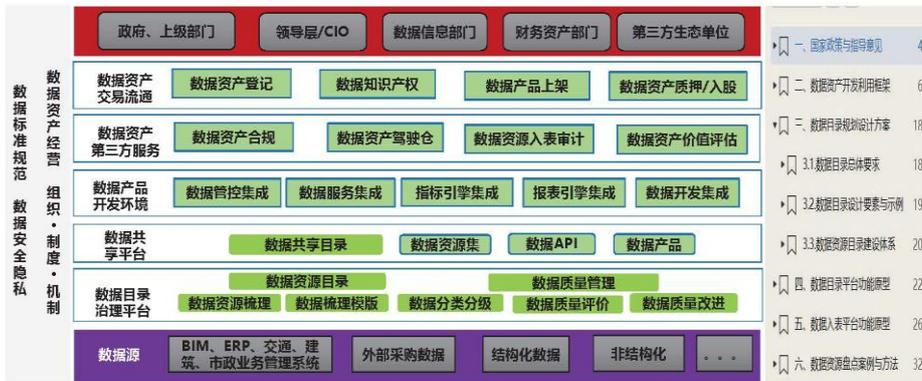


分享红利

06

数据资产技术体系：数据资源治理与开发利用

首席数据官



数据资产入表：数据资产管理能力的成熟

首席数据官

将数据资源确认为资产，不仅需要会计处理，在前期也需要一些基础条件的支撑，从某种角度看，数据资源能够入表代表着企业的数据资产管理成熟度达到了一定程度。



数据资产入表：成本核算、分摊是关键

首席数据官

成本大类	成本小类	具体说明
前期费用	数据规划	数据规划的整体成本，包含数据生存周期整体规划所投入的人员薪资、咨询费用及相关资源成本（人工工资部门预算支出/规划项目费用）
	数据采集	主动获取 ：向数据持有者购买数据的价款、注册费、手续费、服务费等，通过其他渠道获取数据时发生的市场调查、访谈、实验观察等费用，以及在数据采集阶段发生的人工工资、打印费、网络费等相关费用。 被动获取 ：企业生产经营中获得的数据、相关部门开放并经确认的数据、企业相互合作共享的数据等，开发采集程序等相关费用。
	数据汇聚	数据汇聚成本是指合并来自不同数据源的数据的过程中发生的成本，具体包括投入相应的计算资源、大数据科研及技术人员的薪酬、接口费及咨询的成本等。
建设成本	数据存储	数据存储库的构建、优化等费用
	数据开发	信息资源整理、清洗、挖掘、分析、重构和预估等费用 知识提取、转化及检验评估费用 算法、模型和数据等开发费用
	数据应用	开发、封装并提供数据应用和服务等产生的费用
	数据维护	数据质量评价费用，包括识别问题、敏感数据等费用 数据优化费用，包括数据调整、补充、标注、更新、脱敏等费用 数据备份、数据冗余、数据迁移、应急处置等费用 与数据资产相关的软硬件采购或研发以及维护费用
间接成本 (不算数据资产分摊)	软硬件成本	
	基础设施成本	包括机房、场地等建设或租赁以及维护费用
	公共管理成本	包括水电、办公等分摊费用

数据资产入表：系统平台支撑

首席数据官

“数据资产入表管理平台”一级功能模块规划为：数据资产账户、数据资产管理、数据资产自审、数据资产驾驶舱、系统设置等。平台主界面综合统计展示企业所有数据资源、数据资产，数量、金额、成本收益、评估、评价、资产合规等情况。同时主界面列表每项具体数据资产项目，主要应用于数据元通公司数据资产管理经营部门具体管理操作。

The screenshot displays a dashboard for data asset management. It features several key metrics and data points:

- 数据资产 (Data Assets):**
 - 资产总额: 305,731,799.00 元 (18个)
 - 无形资产: 194,519,486.00 元 (17个)
 - 有形: 111,212,313.00 元 (1个)
- 成本与收益 (Cost and Revenue):**
 - 成本总额: 305,731,799.00 元 (16个)
 - 收益总额: 844,825,505.00 元 (10个)
- 数据资源 (Data Resources):**
 - 数据资源个数: 23个
 - 累计使用存储空间: 44,261.20 MB
- 资产评估 (Asset Evaluation):**
 - 资产评估: 78.99 分 (4个)
 - 资产评估标准: 6套 (4个)
 - 资产评估数量: 69个 (4个)
- 资产合规 (Asset Compliance):**
 - 数据资源数量: 1,731条 (22个)
 - 数据资源合规数量: 42条 (547个字符)

The interface also includes a navigation menu on the left, a search bar at the top, and a footer with the company name (数据元通) and the date (2024-10-31 13:05:54).

数据资产目录：关键治理环节

首席数据官



*****数据资源梳理：**梳理企业数据资源管理现状，分析数据可利用价值，进行基本分类分级、来源合规性分析。摸清企业数据家底，**有多少数据、什么数据、存储在哪、数据归属管理责任人。**

*****建立数据目录：**直观展现企业所掌控的数据资源情况，更有效的进行数据资产开发利用，进一步**指导企业构建、完善与业务运营相适宜的数据资产运营体系。**

1. 建立企业**数据治理运营机制**，开展数据资源梳理、数据资产盘点。
2. 手工盘点结合智能扫描企业数据资源，**识别有潜在价值**的企业数据资源。
3. 梳理数据结构及内容，明确**库-表-字段结构体系**，标记存储位置、内容、存储量。
4. 基于**业务进行数据分类**，基于**安全进行数据分级**，为数据开发利用提供规提供参考。
5. 确定数据资产**业务归属和责任人**，为跨部门数据使用、关联分析、安全保护参考。
6. **建立企业数据资产目录**，清晰展现企业所控制的数据资源内容、数据量、数据价值、数据存储位置以及数据归属和责任人，为企业进行**数据资产开发利用、数据价值挖掘以及安全保护**提供依据，支撑企业**数据资产运营体系**构建。

数据资产目录：数据资源梳理模版设计

首席数据官

信息系统目录清单											
编号	项目名称	信息系统名称	服务器地址	数据库类型	数据存储量	数据大类	管理运维部门	业务使用部门	外部使用单位	关联系统	系统功能概述

城市轨道交通数据资产目录清单											
编号	数据集名称	信息系统	数据表	关键字段	容量信息	更新频率	数据来源(类型)	数据状态	数据格式	数据分类	数据分级

入表选型建议：以模型数据查询分析部分数据集合为标的物进行模拟入表，归集数据开发成本入账无形资产，其余部分入账开发支出；
 总体思路：按照数据确权、质量评价、价值评估、入表披露推进数据资产化；按照数据资产盘点、管理、入表、交易、运营开发、产业赋能深入挖掘数据价值，稳健推进数据要素×。

数据目录分类框架思路			
数据主领域	数据分类	数据资源集	信息项(关键字段)
行政区域		行政区	所有行政区
		街道	所有街道
		手机信令网络	150*150/250*250栅格数据
交通设施		道路	所有道路网数据(高速公路、快速路、主干道、次干路、支路、国道、省道、县道)
		轨道交通	所有轨道交通和轨道交通
		公交	所有公交线路、公交站点
		铁路	所有铁路
		客运枢纽	所有机场、火车站、客运站、内部枢纽
兴趣点		所有公交场站、停车场、收费站、加油站、学校、医院、景点	

数据资产目录：系统平台支撑

首席数据官

“数据目录治理平台”主界面综合统计展示接入的共享机构数量、数据资源数量、共享数据产品API接口、资源调用服务以及**数据资源主题分类（例如公路运输、城市客运、轨道交通、桥梁隧道等数据大类）**情况。分栏目列表展示数据资源更新以及共享使用热门数据等情况。一级功能模块规划为：**数据资源主题、数据资源共享目录、数据产品API目录、数据质量管理等。**



数据资产质量：批量数据评测方案

首席数据官

- 1、企业数据资产质量情况直接影响到企业数据资产**治理**、数据资产**价值评估**、数据资产**入表**、**债权融资**、**交易流通**等一系列工作成效。
- 2、企业数据资源往往包含**成百上千个数据表**、**几千个数据字段**、**上亿的数据记录**，人工进行数据质量评价难度大、效率低，根据实际数据情况，不同的数据类型需要采用不同的评价标准。
- 3、数据资产质量评测系统内置科学的评价体系、**成熟的模型和评价算法**，采用**五步法**“**确认数据-匹配规则-执行评价任务-展示结果-生成报告**”进行数据质量评测。

- 1 确认数据
- 2 匹配规则
- 3 执行评价任务
- 4 展示结果
- 5 生成报告



数据资产质量：评测报告作用示例

首席数据官

数据质量评测作用：

- 1) 数据资产价值评估依据
- 2) 数据治理体系建设依据和参考
- 3) 政务数据资源管理考核机制需要。



数据资产估值：价值评估国标（2026年5月1日实施）

首席数据官

评估保障为**前提**，评估依据为**参考**，评估对象要**明确**，数据评价与价值评估为**任务目标**。

数据评价：数据资产**质量要素**或**生成要素**或**应用要素**。

价值评估：收益法、成本法和市场法**三种基本方法**及其衍生方法。

评估保障	评估依据	评估对象	数据评价		价值评估
技术保障	权属依据	属性信息	质量要素 (6)	准确性、一致性、完整性、规范性、时效性、可访问性	收益法
平台保障	取价依据	基本特征	生成要素 (3)	数据生成投入分析、生命周期管理分析、数据安全合规分析	成本法
制度保障	其他依据		应用要素 (5)	使用范围、应用场景、商业模式、供求关系、数据关联	市场法
数据安全保障					

数据资产开发：产品类型

首席数据官

1、数据驱动智能

数据智能产品是指利用数据智能技术，挖掘数据要素价值，赋能传统实体产品提升综合竞争力。例如，电网变压器可能分布在山林、野外、戈壁、荒漠等人烟稀少之地，我们通过**集成数据采集管控芯片实现远程检测变压器运行状态**，加持数字化管控运维能力，延伸了变压器产业链。还有**大型挖掘机产业链延伸**，可以借助物联网技术实时掌控挖掘机运行数据，让挖掘机的运行情况实时在线，及时地提供检测维修服务，而使用挖掘机的客户也可以实时掌握挖掘机所处的位置、业务运作情况，从而掌握商业的全貌，精准进行资源调度和分配。

2、数据分析产品

为企业经营管理决策提供数据支持，如**商业智能工具、管理驾驶舱、数据大屏和能提供决策建议的对话机器人**，解决特定的业务管理需求问题，通过将企业数据加工成支持决策的业务报表，并以语音、文字、图表等形式展示给用户。

3、数据服务产品

数据作为生产要素的主要应用形式之一，典型代表包括大数据交易所挂牌的**各类数据集和算法**。主要形式包括：**API**，最实时、最普遍的提供数据服务方式；**数据订阅服务**，以订阅的方式主动向客户推送数据、信息和知识；**文件数据服务**，比如通过FTP下载统计类数据。

数据资产权属：北京知识产权试点政策

首席数据官

概念总则

1. 与数据产业创新相关**智力成果**所享有权利；**算法、模型**；独创性、复制性、复杂性，法律保护；
2. 保护“**数据持有者**”、“**数据处理者**”权益；引导企业创造、运用、管理和保护数据知识产权能力，推动**释放数据要素价值潜能**；
3. 依法**收集或合同约定获得**，经过**算法处理**、具有**实用价值**及**智力成果**属性的数据集合的权益；可以成为**企业重要的数据资产**；

数据知识产权创造

1. 企业**立足基础和场景需求**，培育数据知识产权**核心竞争能力**，提高数据产品服务**供给能力**；
2. 质量优先、兼顾数量**发展策略**；优化数据创新战略布局，不断增进**数据知识产权价值**；
3. 创新联合体，数据融合特色，数据知识产权纽带，共同打造数据驱动的创新生态和数据价值生态，携手提高数据知识产权创造能力；
4. 通过全市**统一平台功能模块**进行合法登记，鼓励北数所先行区**社会数据资产登记中心**业务；

数据知识产权运用

1. 深挖数据资源价值推动**政企数据融合**应用和场景创新，把数据知识产权作为重要的**无形资产**积极开展**数据资产入表活动及资本化运作**；
2. 鼓励北数所进行**数据知识产权交易**流通；

数据知识产权管理

1. 企业应构建覆盖数据全生命周期的数据管理体系，**增强数据管理能力**，支撑数据知识产权工作；
2. 结合数据运营管理需求，建立健全数据知识产权管理体系；
3. 可以把**数据知识产权合规管理**纳入数据合规和知识产权合规绩效评价体系；

数据知识产权保护

1. 定期分析**经营产品、软硬件设施及业务流程**涉及他人数据知识产权情况；
2. 不合规行为：窃取等非法方式**获取数据**；**盗采滥用**不当数据；**算法规则**等数据**技术手段垄断**、不正当竞争；非法数据**交易**或知识产权交易；

数据资产权属：登记取证查看

企业级数据资产管理平台与行业数据资产管理平台互联互通，快速帮助企业获取数据资产登记证书，并在本地系统进行查看、核验。

提供将数据资产快速登记到全国数据资产登记平台节点平台的入口，并能够一键获得当前数据资产的登记证书，在数据资产系统进行查看。



上市公司前三季度数据资产入表融资概况

首席数据官

发展维度	关键数据/指标	数据可靠性分析	战略意义
数据入表规模	A股上市公司披露规模超18亿元。	该数据基于上市公司公开财报披露，可靠性高，反映了市场活跃度。	内部价值确认：主要由成本法驱动，多数为自研数据资产资本化。
入表结构	三大基础运营商入表13.66亿元，增速高达426%；91%确认为研发支出资本化形成的无形资产。	结构精确，指向了现阶段入表的主流路径：大型央企、成本法、内部使用。	资产质量与挑战：尚未充分体现数据资产的交易价值和外部变现能力。
数据贴标ABS	全国首单贴标ABS（华鑫-鑫欣）储架规模5亿元，已获深交所无异议函。	金融监管机构文件披露，数据精确。	里程碑突破：验证了数据资产作为底层基础资产直接证券化的可行性。
数据赋能ABS	首单赋能ABS（天风中投保）发行规模5.14亿元。	交易所公开信息，数据精确。	价值规模化路径：通过数据赋能提升传统资产信用（风控、定价），是更具普惠性和规模化潜力的创新模式。

省级高速公路企业数据资产化试点

首席数据官



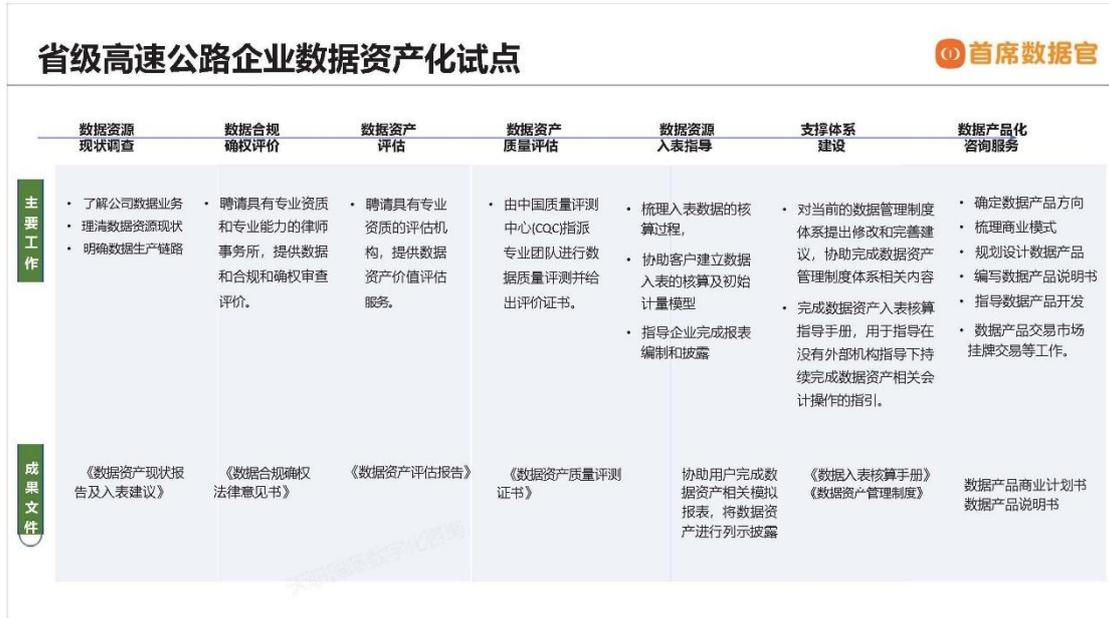
公司经辽宁省政府授权作为全省高速公路唯一的经营主体，享有辽宁省内高速公路特许经营权。目前，公司管辖辽宁全省高速公路4409公里/28条，沿线分布主线桥梁1129公里/14716座（特大桥84座、大桥2242座、中桥3302座、小桥9088座），隧道202公里/224座（特长隧道4座、长隧道68座、中隧道102座、短隧道50座），高速公路路面使用性能指数常年保持优良级别。设置收费站299个，ETC车道覆盖率达到100%，日均服务出入口车辆160万台次。形成了以省会沈阳为中心，连接14个省辖市及100%陆地县的路网格局。

数据资产特点

- 数据质量高，评测S级
- 数据完整，全省唯一
- 人员数据素养高，数据应用场景多

项目产出

- 入表金额140余万，评估1000万
- 数据合规确权意见书
- 数据质量评测报告
- 数据产品开发规划设计报告
- 数据资产管理制度体系
- 数据资产核算手册



省级农业畜牧养殖数据资产化试点

首席数据官



60万头奶牛数据
4000万 各类数据
494.68万 生产数据
15.34万 头体型数据
109.7万 条产犊记录
352.3万 条繁殖记录



饲料中粗蛋白、粗脂肪、粗纤维、淀粉、矿物质、维生素和添加剂等营养成分的配比，涉及**饲料检测数据**
28000 条

数据价值

提质增效

- 体细胞和细菌数逐年降低，牛奶质量达到最好水平，全年体细胞数17.67，细菌数1.28，较2018年，体细胞水平降低了25.5%，细菌数降低了66.3%
- 近5年平均产奶增幅4%，平均产奶量增加425kg
- 近5年平均产奶周期5%，平均头日产增加1.6kg/天
- 成母牛实繁率大幅提升，稳定在较高水平，较2018年，实繁率提高约8%，每年多产犊牛3800头，5年共计多产犊牛19000头，每年可增加生物资产3040万

降本降耗

- 在奶牛日粮营养水平不变的情况下，全群可节省饲料费用15.11万元/日，每年节约饲料达43339吨，饲料费用每年可节约5516.8万元
- 与传统育种技术相比，奶牛基因组选择育种技术能使种公牛培育周期由6年缩短至约2年
- 质检数据实时上链，青贮以质论价——2023年节约采购金额660万元

入表资产

¥330万

无形资产（其中数据资源）

数据资产评估

¥2000万

采用重置成本法

数据资产质押贷款

¥1020万

北京农商银行

北京市属国企“首单”数字资产入表

首农食品集团数字化转型迈出关键一步

估价2000万元，北京国企养殖业大数据变资产

1020万元，首农畜牧数据资产成功获得质押贷款！

关注首农，美好生活守护着你

省会城市公共数据资产管理试点

首席数据官

- 目的范围：**助推福州市产合数据资源高效流通使用；**部分行政事业单位和公用企业；**
- 运营任务：**有条件的运营主体市数据局审查合规性；梳理**数据资产目录**，细化本单位数据资产**编码规则**，形成**数据资产信息卡**，完成数据资产**盘点工作**；选定律所和第三方评估机构出具**合规报告**、**价值评估报告**，大数据服务中心登记**数据资产卡**；各运营主体结合实际选取**数据开发主体**合规开展公共数据开发利用。
- 数据资产盘点：**定期对拥有数据系统化清查、梳理、评估，形成**数据规模、类型、以、价值**等核心内容数据资产目录；
- 数据资产登记：**登记主体将经过第三方评估的合法数据资产进行统一记录，并获得数据资产**登记凭证**的过程；
- 数据资产运营：**在不改变数据生产主体持有**数据和履行数据安全保护**义务前提下，开展数据资产**归集、加工处理、分析挖掘、开发形成数据产品和服务、价值评估和交易流通**等的活动过程。
- 数据开发主体：**按照数据运营主体管理要求，针对应用场景，开展数据资产开发利用的主体。数据开发主体对**协议范围内的数据资产依法依规进行加工使用。**

福州市公共数据资产解决方案主要需求		
方案大纲	方案编制要求	业务概述
项目概述	阐述本项目的概况，包括但不限于以下内容：项目背景、政策依据、建设目标等。	
现状分析	阐述现状分析，包括但不限于以下内容：业务现状、应用系统现状、数据应用现状、建设必要性分析等。	
需求分析	阐述本项目的需求分析，包括但不限于以下内容：业务需求、用户需求、功能需求、数据资源需求、对接需求、网络安全需求等。	
总体设计	阐述本项目的总体设计，包括但不限于以下内容：总体思路、总体架构、数据架构、数据流、网络结构、主要设计原则等。	
建设内容设计	数据资产盘点	明确数据资产目录，对数据资产进行盘点，以全面了解组织内部所有数据资产的分布、数量、质量、使用状况、系统分布等，可自动生成数据资产报告、数据资产评估报告等，确保数据的真实性、有效性。
	数据资产登记	明确“数据三权”登记工作，建设数据资产登记系统，为组织和企业提供数据资产登记受理、审核、公示、登记、上链功能，证书下载等功能，并支持将登记信息公示和登记证书的开放查询，以及登记证书的变更、年检和注销等服务，有序推进数据资产确权工作，推动数据资产的合规使用与价值实现。其中，登记信息在审核完成后登记区块链中，并生成登记证书。
	数据资产卡片管理	建设数据资产卡片管理方案，并按照财政部统一设计标准要求建立数据资产卡片，内容包括但不限于：资产体系、数据资产卡片入库、变更、使用、处置等全生命周期管理，并实现卡片间逐步实现数据资源管理一体化系统资产模块，同步形成会计核算系统中的项目来源数据项目信息。
	数据资产目录汇总	建设数据资产目录公示方案，内容包括但不限于：标准数据资产目录公示、关联数据产品展示和推送、数据产品信息上下管理、需求信息交流等功能。
数据资产收益分配和清分	明确数据资产收益分配和清分方案，内容包括但不限于：查看资产流水、关联收益账单、生成清分账单、资金到账、台账记录，以及与福建非税收入收缴和财政电子票据系统对接，实现数据资产非税收入缴交、开具财政电子票据和资金使用用途清分。	
安全设计	阐述本项目系统安全设计，包括但不限于网络安全、数据安全、隐私应用等。	
信创要求	阐述系统适配信创要求，包括但不限于适配人大金仓达梦等信创数据库。	
系统对接设计	阐述本项目系统对接设计，包括但不限于福建预算管理一体化系统资产模块、用于数据资产登记的区块链平台、福建非税收入收缴和财政电子票据系统等。	
系统部署实施	阐述系统部署实施方案，并实现7天内部署上线。	

CDO数据资产管理三板斧

首席数据官

第一阶段：认知规划，百度搜索“数据资产技术咨询”AI智能体，自主学习、生成解决方案大纲及思路，完善形成**适合企业自身特点的“数据资产全过程试点方案”**。

第二阶段：试点验证，结合市场上系列产品、服务，开展“咨询规划、数据目录、数据治理、数据开发、数据入表、流通交易”等，构建、维护企业数据资产运营框架，按需开展数据治理、产品开发、金融创新等活动。

第三阶段：常态推广，完善、建立企业数据资产全过程管理**框架体系**，机制，生态、工具平台等！



数据资产技术咨询智能体



政务数据要素化实践

IITC工信人才岗位能力评价培训
数据产业高层次人才培训课程

首席数据官
GOVCDO

政务数据要素实践

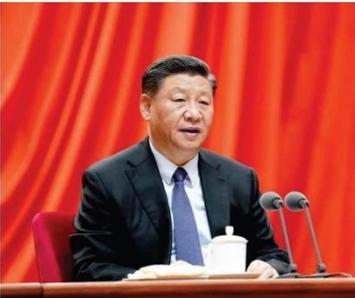
讲师：刘方哲 龙华数据

2025年10月

党和国家高度重视数据要素发展

首席数据官
GOVCDO

将数据作为一种生产要素，是习近平新时代中国特色社会主义思想的一个重大理论创新。以此为基础构建数据基础制度，探索构建与新生产要素、新生产关系、新经济基础等相适应的新上层建筑。



- 2015年10月**
党的十八届五中全会首次提出“国家大数据战略”
- 2017年12月**
中央政治局第二次集体学习时提出“要构建以数据为关键要素的数字经济”
- 2019年10月**
党的十九届四中全会首次提出“将数据作为生产要素之一、参与分配”
- 2020年4月**
中共中央、国务院印发《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、劳动力、资本、技术等相并列，提出“要加快培育数据要素市场”
- 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》（即“数据二十条”），构建起数据基础制度体系的“四梁八柱”
- 2024年7月**
党的二十届三中全会审议通过《中共中央关于进一步全面深化改革、推进中国式现代化的决定》7次提到数据，对数据市场、数据要素、数据基础设施、数据共享、数据产权、数据跨境流动等工作作出部署
- 2024年9月**
中央办公厅国务院办公厅印发《关于加快公共数据资源开发利用的意见》，中央层面面对公共数据资源开发利用首次系统部署

数据要素市场化配置



数据要素要与市场挂钩，数据要素要走向市场才能发挥要素价值，围绕数据要素市场流通要建立相应的市场机制，推进市场化配置改革。

数据要素市场化配置改革

数据“供得出”

落实产权分置制度，理清数据持有权、加工权、经营权，让公共数据“供得出”

数据“流得动”

建设全国一体化安全可信的数据流通基础设施，让公共数据“流得动”

数据“用得好”

针对行业发展痛点堵点，实施“数据要素x”行动计划，让公共数据“用得好”产生倍增效应

数据“能运营”

以公共数据为引领，鼓励探索公共数据、企业数据授权使用新模式；

数据“可交易”

优化全国交易所布局，发展服务型、应用型、技术型数商，建立定价、交易、收益分配制度

数据“保安全”

要防范隐私泄露、数据滥用等潜在风险，实现更高效率、更加安全的发展

深刻认识做好公共数据资源开发利用的重大意义



公共数据是国家重要的**基础性战略资源**

是数据要素市场的**重要组成部分**

加快公共数据开发利用是统筹推进数字中国、数字经济、数字社会建设的**必然要求**

数字政府建设

- 是推进国家治理体系和治理能力现代化的**重要支撑**；
- 有利于加快政府职能转变，推进国家治理体系和治理能力现代化。

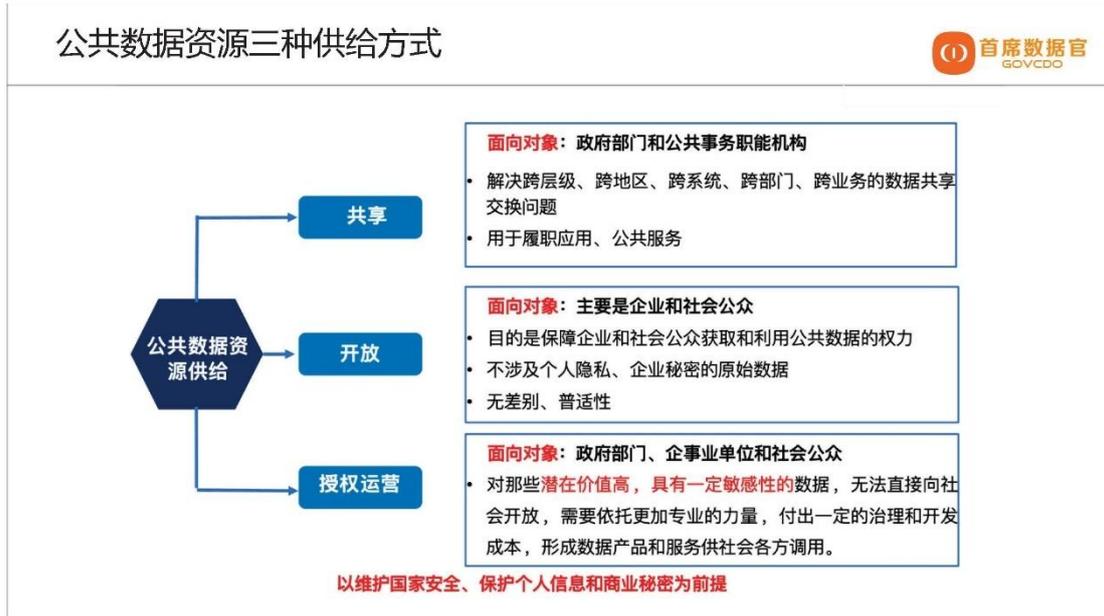


数据要素市场建设

- 是深化数据要素市场化配置改革的**关键举措**；
- 有利于激发市场创新活力，培育数据产业，形成新的经济增长点。

产业数字化转型

- 有利于加快传统产业转型升级，优化经济结构和运行机制。



我们用“数据要素”办三件实事

- 免申即享**
把企业“找政策—兑政策”从6步压成3步，材料更少、时限更短。
- 龙数贷**
把政务可信记录变成“增信因子”，帮中小微更快拿到贷款。
- 数据入表**
把沉睡数据写进报表与资产负债表，撬动授信与质押融资。

龙华区数据要素市场化探索

首席数据官 GOVCDO

公共数据授权运营

推出“龙数贷”“龙数贸”等公共数据产品，探索“数据要素×金融”和“数据要素×外贸”的应用模式。

数据资产入表

对接龙华排沁有限公司开展数据资产入表工作，成为全市首个“入表+授信”完整闭环的国企案例，为国有企业融资探索了新路径。

抢滩低空经济

抢滩“低空经济”蓝海，如期高质量完成20个低空起降点建设任务，助力龙华区成功跃居全市低空起降设施建成数量首位。

土地 在推进土地要素市场化配置方面，提出建立全域多规合一的用途管制市场。

劳动力 在引导劳动力要素合理流动和配置上，深化户籍制度改革。

资本 在推进资本要素市场化配置上，完善信贷市场基础制度。

技术 在加快发展高技术要素市场上，健全知识产权保护成果产权制度。

数据 在加快培育数据要素市场上，健全数据产权、开放共享。

- **建市场：**培育规范化、高效化的数据要素市场。
- **促潜力：**挖掘数据价值，维护数据安全。
- **强运营：**构建“组织、资金、技术与政策”相结合的一体化体系。
- **稳保障：**创新“整合与轻资产”资金模式。
- **育生态：**共同构建“政产学研用”协同发展的新格局。

01

政务数据 共享开放

龙华区以公共数据开发利用赋能政务服务创新



在数字化时代，数据已成为重要的生产要素，公共数据的开发利用对于提升政府治理能力、优化营商环境、服务民生福祉具有关键作用。龙华区积极响应国家数字化发展战略，深度挖掘公共数据价值，通过创新政务服务模式，推动公共数据在惠企利民领域的广泛应用，走出了一条具有特色的数字化政务服务之路。

秉持先进理念 政策兑现流程

以“免申即享”重构兑现链路，将传统6环节压缩为“智能比对—意愿确认—自动兑付”3环节，以数据流转替代人工传递，实现“政策找人”、材料减负与时限压缩。

依托“龙脑”引擎 打破数据壁垒

依托“龙脑”整合市区多部门数据，构建标准化接口与共享协议的跨部门融合生态，为“免申即享”和“政策AI计算器”提供精准、按需的“滴灌式”数据支撑。

数据开放共享 实现“免申即享”

以“数据整合—算法驱动—智能决策”为主线，按“一事项一算法”与政策特征标签自动识别靶向对象，联审比对并开通“信用即资格”零干预通道，推动“人工筛选”向“数据兑现”转变。

践行“换位思考”， 打造“企业条件导办”

构建“企业画像×政策图谱”匹配引擎，将300+政策拆解为188个条件标签与8大图谱，企业勾选条件即可获得“条件勾选—政策推荐—一键申报”的全流程指引。

引入AI算法 实现补贴政策精准推送

汇聚431项政策要件并学习近5.6万条历史申报数据，形成覆盖近60万家企业的画像与三级匹配模型，企业仅输入名称即可在8秒内生成约30项可申报清单并联动税收优惠。

区大数据管理建设情况



数据共享通道搭建

龙华区大数据管理平台打通市区两级跨部门的数据共享通道，建立全区政务信息资源共享、复用的统一管理机制。这一举措有助于打破数据壁垒，促进各部门间的数据流动，为后续的数据收集和应用程序奠定了坚实基础。

数据需求收集依据

平台已根据国务院办公厅印发的《六批国务院部门数据共享责任清单》《四批国务院部门垂直管理业务系统与地方数据平台对接责任清单》收集区内各单位数据共享需求和供给信息。严格依据相关清单，确保数据收集的规范性和全面性。

数据汇聚成果

截至2025年5月，已汇聚约90个市区两级部门共计3500余张表约90亿条原始政务数据。如此庞大的数据量体现了平台强大的汇聚能力，为全区多种场景落地应用提供了丰富的数据资源。

数据支撑作用

这些汇聚的数据为全区多种场景落地应用提供数据支撑，为授权运营提供坚实的数据资源基础。无论是政务服务优化，还是其他相关业务开展，都能从这些数据中获得有力支持，推动区域发展。



公共数据授权运营探索



现状

- 国家导向：将“数据要素×金融服务”列为重点场景。
- 数据基础：龙华区政务数据丰富，但分散沉睡、利用不足。
- 现实痛点：中小微企业征信覆盖有限，融资难、融资贵、融资慢问题突出。

探索

与龙华区政务和数据局共同推出“龙数贷”“龙数贸”等公共数据产品，探索“数据要素×金融”“数据要素×外贸”的应用模式。

整合九大产业部门区级补贴数据
(累计10万条、金额超百亿元)

覆盖全球 200 个国家和地区的
5.6 亿家企业商业数据。

成效

龙数贷

- 累计放款 234 笔，额度 4.34 亿元
- 对接授信企业 24,094 家，首贷户 970 家，累计授信金额 804.11 亿元。
- “龙数贷”数据产品在2024年中国(深圳)数据要素产业创新大会上获得了卓越领航案例。

龙数贸

- “龙数贸”已输出超50份市场分析报告、提供15000余条海外销售线索。
- 助力企业潜在客户数量提升3倍、初始转化率由25%提升至50%。

“龙数贷”介绍



“龙数贷” 定位与价值



以政务数据、企业数据如政府奖补资金、纳税、社保等为中小微企业精准增信，辅以征信、工商等数据，为企业提供信用贷款，单家企业额度最高 1000 万元，主要服务于科技型中小企业，满足其资金需求。

产品定位介绍

龙数贷作为区级金融公共数据产品，以“政务数据×金融风控”为底座，面向科创型、中小微与园区企业的融资场景，提供数据增信、在线授信与风险控制服务。产品与银行机构深度对接，支持在深圳数据交易所上架与合规流通，形成“数据要素驱动金融”的业务形态。

价值点分析

将分散的纳税、社保、奖补、人才等政务可信记录转化为可计算的信用因子，显著降低尽调成本与授信时长；提升授信命中率与额度，缓解“首贷难、首贷贵、续贷烦”。同时与园区招商、产业链金融联动，形成“园区画像—银行风控—融资服务”的闭环，加速企业成长。

“龙数贷” 数据产品，为银行风控降本增效提供深圳样本



01 交易情况

在深圳数据交易所平台合规上市的“龙数贷”数据产品连续完成两笔交易。

交通银行股份有限公司深圳分行、中国农业银行股份有限公司深圳龙华支行，先后与深圳市龙华数据有限公司签约，以数据流通破解银企信息不对称难题，为普惠金融装上“数字引擎”。

02 产品开发

“龙数贷”由深圳市龙华数据有限公司开发，在龙华区发改局、财政局和政数局的指导下，创造性复用政府产业部门惠企政策补贴数据，通过深圳数据交易所系列场景评估，与金融机构的风控贷款模型相融合。

03 产品作用

“龙数贷”为金融机构创新相关金融产品提供数据支撑，助力提升龙华区科技中小微企业的贷款额度，提升实体经济金融服务水平，实现政务数据与金融信贷场景深度对接，是一款公共数据产品。

04 行业影响

随着普惠金融进入“数据要素×金融服务”深水区，以可信数据流打通银企对接“最后一公里”已成为行业共识。此次“龙数贷”连续成交，为银行风控降本增效提供了可复制、可推广的深圳样本，为实体经济高质量发展注入金融活水。

“龙数贷” 创新模式

首席数据官
GOVCDO

创新实践背景

在科技型中小微企业发展中，“融资难”是一大难题。轻资产、缺抵押、财务数据不完整，使企业难以跨过银行传统风控门槛。而深圳龙华区开展的“数据要素×金融服务”创新实践，正改变这一局面。

“龙数贷”介绍

由深圳市龙华数据有限公司开发的公共数据产品“龙数贷”连续落地成交，是深圳首个通过政务数据赋能金融机构风控、精准滴灌科技中小微企业的典型案列。其将惠企政策补贴数据激活，与金融机构风控模型融合。

评估体系变化

过去银行评估企业贷款资质依赖财务报表、抵押物等传统信息。现在“龙数贷”把企业在龙华区享受的政策补贴数据纳入评估体系，这些数据反映企业政策契合度、发展潜力与经营稳定性，转化为风控参数提升审批效率。

模式意义

该创新模式意义重大。对金融机构，提供可信数据流支撑，降低风控成本与尽调难度；对企业，提升贷款额度，获更精准高效金融服务；对龙华区，是政务与金融场景对接突破，提供可复制深圳样本。



“龙数贷”：获奖情况

首席数据官
GOVCDO

积极推进“数据要素×”应用场景试点，完成了“龙数贷”数据产品在深圳数交所上架交易，在2024中国（深圳）数据要素产业创新大会，荣获“数据要素卓越领航案例”。

以“基于隐私计算的公共数据产品‘龙数贷’”，获评2024年大数据“星河（Galaxy）”的数据要素流通专项潜力案例；



龙华数据公司与行业主管部门联合举办“龙数贷”产品推介会



- ✓ **“龙数贷”推介会举办**
在龙华区发改局、区政数局指导下，龙华数据公司成功举办“龙数贷”数据产品推介会。会议以“数字赋能金融 助力企业发展”为主题，辖区25家银行龙华支行负责同志参会，推动金融机构深度应用“龙数贷”产品。
- ✓ **“龙数贷”运营成果**
“龙数贷”是深圳首款金融公共数据产品，面向银行类金融机构，复用政府产业部门惠企政策补贴数据，与金融机构风控贷款模型融合，为中小微企业精准增信提供信用贷款。自2024年9月5日上线，已累计发放贷款5.18亿元。
- ✓ **“龙数贷”荣誉与备案**
2024年11月，“龙数贷”创新案例获深圳数据交易所2024中国（深圳）数据要素产业创新大会卓越领航案例奖。2025年4月，它成功通过国家平台登记备案，成为深圳首个在国家层面完成备案的公共数据产品。
- ✓ **未来发展规划**
龙华数据公司董事长表示要推动数据要素赋能实体经济，丰富“龙数贷”产品打造2.0升级版，试推“人才贷”为人才型企业提供无抵押贷款。构建区域可信数据空间，联合打造跨部门数据流通基础设施，促进实体经济发展。

“龙数贸”介绍



“龙数贸” 定位与价值





“龙数贸” 数据产品覆盖全球200多个国家和地区，
超过5.6亿家全球企业商业数据。

产品定位介绍


龙数贸作为区级外贸数据产品，具有广泛的覆盖范围，涵盖 200+ 国家/地区、5.6 亿+ 全球企业商业数据，其定位明确，主要面向找订单、配资源和外贸拓展等业务场景，与政务低空平台同场发布，具有较大的市场潜力。

价值点分析


该产品填补了跨境数据产品场内交易的空白，为跨境贸易提供了有力的数据支持。同时，它能够对接产业链与园区招商，促进产业的协同发展，为区域经济增长做出贡献。

“龙数贸” 上线：助力企业匹配全球市场订单数据



01

“五外联动” 组合拳

- 近年来，龙华区着力打好外贸、外资、外包、外经、外智“五外联动”的组合拳。
- 龙华区在经济发展上的全面布局，通过多方面协同，提升区域经济的综合竞争力，积极适应全球经济发展的新趋势。

02

应对产业链重构

- “龙数贸”是龙华区应对全球产业链重构的主动作为，也是探索数据要素市场化配置的改革实践。
- 龙华区以积极姿态应对国际经济格局变化，通过创新举措，推动数据要素合理配置，为区域经济发展注入新动力。

03

创新技术应用

“龙数贸”创新采用隐私计算技术，通过“可信执行环境 (TEE)” “联邦学习 (FL)” “多方安全计算 (MPC)” 三大主流技术方案，将海关进出口统计数与海外合规商业数据融合计算，生成核心产品，为企业海外市场开拓方案。

04

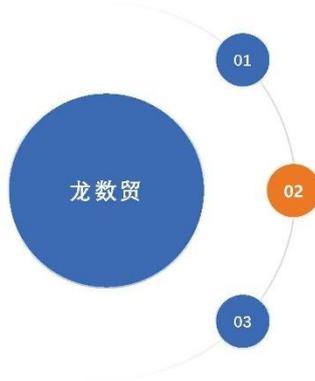
数据安全防护

结合关键行为数字验签和区块链安全日志审计技术，通过平台安全、传输安全、存储安全、计算安全、审计安全五位一体化的模式对数据进行全生命周期的防护，确保数据“可用不可见、可控可计量”，保障企业数据在使用过程中的安全性。

“龙数贸”功能优势



“龙数贸”致力于解决这一核心痛点。作为一款专为出海企业量身打造的智能销售线索核验工具，它依托强大的全球数据整合能力与AI驱动的智能分析引擎能够：



深度核验线索真实性

“龙数贸”作为专为出海企业打造的智能销售线索核验工具，依托强大全球数据整合能力与AI智能分析引擎，可快速验证潜在客户企业信息合法性、经营状态及信用风险等，大幅降低企业“踩坑”概率，为出海企业的业务开展提供可靠保障。

洞悉线索潜在价值

该工具结合行业数据、市场趋势、竞争格局等多维度信息，对线索进行深度分析，能精准把握线索的业务匹配度和商业潜力，让出海企业清晰了解每个线索的价值，从而更好地规划业务方向，挖掘更多的商业机会。

提升决策效率与精准度

“龙数贸”为销售和决策层提供数据支撑，改变以往盲目跟进的状况。它能帮助企业将资源集中在最有潜力的目标上，避免资源的浪费，使得决策更加科学、精准，提高出海企业的运营效率和市场竞争力。

“龙数贸·出海通GoGlobal”项目



项目成功获得2025年“数据要素×”大赛广东分赛——数据要素×商贸流通“特色奖”

项目亮点：

多源数据智能线索服务“龙数贸·出海通GoGlobal”聚焦企业全球化发展中的精准出海需求，通过整合多源数据资源和人工智能技术，为企业提供：

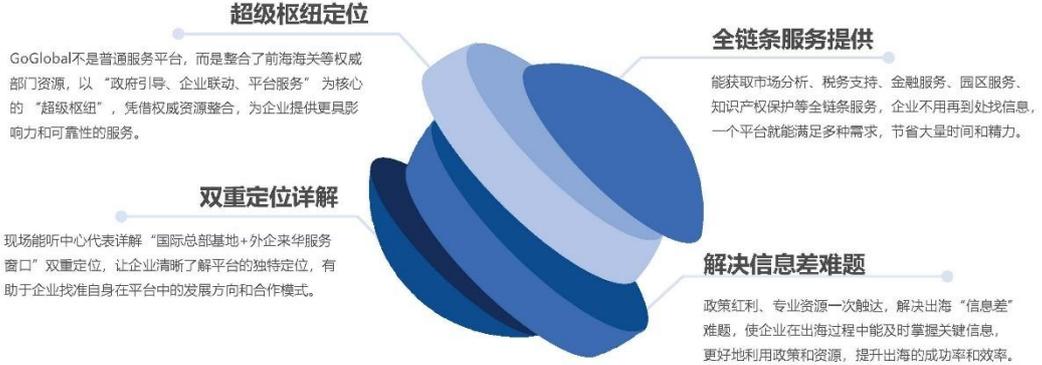
- 精准市场洞察：基于全球贸易、政策、供应链等多维数据构建分析模型；
- 智能商机挖掘：通过动态数据追踪与匹配，生成高价值出海线索；
- 一站式解决方案：覆盖市场定位、渠道分析、风险预警等关键环节，帮助企业降低出海成本，提升决策效率。

强强联合 共创价值

本项目由深圳九鑫软件有限公司与深圳市龙华数据有限公司共同打造，融合双方在数据技术、跨境服务与地方产业资源方面的优势，致力于为企业提供可信、可靠、可控的数据智能服务，助力中国品牌走向世界



GoGlobal平台优势



超级枢纽定位

GoGlobal不是普通服务平台，而是整合了前海海关等权威部门资源，以“政府引导、企业联动、平台服务”为核心的“超级枢纽”，凭借权威资源整合，为企业提供更具有影响力和可靠性的服务。

全链条服务提供

能获得市场分析、税务支持、金融服务、园区服务、知识产权保护等全链条服务，企业不用再到处找信息，一个平台就能满足多种需求，节省大量时间和精力。

双重定位详解

现场能听中心代表详解“国际总部基地+外企来华服务窗口”双重定位，让企业清晰了解平台的独特定位，有助于企业找准自身在平台中的发展方向和合作模式。

解决信息差难题

政策红利、专业资源一次触达，解决出海“信息差”难题，使企业在出海过程中能及时掌握关键信息，更好地利用政策和资源，提升出海的成功率和效率。



数据资产入表探索



现状

- **政策指引：**国家出台系列制度，推动数据确权与入表。
- **数据基础：**龙华排水公司具备丰富业务数据可入表。
- **价值需求：**盘活数据资产，提升融资授信与市值价值。

探索

探索开展数据资产入表试点，建立覆盖登记、估值、入管的全流程管理体系。优先在非水、交通等基础民生领域推进数据资产价值评估工作，形成可复制推广的经验做法。



成效

数据资源列报及披露
龙华排水公司在资产负债表中，项目开发阶段将“河道水位监测预警分析数据集”列入“开发支出”开发完成后转入“无形资产”，反映其期末账面价值。

数据资产评估及入表转化
龙华排水公司资产入表项目已进入交付阶段，首批高价值数据资产将完成评估并纳入财务报表，实现了数据价值从资源向资产的转化。

为“数字龙华”战略深入推进注入新动能



获得融资额度

龙华排水公司凭借数据资产成功获得光大银行深圳分行 1.1 亿元融资授信额度。这笔资金将作为公司的后续发展提供有力的资金支持，助力其在相关业务领域进一步拓展。

打通转变路径

该项目打通了数据要素资源化向资产化、资本化的转变路径。意味着数据不再仅仅是资源，而是可以转化为实实在在的资产和资本，为数据要素的价值实现提供了新途径。

开辟发展空间

该项目为推动数据要素市场化配置改革、探索数字经济发展新空间。有助于促进数据要素的合理流动和高效配置，激发数字经济的创新活力，推动相关产业的发展。



数据资产入表融资

龙华排水公司以“河道水位监测预警分析数据集产品”作为增信条件，完成全市首例国有企业“数据资产入表 + 融资授信”。这一举措是数据要素在金融领域应用的创新实践，展示了数据资产的经济价值。

龙华排水公司合规操作



入表前：筑牢数据基础

入表前要筑牢数据基础。首先建立数据质量评价体系，明确数据准确率需≥98%、更新频率≥每日1次，安排专人负责数据清洗，像剔除重复管网巡检记录、补全缺失水位监测值。其次梳理授权链条，针对市政设施信息获住建局授权文件，针对员工采集数据签归属协议，避免权属纠纷。

入表中：规范流程确权

入表中要规范流程确权。在会计处理上，设立“数据资产”专用会计科目，单独核算数据采集、清洗、存储的成本。同时在深圳数据交易所完成确权，提交数据样本、授权文件、质量报告等材料，交易所审核后通过后出具《数据资产登记证书》，这是后续融资、信托的关键。

入表后：风险隔离保护

入表后要风险隔离保护。龙华排水公司联合华润深国投设立数据知识产权信托，将6660公里检测数据的收益权转让给信托计划。这样既实现了资产增值，又隔离了风险，若未来数据合规政策调整或出现数据泄露纠纷，仅影响信托资产，不波及公司核心业务。



龙华排水落地全市首单国企数据资产质押融资



首单数据资产质押融资

在数字经济加速发展背景下，龙华排水公司突破传统融资模式，以“河道水位监测预警分析数据集”为质押物，成功落地深圳市首单国企数据资产质押融资业务，获融资额度6000万元，标志着深圳市国企数据资产资本化取得实质性突破。

数据资产领域三大突破

作为龙华区属国企，龙华排水公司长期积累大量管网运行与环境监测数据。在龙华区政府指导下，于2024年11月起率先实现三大突破，即完成龙华区首笔国企数据资产入表、落地“数据入表 + 融资授信”组合业务、获发供水行业数据资产登记证书，为数据金融化奠定基础。

构建全流程服务体系

龙华排水公司与上海银行深圳分行开展战略合作，创新构建数据资产“确权 - 评估 - 质押”全流程服务体系。依托专业评估模型量化数据价值，以数据资产质押替代传统保证金模式，降低融资成本、提升审批效率，激活企业沉淀数据资产价值。

未来发展方向

未来，龙华排水公司将聚焦完善数据全生命周期治理体系、深化数据与主营业务融合、探索“资源 - 资产 - 资本”转化机制三大方面。通过释放数据潜能实现增长拓新、运营提效、锻造竞争硬实力，提升公司核心竞争力。

03 抢滩 低空经济

龙华区低空经济现状



企业数量增长

如今龙华成为低空经济创新热土，截至7月，拥有低空经济相关企业445家，较去年新增22家，占全市23%。企业数量的增加体现了龙华在该领域的吸引力和发展活力，为低空经济的进一步壮大奠定了坚实的基础。



创新平台汇聚

目前龙华全区汇聚23个与低空经济紧密相关的创新平台，包含国家级技术创新中心1个，省级技术研究中心8个等。这些创新平台构筑起坚实的“空中基础设施”，为低空经济的技术研发、创新突破提供了有力支撑，推动产业不断升级。



产值增速可观

2025年上半年龙华区低空经济产值（营收）达38亿元，同比增速14%。这一良好的增长态势表明龙华区低空经济发展势头强劲，不仅在规模上有所扩大，在经济效益上也取得了显著提升，发展前景十分广阔。



筹建国家级中心

大疆与北京航空航天大学强强联合，正积极筹建文旅部旗下的国家级无人机集群表演技术创新中心。该中心聚焦文旅行业痛点与前沿技术布局，有望解决行业难题，提升无人机集群表演技术水平，为低空经济在文旅领域的应用带来新突破。

解锁“天空之城” 龙华找到“关键钥匙”



《深圳市低空基础设施高质量发展建设方案（2024-2026年）》发布，表示将建成千条航线与千个起降点，进一步完善低空交通网络布局，多区协同共建“天空之城”。其中，龙华区与南山、龙岗、宝安、光明等区被确立为低空经济发展重点区域，在基础设施布设、应用场景拓展、促进科技创新、推动产业集聚等方面实现突破。



- **数字底座：**依托全信创数字孪生城区+全域高精度三维，融合300余万人口、48万家企业与物联感知数据，持续动态更新。
- **核心能力：**形成政务低空飞行服务平台与“一张图”，禁区自动识别、路径校核与任务前评估，已在重大工程监测、平安校园落地。
- **管理机制：**制度化流程与数据安全并重，飞前审批—飞中监控—飞后留痕闭环，推动政务低空从响应式到主动智能管理。

以“低空空域一张图”为底座，政务飞行从被动响应跃迁为主动智能、零事故运行。

聚集创新平台龙华激活低空经济新动能



- **规模动能：**截至7月，相关企业445家（较去年+22，占全市23%）；2025年上半年产值38亿元，同比+14%。
- **平台矩阵：**23个创新平台（国家级1、省级8、市级9、其他5）；筹建国家级无人机集群表演技术创新中心（大漠大×北航）；与南航共建低空智联产业科技创新中心，引入4个国家重点实验室分支。
- **技术标杆：**大漠大“专精特新”小巨人+国家高新，100+核心IP；上万架三维编队、连续6年登春晚、屡破吉尼斯，贯通“研—转—创—用”。



“数据换空间” 驱动低空经济稳健发展



自然资源部、国家数据局联合发布“2025年实景三维数据赋能高质量发展创新应用典型案例”，由深圳市规划和自然资源局数据管理中心、龙华区政务服务和数据管理局、深圳市规划和自然资源局龙华管理局三方联合申报的《实景三维赋能数字孪生低空管理和政企飞行服务》成功入选。

- 市级底座×龙华试点：深圳完成全市域3.5cm级Mesh，已赋能50+领域/100+场景；2025年以龙华为试点，市区协同打造低空经济“双精品”示范。
- 政企成效双线开花：政务侧“一键启动、全网覆盖、空地联动”，4000+监测点联动编队，交通/应急/安防15分钟自动监测；企业侧“云勘察”+高精导航，配送时长≈30→≈12分钟，效率≈+150%。
- 数据换空间落地：龙华将实景三维与人口/法人/物联等（28万+设备）深度融合，构筑“地上-地下/室内-室外/动-静”孪生底座，12大模块95细项场景运行，连续入选国家级典型案例



以实景三维为底座，“数据换空间”驱动低空经济从示范走向规模化、可管可用可复制。

THANKS
感谢观看！

中医药数据空间与利用

IITC工信人才岗位能力评价培训
数据产业高层次人才培训课程

首席数据官

中医药数据空间与利用

讲师：尤良震 北京中医药大学数智中医慢病防治北京市重点实验室 2025年10月

提纲报告

一、数据要素基础制度构建与方向

二、中医药数据要素特征与基本属性

三、中医药数据基础设施建设与创新

四、中医药数据要素赋能精准诊疗

党中央、国务院高度重视数据资源开发利用

首席数据官

- ▶ 以习近平同志为核心的党中央着眼时代发展大势和国内国际发展大局，高度重视、系统谋划、统筹推进数字中国建设，擘画新时代数字中国建设的宏伟蓝图。
- ▶ 习近平总书记指出“谁掌握了数据，谁就掌握了主动权”“要构建以数据为关键要素的数字经济”“充分发挥海量数据和丰富应用场景优势”“不断做强做优做大我国数字经济”。



总书记论数据

首席数据官



要构建以数据要素为关键要素的数字经济。建设现代化经济体系离不开大数据的发展和应用。我们要坚持以供给侧结构性改革为主线，加快发展数字经济，推动实体经济和数字经济融合发展，推动互联网、大数据、人工智能同实体经济深度融合，继续做好信息化和工业化深度融合这篇大文章，推动制造业加速向数字化、网络化、智能化发展。

——2017年12月8日 习近平在十九届中央政治局第二次集体学习时强调



中国高度重视大数据产业发展，愿同各国共享数字经济发展机遇，通过探索新技术、新业态、新模式，共同探寻新的增长动能和发展路径。

——2019年5月26日 习近平向2019中国国际大数据产业博览会致贺信



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，维护个人信息和商业机密，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6月22日 习近平主持召开中央全面深化改革委员会第二十六次会议强调

中共中央关于进一步全面深化改革推进中国式现代化的决定

首席数据官

——2024年7月18日中国共产党第二十届中央委员会第三次全体会议通过

“构建高水平社会主义市场经济体制”

- 培育全国一体化技术和数据市场
- 健全数据等生产要素由市场评价贡献、按贡献决定报酬的机制

“构建支持全面创新体制机制”

- 推荐教育数字化，赋能学习型社会建设，加强终身教育保障

“构健全宏观经济治理体系”

- 深化金融体制改革，积极发展科技金融、绿色金融、普惠金融、养老金融、数字金融

“健全推动经济高质量发展体制机制”

- 以国家标准提升引领传统产业优化升级，支持企业用数智技术、绿色技术改造提升传统产业
- 健全促进实体经济和数字经济深度融合制度。加快推进新型工业化，培育壮大先进制造业集群，推动制造业高端化、智能化、绿色化发展
- 加快构建促进数字经济发展体制机制，完善促进数字产业化和产业数字化政策体系
- 建设和运营国家数据基础设施，促进数据共享。加快建立数据产权归属认定、市场交易、权益分配、利益保护制度，提升数据安全治理监管能力，建立高效便利安全的数据跨境流动机制

“完善高水平对外开放体制机制”

- 积极应对贸易数字化、绿色化趋势
- 创新发展数字贸易，推进跨境电商综合试验区建设
- 继续实施“一带一路”科技创新行动计划，加强绿色发展、数字经济、人工智能等领域的多边合作平台建设

“深化文化体制机制改革”

- 完善生成式人工智能发展
- 加强网络空间法治建设健全网络生态治理长效机制

“推进国家安全体系和能力现代化”

- 全面加强网络安全体制建设，建立人工智能安全监管机制

我国培育数据要素政策脉络日益清晰

首席数据官

◆近年来，**中国将数据要素作为一种新型生产要素写进国家政策文件**，并不断细化

□ 中国推动数据交易流通工作

- 2023年11月10日，**国家数据局党组书记、局长刘烈宏**出席北京数据基础制度先行区启动会议并作重要讲话。
- 一是**探索数据“三权”分置落地**，让数据放心“供”出来。
- 二是**培育多层次数据流通交易体系**，让更多数据“活”起来。
- 三是**推动数据基础设施建设**，让数据安全“动”起来。

🔗 制度健全

2020年：首次将“数据”与土地、劳动力、资本、技术等传统要素并列列为要素之一写入国家文件，明确数据要素市场培育新格局。

🔗 机制深化

2022年：加快推进大数据产业发展，制定数据要素价值评估框架和评估指南，发展数据资产评估、登记结算、交易撮合、争议仲裁等市场运营体系。

📈 经济赋能

2023年：加速数据资产入账的进程，提升企业财务报表的盈利能力和质量；推动数据在不同场景中发挥千姿百态的乘数效应，促进数据基础资源转化为经济发展新优势。

我国培育数据要素政策脉络日益清晰

首席数据官

◆近年来，**中国将数据要素作为一种新型生产要素写进国家政策文件**，并不断细化

2020年4月，中共中央、国务院《关于构建更加完善的要素市场化配置体制机制的意见》
提出**数据要素化**，加快培育**数据要素市场**。

2022年1月，国务院办公厅《“十四五”数字经济发展规划》
鼓励市场主体探索**数据资产定价机制**，推动形成**数据资产目录**，逐步完善数据定价体系。

2022年10月，国务院办公厅《全国一体化政务大数据体系建设指南》
规范**数据权属、数据定价**、交易规则、建立权责清晰的**数据要素市场化**运行机制。

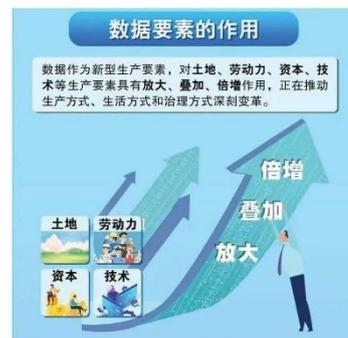
2022年12月，中共中央、国务院《关于构建数据基础制度更好发挥数据要素作用的意见》**数据20条**
创造性地提出建立**数据资源持有权、数据加工使用权和数据产品经营权**，构建**中国特色数据产权制度**体系。

2023年8月，财政部《企业数据资源相关会计处理暂行规定》9月中国资产评估会《**数据资产评估指导意见**》
《**暂行规定**》标志我国向“数据资产”时代迈出实质性一步，《**数据资产评估指导意见**》为数据资产价值评估赋予统一标准。

2023年12月15日，国家数据局《“数据要素×”三年行动计划》
强化场景需求牵引，带动数据要素高质量供给、培育新业态新模式，带动数据产业年均增速超过**20%**，数据交易增长**1倍**。

《关于构建数据基础制度更好发挥数据要素作用的意见》

首席数据官



《关于构建数据基础制度更好发挥数据要素作用的意见》

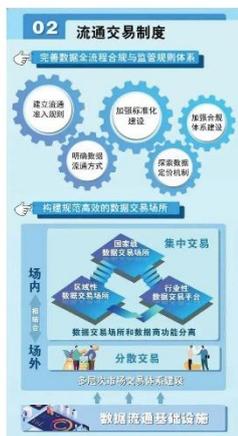
首席数据官



- 1. 数据分类分级制度**
目的在于区分公共/企业/个人数据，制定差异化管理规则。
- 2. 数据确权登记体系**
已经在北京/上海等地为代表试点数据产权登记。
- 3. 数据交易场所规范**
初步建立了国家级+区域性数据交易所建设标准。
- 4. 数据要素价格形成机制**
正逐步探索探索数据资产定价模型。
- 5. 数据要素统计核算制度**
积极开展数据资源入表、GDP核算试点工作等。
- 6. 数据安全合规体系**
围绕合规，出台了大量《数据出境安全评估办法》等法规、规范。
- 7. 公共数据授权运营机制**
推进政府数据开放与开发利用规则，已经出台了《政务数据共享条例》等重要规范。
- 8. 数据要素生态培育政策**
积极培育数据商、第三方服务机构，扶持相关行业发展。

《关于构建数据基础制度更好发挥数据要素作用的意见》

首席数据官



2023年国家数据局成立.....

首席数据官

2023.12.19.《关于加快生活服务数字化赋能的指导意见》

2023.12.23《数字经济促进共同富裕实施方案》

2023.12.25《关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》

2024.1.4《“数据要素”×三年行动计划（2024-2026年）》

2024.2.6.《关于组织开展基本养老保险综合平台试点的通知》

2024.03《关于征集数据基础制度相关问题及对策建议的函》

2024.4《加快数字人才培养支撑数字经济发展行动方案》

2024.4《数字经济2024年工作要点》《数字社会2024年工作要点》

2024.5.14.《关于深化智慧城市发展推进城市全域数字化转型的指导意见》

2024.5《数字中国建设2024年工作要点清单》

2024.5《关于建设碳足迹管理体系的实施方案》

2024.6《数字中国发展报告》

2024.7.3《数据中心绿色低碳发展专项行动计划》

2024.9.21《关于加快公共数据资源开发利用的意见》

2024.9.25《国家数据标准体系建设指南》

2024.10《关于开展向平台企业开放使用监督数据试点推动平台经济规范健康发展的通知》

2024.10《数据领域名词解释》向社会公开征求意见

2024.11《可信数据空间发展行动（2024-2028年）》

2024.11《推动数字金融高质量发展行动方案》

2024.12《关于促进企业数据资源开发利用的意见》

2024.12《关于促进数据产业高质量发展的指导意见》

2024.12《国家数据基础设施建设指引》

2025.1《关于完善数据流通安全治理更好促进数据要素市场化价值化的实施方案》

2025.1《关于促进数据标注产业高质量发展的实施意见》

关于加快公共数据资源开发利用的意见

首席数据官

充分发挥数据要素放大、叠加、倍增效应，
为不断做强做优做大数字经济、构筑国家竞争新优势提供坚实支撑

主线：促进公共数据合规高效流通使用

目标：提高资源开发利用水平

政府指导、市场驱动

尊重规律、守正创新

系统推进、高效协同

加快发展、维护安全

五方面17项具体举措

深化数据要素配置改革 扩大公共数据资源供给	加强资源管理 规范公共数据授权运营	鼓励应用创新 推动数据产业健康发展	统筹发展和安全 营造开发利用良好环境	强化组织实施
<ul style="list-style-type: none"> 统筹推进政务数据共享 有序推动公共数据开放 鼓励探索公共数据授权运营 	<ul style="list-style-type: none"> 健全资源管理制度 完善运营监督 建立健全价格形成机制维护公共利益 	<ul style="list-style-type: none"> 丰富数据应用场景 推动区域数据协作 加强数据服务能力建设 繁荣数据产业发展生态 	<ul style="list-style-type: none"> 加大创新激励 加强安全管理督 鼓励先行先试 	<ul style="list-style-type: none"> 加强组织领导 强化资金保障 增强支撑能力 加强评价监督

关于促进数据产业高质量发展的指导意见

首席数据官

加强数据展业规划布局

数据采集，推动基于5G、物联网等技术的数据实时采集方式创新，支持高精度数据采集，合规采集，自动识别采集等采集技术创新，提升数据采集自动化智能化水平。

数据存储，加快发展的闪存存销，推动湖仓池一体，数据编织、数据压缩等技术创新。支持面向数据分类分级管理使用需求，提供全栈数据存储产品和解决方案。

数据治理，推动数据清洗、数据加工、数据标注、数据集成等业态发展，创新数据开发治理一体化模式，支持人工智能技术在自动化数据处理、数据标注、模型构建、预测分析等领域的应用。

数据分析，大力发展云计算、边缘计算，推进商业智能，数据引擎，数据融合，数据可视化，大数据平台等技术创新和产业化应用，支持机器学习，预训练大模型，深度合成服务算法等迭代创新，加速向量数据库，多模态数据分析等实时检索分析技术突破。

数据交易，发展数据流通交易市场，培育数字经济、数据托管等新业态，提高第三方服务机构专业服务能力。

数据应用，深化产业发展、社会治理，公共服务等领域数据应用，培育“数据即服务”等数据智能应用新产品新服务新业态，支持大模型应用创新发展、大力重点行业**高质量数据集**建设。

数据安全，支持数据加密、防勒索、容灾备份、数据冗余等技术产品推广应用，加强多因子身份认证，端到端加密、零信任安全等技术创新，加快突破数据空间、区块链、隐私计算、匿名化等数据可信流通技术，发展数据安全监测预警，**数据合规检测**，人工智能数据安全等服务业态。

数据基础设施，推动智能芯片，服务器和存储器等持续迭代创新，发展通算，智算和品算，加快网络、算力、**数据流通利用**、数据安全等基础设施建设。

提纲报告

一、数据要素基础制度构建与方向

二、中医药数据要素特征与基本属性

三、中医药数据基础设施建设与创新

四、中医药数据要素赋能精准诊疗

数据要素与数据要素化

数据要素化：数据资源化，具有数据资产与数据资本

术语	含义
数据	是指任何以电子或其他方式对信息的记录。数据在不同视角下被称为原始数据、衍生数据、数据资源、数据产品和服务、数据资产、数据要素等
数据资源	是指具有价值创造潜力的数据的总称，通常指以电子化形式记录和保存、可机器读取、可供社会化再利用的数据集合。
数据要素	是指投入到生产经营活动、参与价值创造的数据资源
数据资产	是指特定主体合法拥有或者控制的，能进行货币计量的，且能带来经济利益或社会效益的数据资源。
数据要素市场化配置	是指通过市场机制来配置数据这一新型生产要素，旨在建立一个更加开放、安全和高效的数据流通环境，不断释放数据要素价值。

数据要素的特征

数据要素化的“五新”特征

数据要素化“五新”特征

- 新要素
- 新空间
- 新要求
- 新设施
- 新产业

数据要素的二十大特征

数据要素的二十大特征

- 非稀缺性
- 虚拟性
- 易复制性
- 原始性
- 非消耗性
- 多环节性
- 非均质性
- 非排他性
- 交易不稳定性
- 边际效应递增性
- 场景依赖性
- 时效性
- 价值不确定性
- 阿罗信息悖论
- 协同性
- 功能替代性
- 规模性
- 多样性
- 流动性
- 可加工性

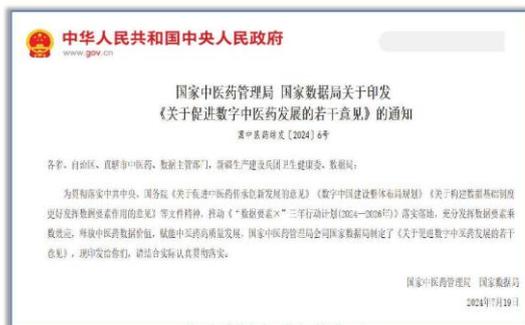
数据要素五大理论问题

首席数据官



数智要素赋能中医药高质量发展

首席数据官



充分发挥数据要素乘数效应，释放中医药数据价值，赋能中医药高质量发展。

要深刻把握数据要素基本特征和战略地位，以发挥中医药数据的赋能作用为导向，推动大数据、人工智能等新兴数字技术逐步融入中医药传承创新发展全链条各环节，用3-5年时间，全力打造“数智中医药”，为中医药现代化发展提供有力支撑。

2024年《关于促进数字中医药发展的若干意见》

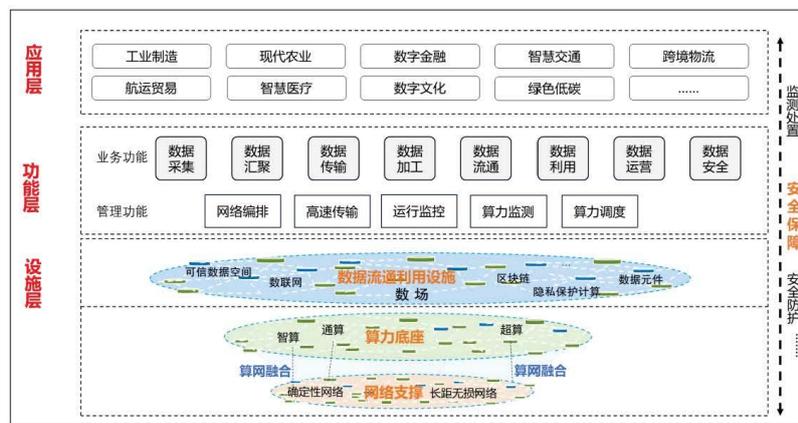
提纲报告

一、数据要素基础制度构建与方向

二、中医药数据要素特征与基本属性

三、中医药数据基础设施建设与创新

四、中医药数据要素赋能精准诊疗



数据基础设施及网络、算力设施总体架构图

可信数据空间：数据流程基础设施，是基于共识规则，联接多方主体，实现数据资源共享共用的一种数据流通利用设施，是数据要素价值共创的应用生态，是支撑构建全国一体化数据市场的重要载体。可信数据空间须具备数据可信管控、资源交互、价值共创三类核心能力。



数场：是依托开放性网络、算力和隐私保护计算、区块链等各类关联功能设施，面向数据要素提供线上线下资源登记、供需匹配、交易流通、开发利用、存证溯源等功能，支持多场景应用的一种综合性数据流通利用设施。



数联网由数据流通接入终端、数据流通网络、数据流通服务平台构成，提供一点接入、广泛连接、标准交付、安全可信、合规监管、开放兼容的数据流通服务。



中医药可信数据空间建设意义

首席数据官

1 是构建高质量的数据要素市场的重要组成部分

- 中医药行业数据要素市场具有专业特性：数据来源、类型、价值、隐私敏感及标准等；
- 创新中医药数据流通模式与交易场景，释放更丰富的数据要素价值，支撑全国一体化数据市场建设。

2 是遵循中医药发展规律，破解中医药发展难题的需要

- 解决中医药行业核心痛点，推动中医药数字化、智能化转型，推进中医药领域的标准化建设；
- 发展新质生产力构建数据驱动的中医药科学研究范式，突破中医药传承创新发展难题；

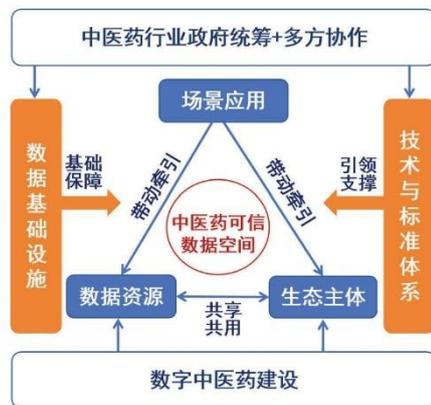
3 是加快推进中医药现代化、产业化与国际化，应对风险挑战的需要

- 为中医药现代化提供数据支持和创新动力，推动中医药在临床、科研、产业等领域的现代化进程；
- 将加快推进健康数据产业生态构建，协同推进数据产业发展；
- 助力中医药传承与文化传播，支持中医药数据要素跨境流通，在中医药国际化中发挥重要作用。

中医药可信数据空间建设路径与实施思路

夯实中医药可信数据空间基础设施建设

首席数据官

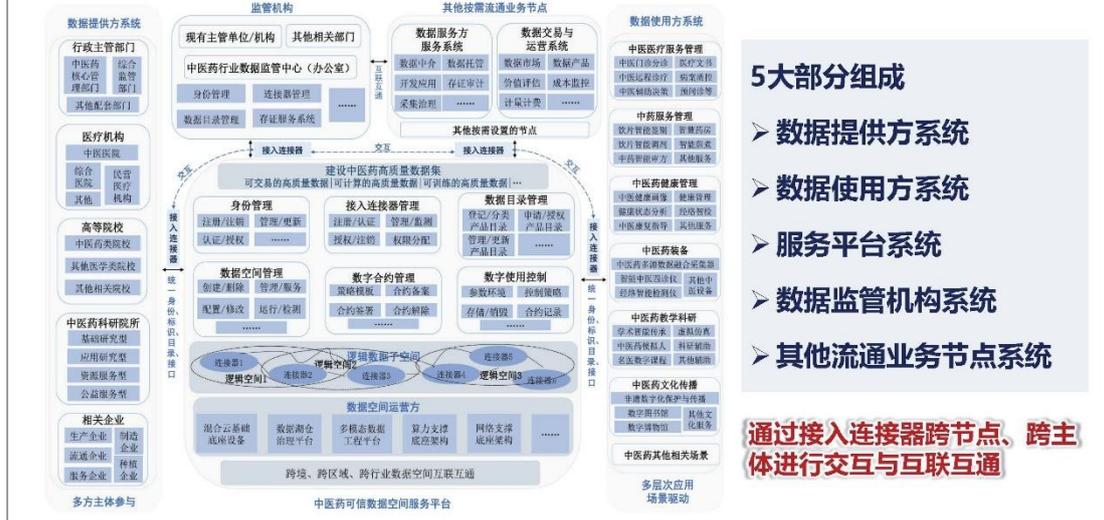


中医药可信数据空间基础设施建设架构

- 数字中医药建设促进数据资源供给
- 政府统筹构建中医药数据基础设施
- 技术与标准体系创新促进中医药数据能力实现
- 场景应用驱动中医药数据生态协同共建

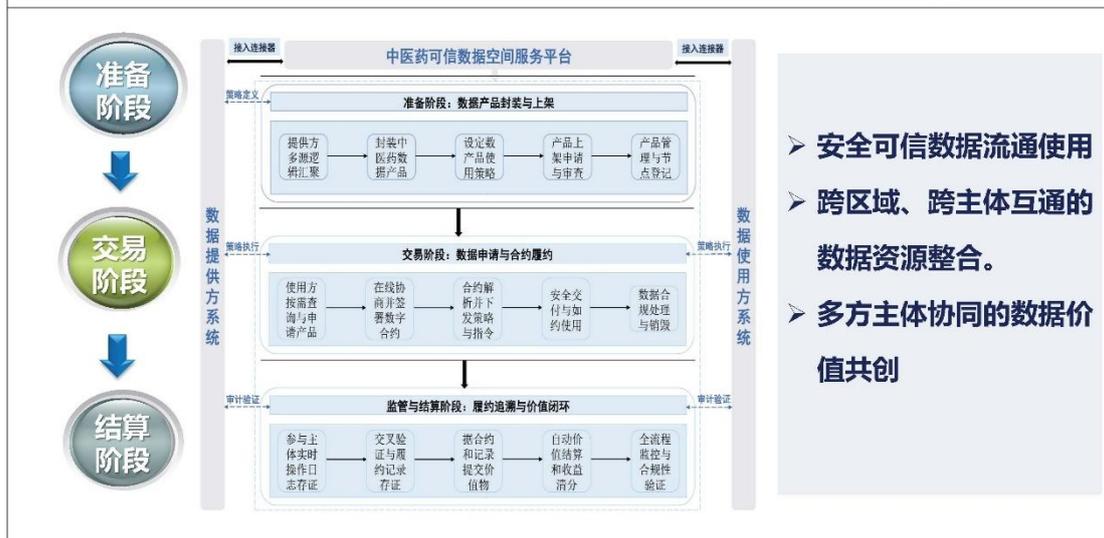
构建中医药可信数据空间的功能架构

首席数据官



明晰中医药可信数据空间技术特点与业务流程

首席数据官



提纲报告

一、数据要素基础制度构建与方向

二、中医药数据要素特征与基本属性

三、中医药数据基础设施建设与创新

四、中医药数据要素赋能精准诊疗

发展病证结合精准诊疗研究模式

首席数据官

· 14 · WORLD CHINESE MEDICINE Nov. 2009, Vol. 1, No. 1

病证结合是中西医结合临床的最佳模式

张京春¹ 陈可冀²

(¹ 中国中医科学院西苑医院心血管病科, 北京市海淀区西苑操场1号, 100091)

摘要 “病”反映疾病发生、发展的全部过程,“证”反映疾病过程中某一阶段的病理特征。在疾病诊断、治疗过程中,辨病与辨证结合已经得到广泛的认可。通过多年来临床、科研工作的体会,认识到病证结合是中西医结合的最佳模式。

关键词 病;证;中西医结合

西医诊疗
以病变局部为主要关注点的还原论下系统论诊疗模式,多精于诊断而略于防治。

中医诊疗
在系统性象思维指导下的认为机体内在本质状态可由外在神、色、形、态综合反映,据象言证,随证治之。

病证结合精准诊疗研究模式

“显象”容易把握,“隐象”相对困难。通过数智融合的方式将“显象”表征“隐象”,病证结合精准诊疗是中医象思维指导下的病证结合模式,是中西医结合走向中西医融合的高级阶段。

病证结合精准诊疗研究模式是在中医理论指导下,借助智能化诊疗装备,准确把握潜在中医脏腑功能状态与健康疾病转归变化,进行中医证候与辨证论治过程科学化表达以及疾病过程的“病-证”规律定量表征,进而实现病证结合的精确诊断、精确治疗与精确评价。

发展病证结合精准诊疗研究模式

首席数据官



中医药诊疗现状与需求：传统中医药--循证中医药--数智中医药

首席数据官



数智融合中医精准诊疗发展新动能

首席数据官



中医数智化精准诊疗研究现状

首席数据官

智能中医学概论
主编 田洪华 商洪才
Generality of Intelligence Traditional Chinese Medicine

中医智能诊断装备 (“望、闻、问、切”四诊)

发展历程: 单诊仪 → 多诊仪 → 云中医 → 云智诊

发展需求: 可视化、客观化、小型化、移动化、可穿戴化、智能化

发展目标: 临床诊断三类医疗器械产品 (尚未满足临床诊断要求)

中医智能治疗装备

发展现状: 针灸、推拿、热敷等非药物智能治疗装备以及中医辅助治疗云平台技术发展迅速

技术问题: 智能治疗与诊断技术衔接不够 数字化与智能化程度不高

发展目标: 中医智能“诊疗一体化”设备

发展趋势 应以疗效为导向中医辨证论治数智化能力提升

- 发展更准、更全、更快的信息感知处理技术
- 多源、多维、多态的融合分析诊断技术方法
- 探索中医诊疗无边、无时、无人的未来医疗场景

中医数智化精准诊疗研究问题与技术瓶颈

首席数据官

多元信息来源与采集技术不足

- 人体病症状态信息的**标准化、规范化不够**
- 中医“四诊”信息规范可靠性与精确度不足、**四诊融合度较低**
- 多元信息采集**如何真实反映中医辨证论治过程**

智能诊疗有效性无法保障

- **临床疗效是中医精准诊疗的核心**
- 当前在新技术、新方法的研制中往往忽视疗效评价，盲目推广
- 难以证明能否使患者真正受益，不利于中医诊疗的长期发展

真实世界疾病状态关注不足

- **慢性共病成为疾病人群真实世界疾病状态**，且表现为慢性持续进展性往往都是由“未病”转变发展而来
- “未病”状态具有多维性、动态性、系统性、复杂性及显著的个体差异性
- “未病”状态发生、发展的生物学基础尚不明确，缺乏可定量表征的客观指标

面向中医精准诊疗的中医辨证论治数智化研究技术瓶颈

- 如何收集真实反映辨证论治过程的有效诊疗信息，实现诊断信息的准确性与可靠性
- 如何形成多元诊疗数据与可靠证候诊断的连接机制，实现“病-证-方药”数字化表征以提高智能辨证论治水平
- 如何定量表征“未病-已病-变病”病证演变规律，提升中医精准诊疗能力

关键技术瓶颈：中医证候与辨证论治过程难以科学量化表达以及疾病过程的“病-证”规律难以定量表征

“人工智能+中医药”高质量数据集建设与应用能力不足

首席数据官

● 通用领域、垂直领域以及具身智能领域的高质量数据供给仍有不足

中医药公共数据开放利用存在**广泛壁垒**，专门面向人工智能发展的高质量行业**数据集匮乏**。具身智能领域真实交互数据采集不足，主要原因在于智能机器人与环境的交互数据获取困难且成本高昂，**同时，中医药行业数据采集缺乏统一的参照标准。**

● 高质量数据的合成、处理和利用技术亟待提升（数据治理能力亟待加强）。

利用深度学习和强化学习生成高精度、多样化合成数据的技术在成熟度和应用范围上急需突破。随着社会自动化和智能化程度的不断提高，对数据处理的要求也不断提升，因此急需针对结构化、半结构化和非结构化数据的技术进行迭代优化，进一步提高数据处理效率。

● 数据主体、商业模式、应用场景开发尚不成熟

目前，缺乏“数据+人工智能”模式的高质量数据汇聚和治理主体，具备大规模数据汇聚管理分析能力的公司数量不足。中医药领域的公共数据授权运营主体目前仍在培育中，数据集构建和运营利用的商业模式发展还不够成熟。

● 政策制定以及解读能力不足，高质量数据集建设思路不够明确

面向新一代人工智能模型训练和场景应用的高质量数据集专项规划和支持政策尚未出台，其建设、运营、流通、利用等方面举措有待进一步细化。在数据采集方面，各领域数据缺乏适用的标准规范；在数据使用方面，缺少面向大模型和具身智能模型训练的数据共享和流通促进机制，一定程度上限制了模型能力的快速提升。



中国电子安全可信数据空间建设与运营之路

IITC工信人才岗位能力评价培训
数据产业高层次专业人才培养课程

首席数据官
GOVCDO

破局与实践

中国电子安全可信数据空间建设与运营之路

——构建“安全合规的数据高速公路”的核心基石

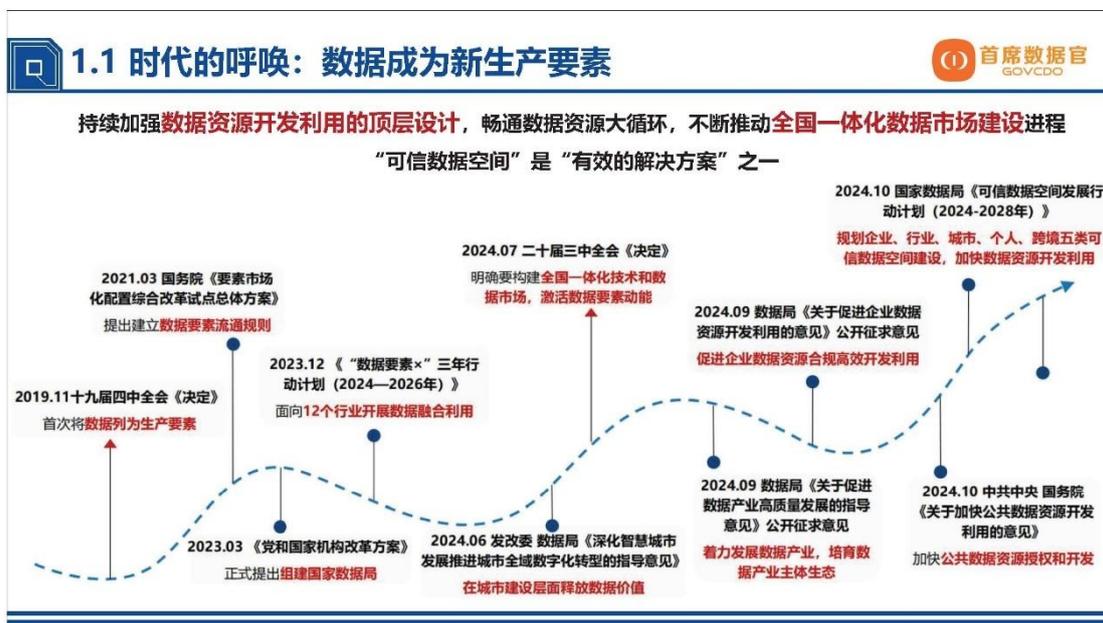
讲师：刘颖 中国电子

2025年10月

目录

CONTENTS

- 一、共识与困境
- 二、破局之道
- 三、中国电子实践
- 四、运营之要





1.2 现实的困境：数据流通的“三不”难题





信任机制缺失
导致数据 **“不敢”** 共享

各方对参与**主体身份、资质**存在疑虑，且提供方无法有效控制、记录数据流通使用全过程，导致**数据滥用、泄露、篡改风险较高，各方不敢共享数据**

例如，医院诊疗核心数据，“理、法、方、药”数据链条割裂。辨证信息（面诊、舌诊、脉诊影像）、方剂、中药溯源数据分属不同系统，形成数据孤岛，该类数据的安全和隐私合规要求极高，数据所有权与使用权模糊，各方不敢共享。

价值牵引缺失
导致数据 **“不愿”** 共享

各主体缺少**高价值的共性应用场景**牵引，不明确数据共享带来的利益与价值，因此**不愿共享数据**

例如，由于缺少共性价值场景牵引，目前在数据接入、数据开发、数据应用等数据价值创造环节各主体之间无法进行合理的利益分配

流通机制不畅
导致数据 **“不能”** 流通

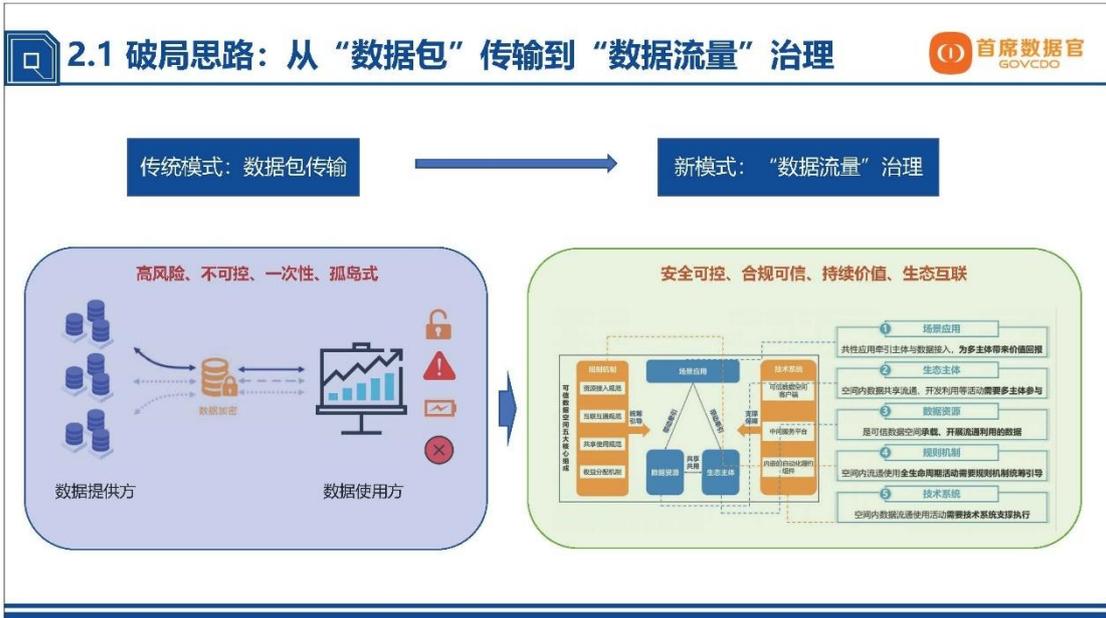
由于各**机构间数据模型不统一、数据语义难互通**，导致跨主体数据难以互联互通，因此**数据不能流通**

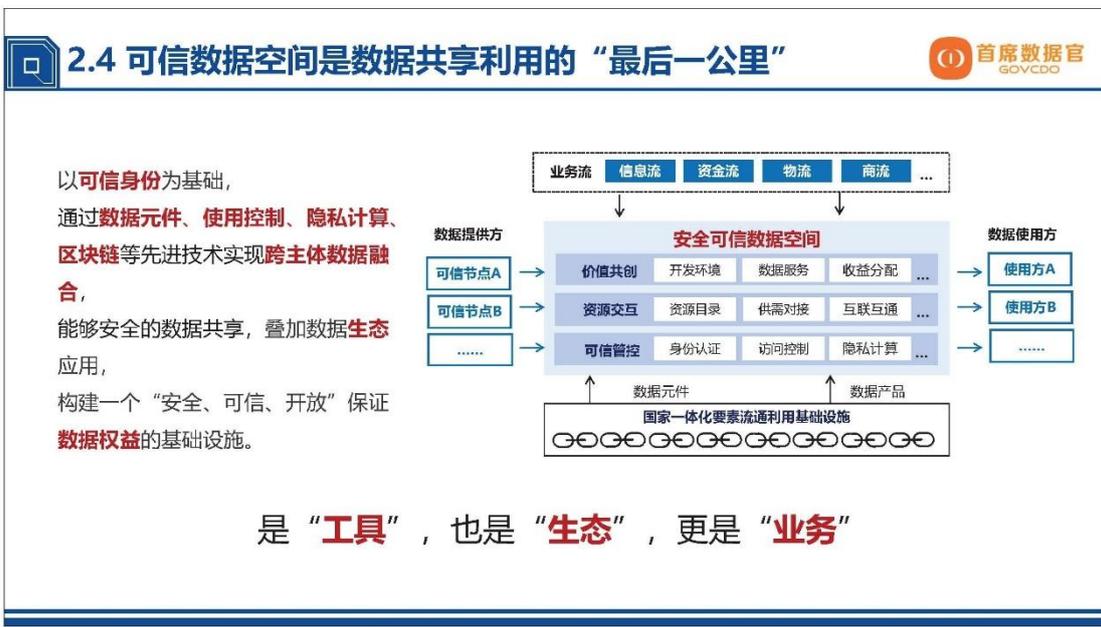
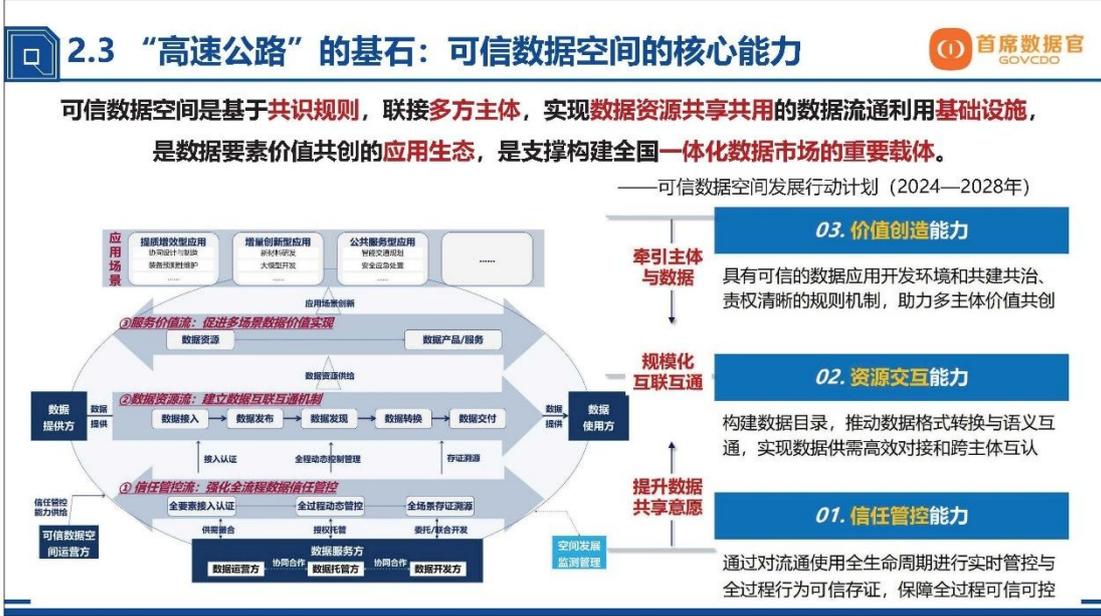
例如，医院的诊前健康状态数据分散于智能穿戴设备、健康APP、社群问诊等地方，该数据格式不一、难以结构化

02

破局之道

—通往未来的数据“高速公路”—







3.1 我们的使命与定位：国家队的选择

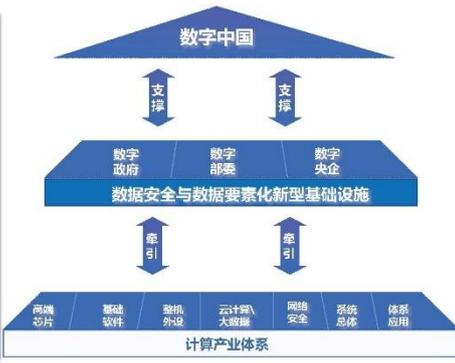
 首席数据官
GOVCDO

国内首家由中央企业设立的数据产业集团

定位：数字中国建设和数字经济发展 的核心战略科技力量，数字政府 建设、数据安全与数据要素化治理的主力军，数据安全与数据治理技术 的策源地，数据产业链 的链长。

2022年12月25日揭牌 2023年4月12日正式成立





3.2 中国电子与清华大学、国家智库联手破题

中国电子联合清华大学成立**数据治理工程联合研究院**，联合中国经济体制改革研究会、中国经济改革研究基金会成立**数据要素市场化配置综合改革研究院**，采用工程化路径，率先开展数据要素市场化实践探索。

校企合作

理论策源

政策智库

要素改革

工程路径

- 特点：制度、市场、技术三位一体
- 优点：体系性安全、规模化开发、产品化流通、平台化运营

3.3 中国电子破解数据要素安全流通的关键

定义**数据金库**和**数据元件**两个产品，是破解数据安全和数据要素市场化痛点的关键。

数据金库

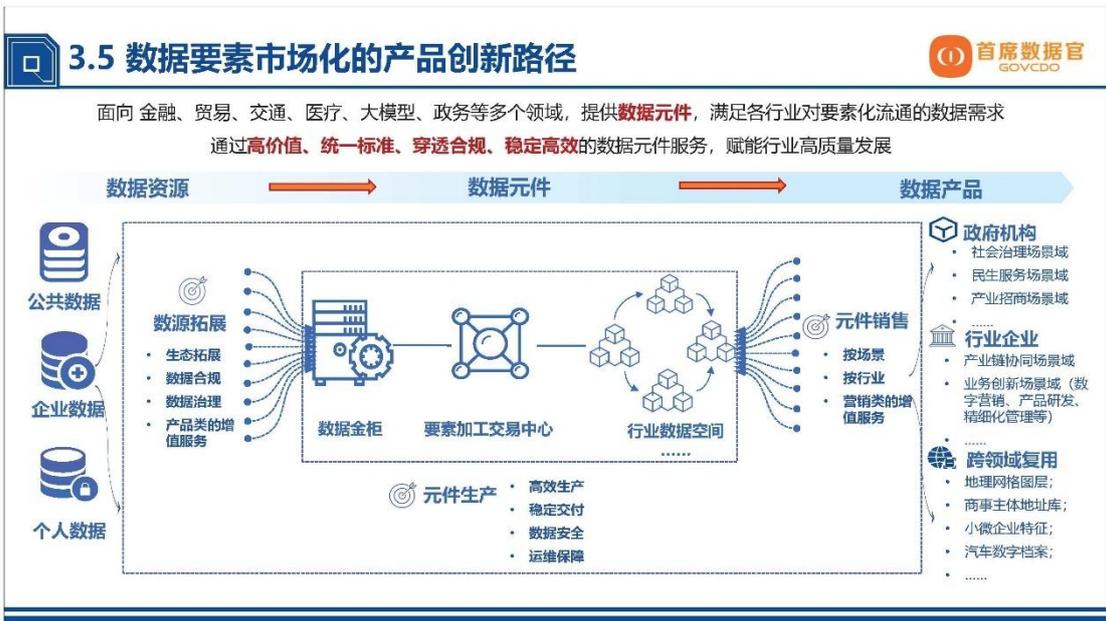
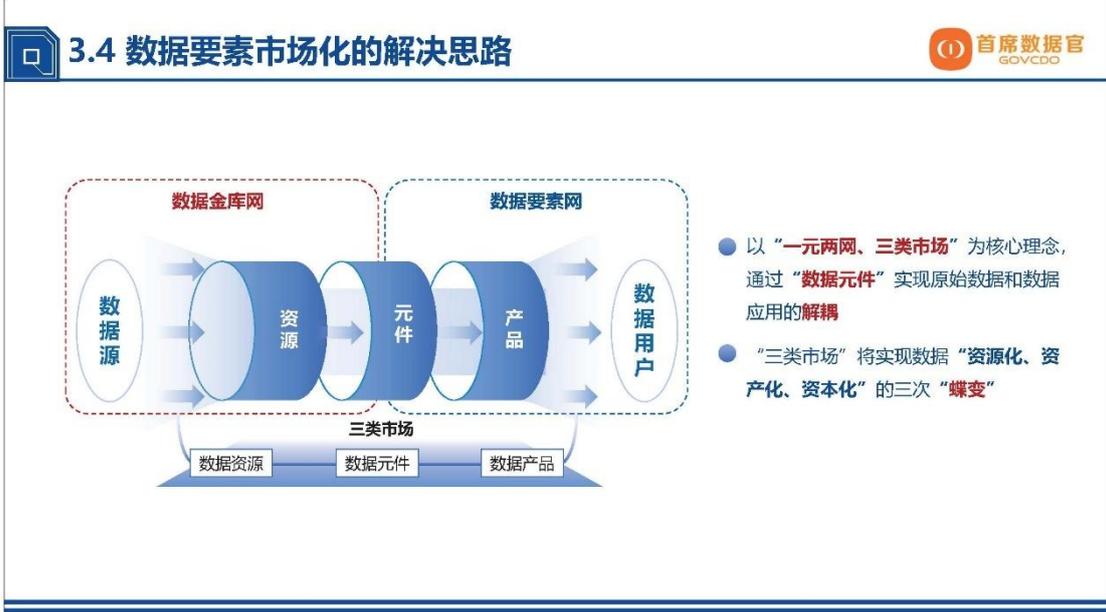
存储关键数据，保障数据安全

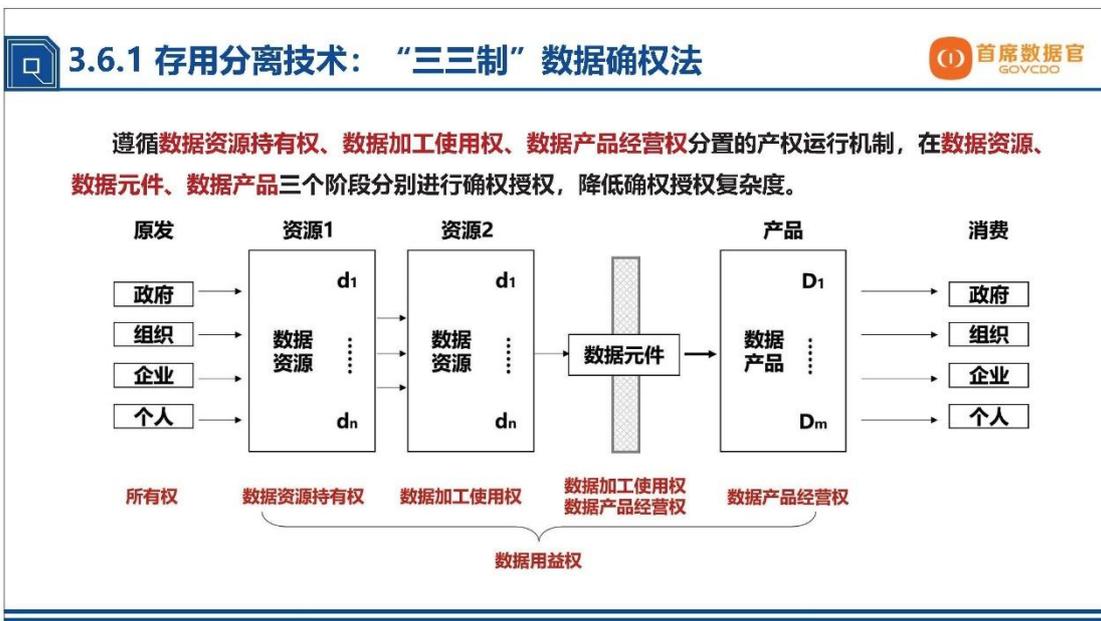
定义：由主管部门监管，统一标准、自主可控、安全可靠，存储核心数据、重要数据、敏感数据和数据元件的数据存储和管理设施。主要部署在政府、组织、大型企业。

数据元件

有效实现原始数据和数据应用“解耦”
推动数据低成本、高效率的流通和交易

定义：数据元件是通过对数据脱敏处理后，根据需要由若干相关字段形成的数据集或由数据的关联字段通过建模形成的数据特征。

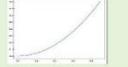




3.6.2 计量交易技术：数据价值计量模型

数据信息量根据香农信息论表达式计算，通常信息量越大，数据的价值越高，不同行业基于信息量的构建与价值相关的表达式，结合数据体量和数据质量系数，给出基于信息量的指导价。

$$\text{信息价值 } I(X) = V(N) \cdot Q(Z) \cdot D(X)$$

元件体量系数	元件质量系数	信息密度
$V(N) = \frac{1}{1 + e^{-N/C}}$ <p>式中： N—元件生产使用的数据体量 C—元件标准增量</p> 	$Q(Z) = (Z^T \beta)^2$ <p>式中： Z—数据质量评估指标矩阵 β—指标权重系数</p> 	$D(X) = h(E(X))$ $E(X) = -\sum_{i=1}^n p(x_i) \log_2(p(x_i))$ <p>式中： E(X)—数据元件X在不同领域的信息熵 h——价值与信息熵的函数</p> 

- 数据元件价值与数据体量、数据质量、信息密度成正比相关
- 元件体量与数据元件使用的数据量成正比相关，体量到达一定值，有一个衰减的过程
- 元件质量由质量评估指标矩阵，以及各指标的权重系数决定，前期质量差，质量变动带来的价值变化较小
- 信息密度可基于信息熵进行建模评估

3.6.3 计量交易技术：三阶段区别定价

1 数据资源市场

- 以价值为基础，成本法为主、收益法为辅

2 数据元件市场

- 以价值法、成本法为基础，市场法为主

3 数据产品市场

- 以价值为基础，按照市场法定价

协议指导价模型

元件价值评估 <ul style="list-style-type: none"> 元件体量算法 元件质量算法 信息密度算法 	元件成本评估 <ul style="list-style-type: none"> 数据成本 模型成本 领域调节系数
--	--

$$P_y^1(X) = (1 + r) \cdot g(I_y(X), C_y(X))$$

式中：
y: 预期收益率 $I_y(X)$: 元件X的信息价值
 $C_y(X)$: 元件X的成本 g: 成本价和价值反映价格的联合定价函数

竞价指导价模型

元件市场评估价

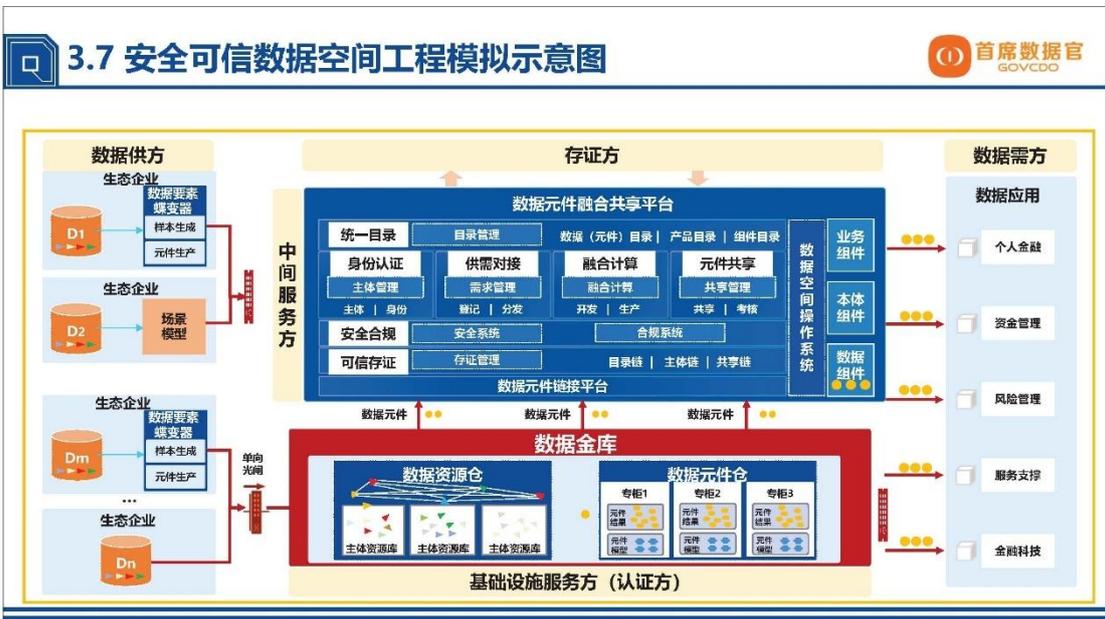
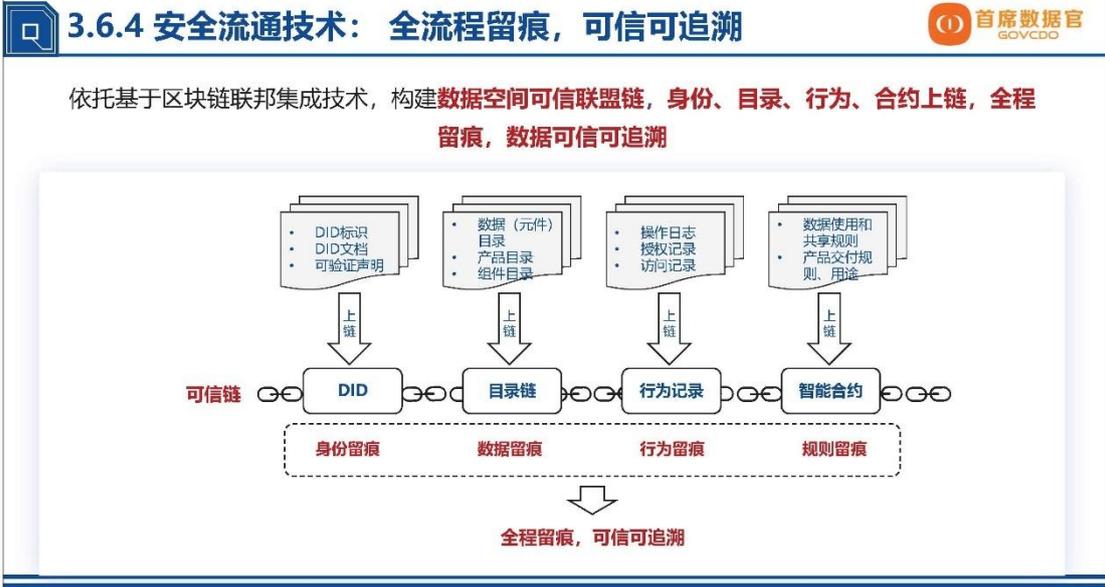
稀缺性	时效性
完整性	规模大小
历史参考价	市场调节因子

$$P_y^2(X) = \beta P_y''(X)$$

$$P_y''(X) = f(S, T, N, W, C_0, P_y'(X))$$

式中：
S: 稀缺性 T: 时效性 N: 规模大小 W: 完整性 C_0 : 历史参考价

- 协议指导价模型：包括元件价值评估和元件成本评估两个维度
- 竞价指导价模型：市场评估价乘以市场调节因子，由数据元件的市场价格影响因子经过估价模型计算产生



3.10 加快技术创新流，破解“不能用”的问题

以人工智能、大数据、云计算、区块链等数智技术的应用全面**优化和重塑**数据价值链结构、数据关系和服务监管方式，大模型等智能体及其工具成为数据产品形态、规模与类型的物理表征，数据流通利用对**数据流通基础设施**的需求进一步增强。

【基于数据空间互联互通提供高质量数据集有通用基础大模型，同时搭建大模型与智能体开发框架，助力智能化升级】

智能体工厂

Agent 编排 | Agent 应用

任务下发

交互对话

智能应用

- 智能调度大屏
- 智能图表问答
- 智能文档问答
- 智能写作
- 智能导办
- 智能以图搜人

智能决策

- 市场拓展
- 经营建议
- 流程优化

智能化升级

- 行业分析
- 知识图谱
- 智能报告
- 智能问答

API 调度接口 | 行业小模型 调度小模型 | 知识库 知识获取

数据供给 | 分析推理

高质量数据集 | 通用基础大模型

城市数据空间/行业数据空间/企业数据空间

GOVCDO 首席数据官

3.11 CEDC “安全可信数据空间产品”定义

数据元件+隐私计算+使用控制
打造覆盖可信管控、资源交换、价值共创等核心能力的**数据空间产品**

唯一融合6条技术路线产品

CEC生态自研全国产化一体机

安全可信数据空间平台

安全可信数据空间一体机

数据资源方

- 市场数据: 价格行情, 供需信息, 交易订单, 结算数据
- 公共数据: 法人信用, 企业能耗, 证券信息, 知识产权
- 企业/行业数据: 生产数据, 物流数据, 财务数据, 设备数据

数据供给

数据需求方

- 核心企业: 产业链协同、降本增效
- 核心企业数字化转型
- 产业链数智大脑
- 行业标准及数据规范
- 商业银行: 风险管控及额度增益

空间合作生态

- 140+ 数据生态
- 800+ 应用生态
- 3600+ 数据元件

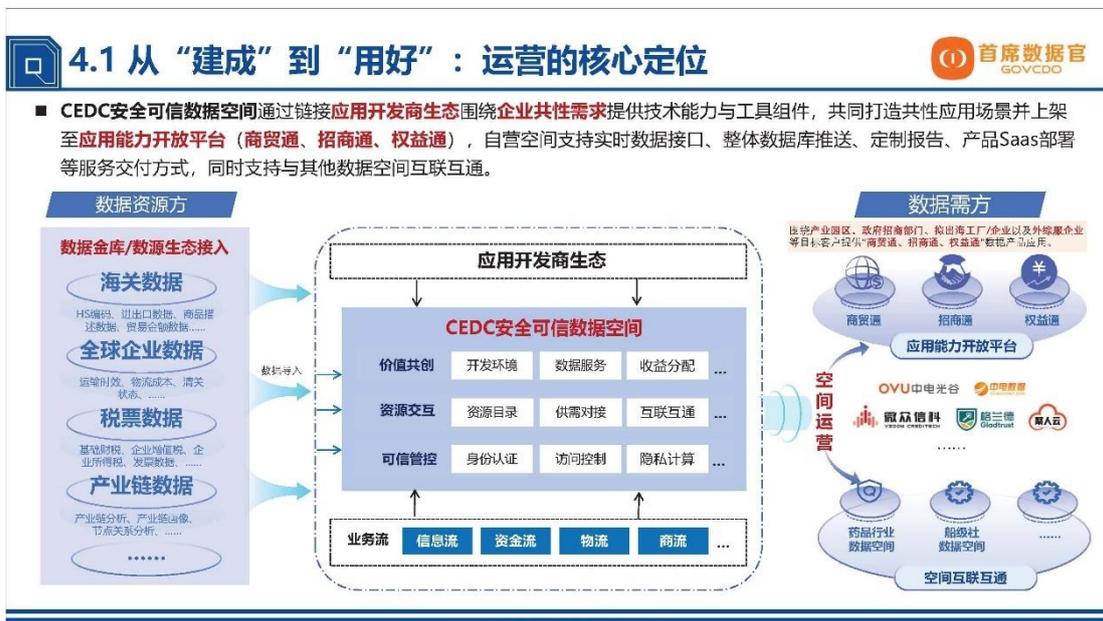
空间行业实践

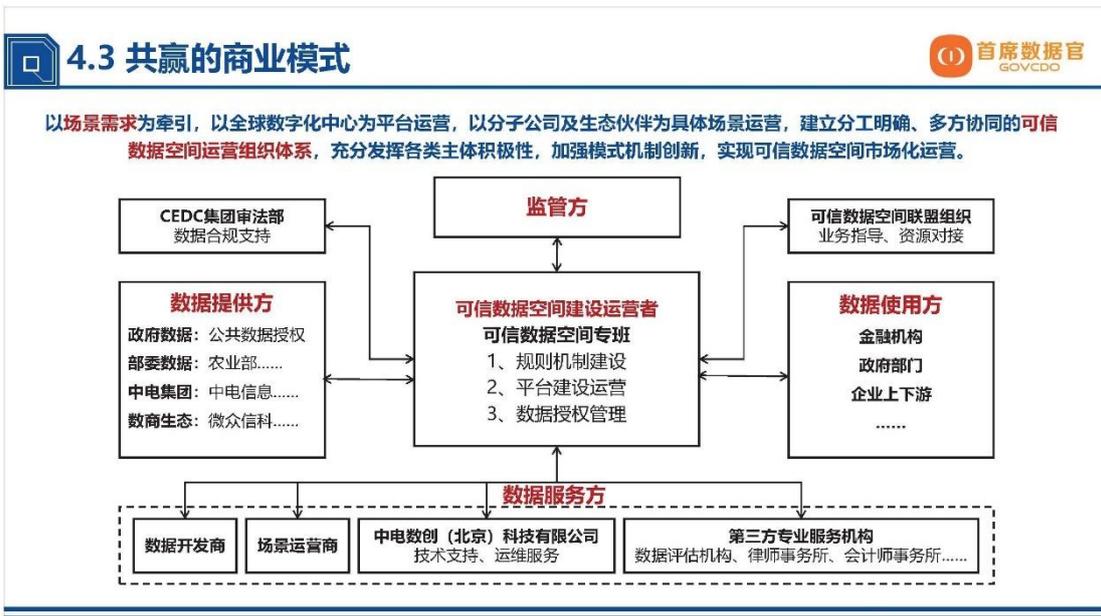
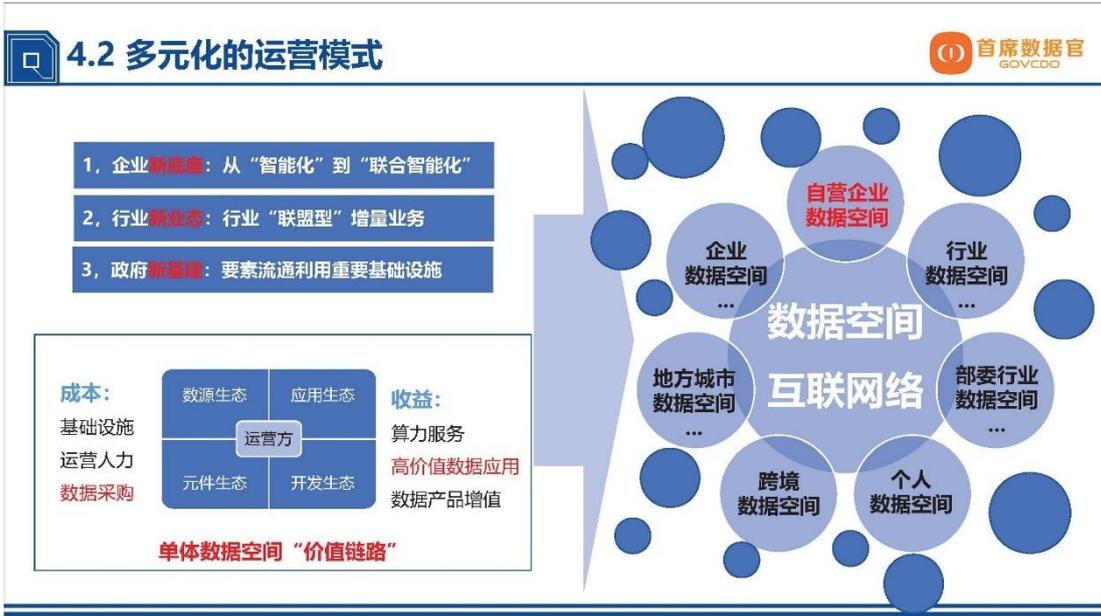
供应链、金融、交通、医疗、外贸、园区等**多行业场景**实践

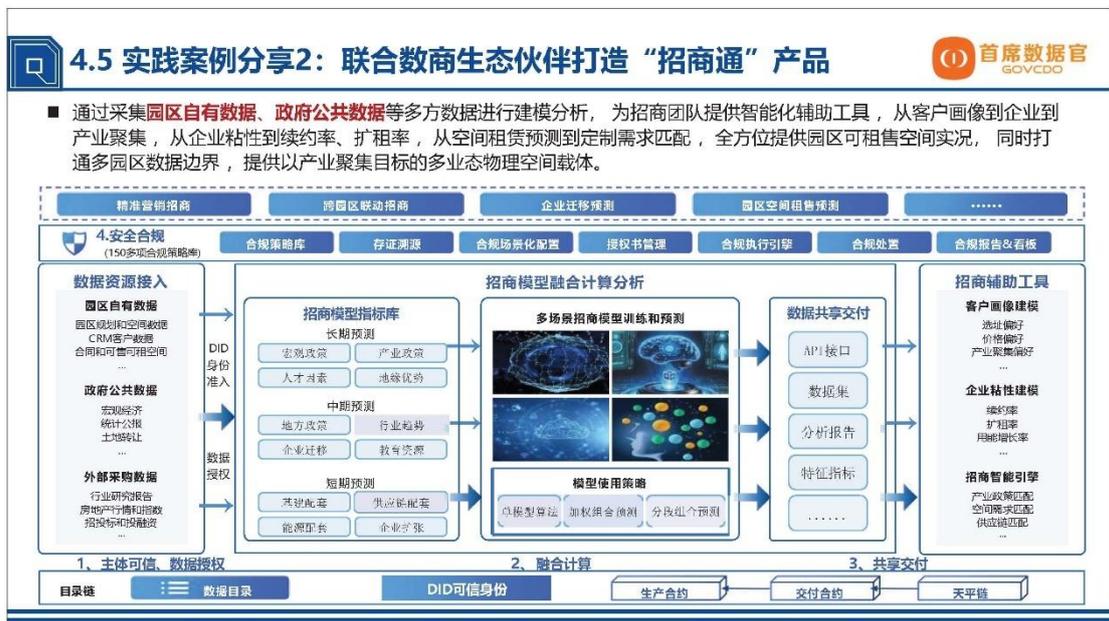
GOVCDO 首席数据官

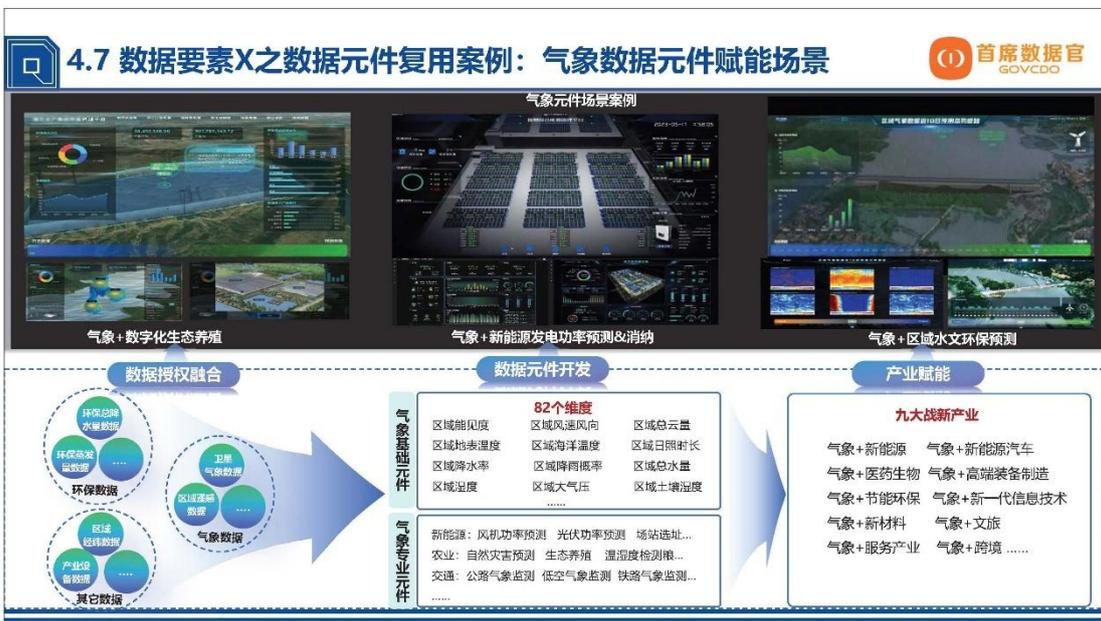
04 运营之要

——如何让“高速公路”产生经济价值









共建·共创·共享

—— 欢迎更多数商加入生态 ——



刘颖
空间产线解决方案专家
TEL: 17600122199
E-mail: liuying@cecdt.com.cn
CEDC安全可信数据空间试用地址申请: t-space.cecdt.com.cn