



业务数据治理中台实践

贝壳找房 李清华 2021/08/28

内容概览

01 介绍

02 数据治理目的及内容

03 结合公司特点聚焦治理范围

04 中台侧实践的建设内容及思路

05 治理项目的目标管理

06 产品及运营落地经验

01 公司及个人介绍



贝壳找房

科技驱动的新居住服务商，致力于为3亿家庭提供全面、可靠的品质居住服务



二手房



新房



租赁



装修



其他

02 治理目的及内容

数据治理目的：提升数据的使用价值

共享

知道公司有哪些数据

- 1、业务视角：业务内部多载体（PRD、邮件、wiki、共享文档）知识记录，受入离调转的影响，“考古成本”高
- 2、公司视角：跨组织数据流通黑盒，数据不流通

准确

业务知道该用哪个数据

- 1、数据语义记录不明晰
- 2、业务自我迭代变更过程中数据记录不同步
- 3、好多数据不知道该用哪个？
- 4、业务迭代数据定义变更如何准确修改定义及触达关注方

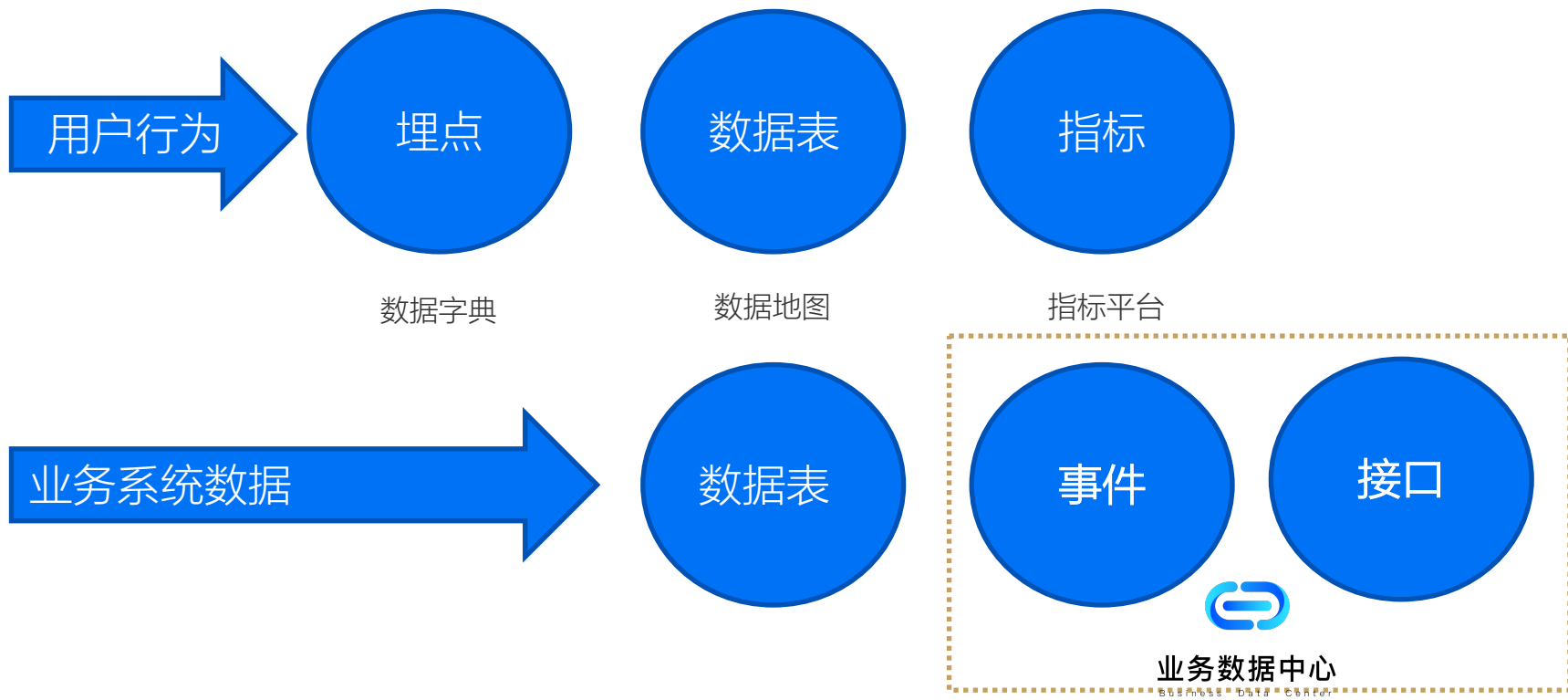
可用

数据质量需要保障

- 1、已经下线了文档还存在
- 2、不同数据健康状态监控
- 3、数据血缘追溯问题排查
- 4、不同数据格式导致下游消费解析成本过高
- 5、上游变更下游无感知

数据治理范围

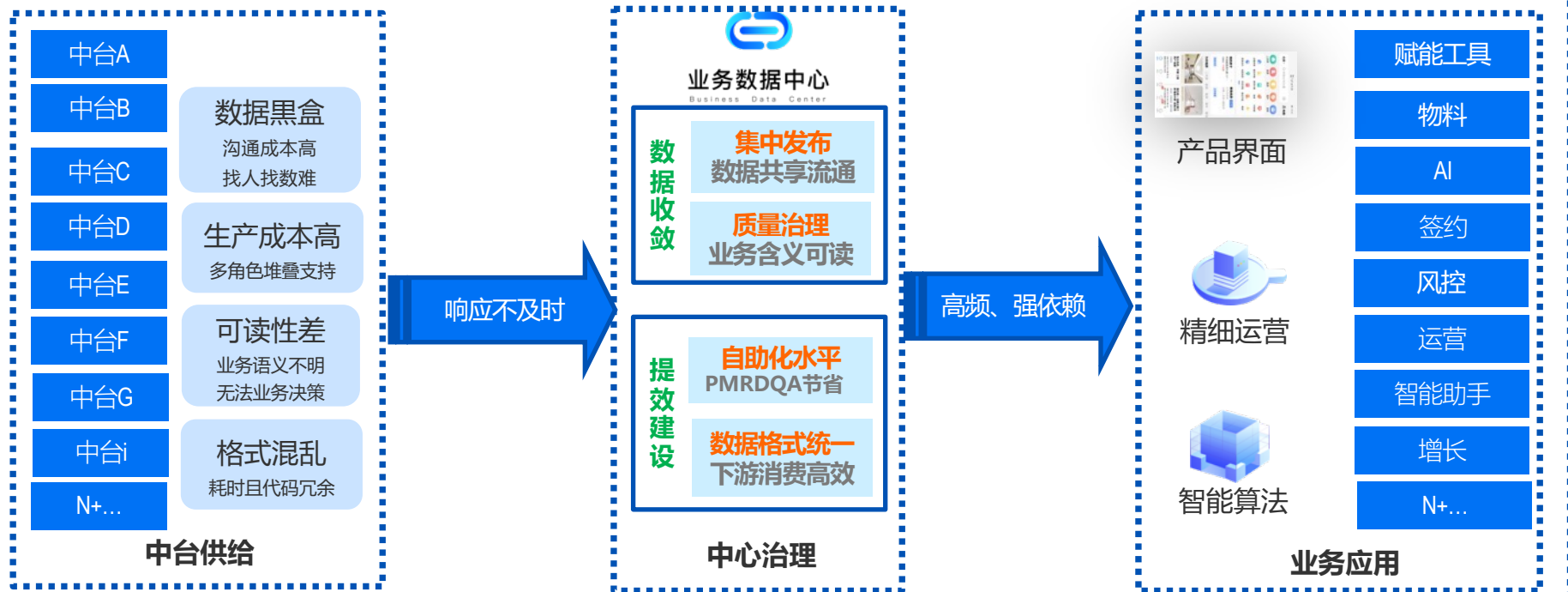
- 数据治理贯穿整个数据生命周期，是复杂且需要长期建设的学科，**聚焦**当下的目标很重要；
- 每家公司数据治理面临的核心问题千差万别，了解后才能更容易把控好自己的方向，不至于走偏；



03 结合公司特点聚焦治理范围

数据治理的公司环境-系统特点

数据治理：保障中台数据可被业务广泛准确、高效使用

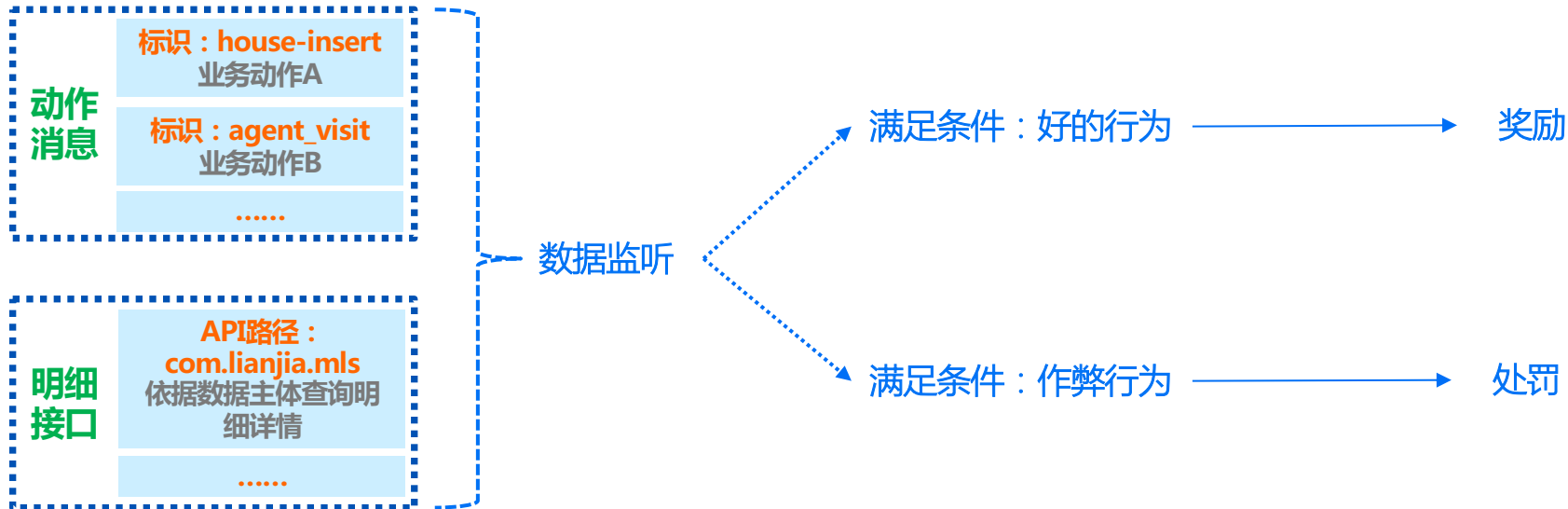


数据治理的公司环境-业务特点

业务运营管理的诉求驱使数据需要完备，实时数据要求高+数据量丰富+共享+业务语义可读+可追溯

重点治理对象 📌

- 接口：围绕业务数据主体关联明细的通用接口；
- 事件：围绕核心业务SOP作业节点的Kafka消息；



数据收敛必要性

取数频率高

半年内上千+增量的接口和事件请求

单次数据获取成本高



一个场景举例：业务动作A，完成后可获取奖励

问题1:判断所属中台——联系名册——官网门户——历史PPT——工作大群——接口人



原因：高流通数据中心化收口的范围不够
目标：数据集中查询申请，降低找人找数成本

问题2:

-部分证件数据收口在A系统，部分在其他系统B；
-数据业务语义无备注，拿过来数据不知道怎么用，不知道数据哪个字段代表什么含义。找产研约会、沟通响应不及时



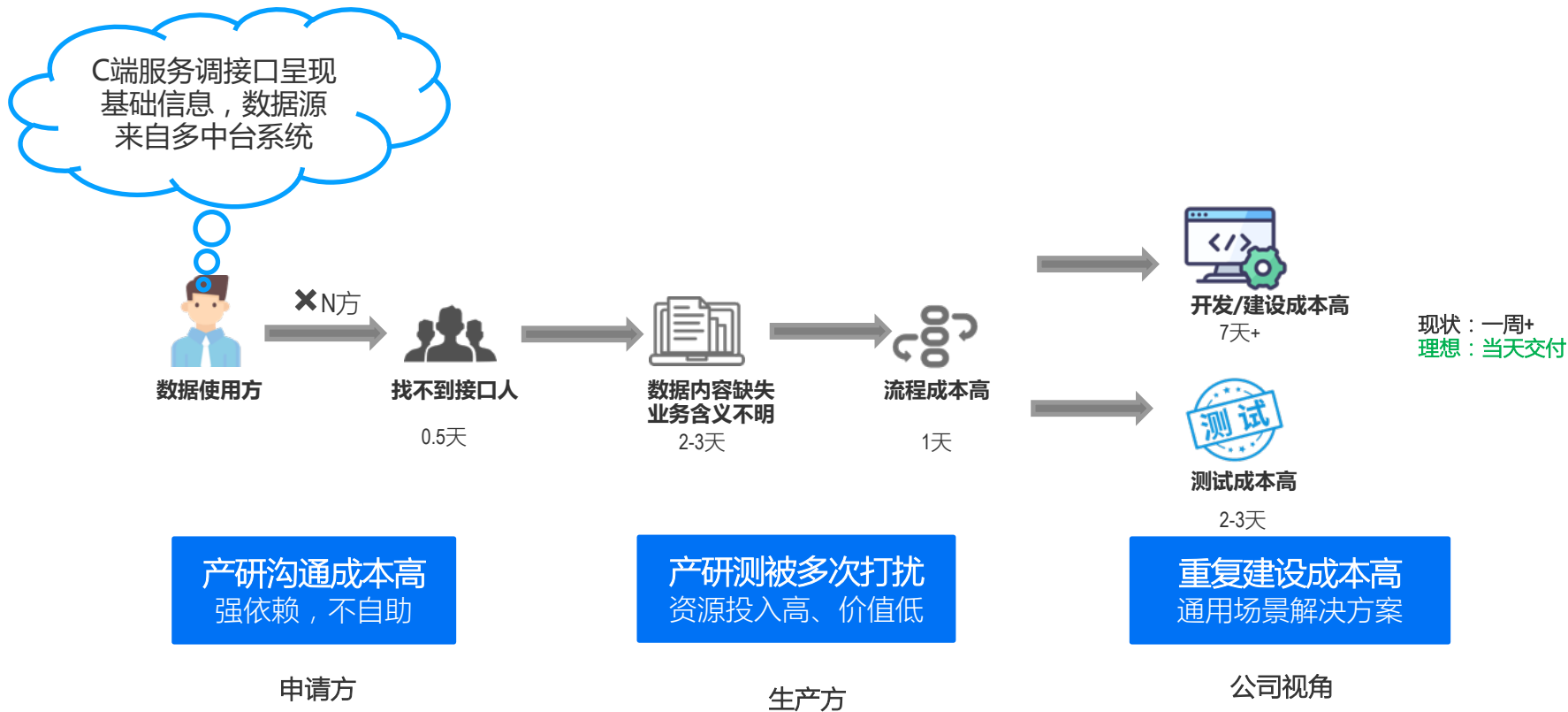
原因：现有数据描述质量低，不足以支持业务自助判断
目标：
• 多维度定义数据
• 需要对数据质量做总控、发布审批

数据智能化助力业务，AI智能、风控品质、精细化运营等等需从各中台一次性获取大批量数据，取数成本成为卡点；

必要性

紧迫性

提效建设必要性



04 中台侧实践的建设内容及思路

建设内容

降低开发学习成本和代码复杂度
进一步省时提效



发布测试



省去日常需求中对测试资源的占用

沟通3天+变为自助查询判断



线上化申请



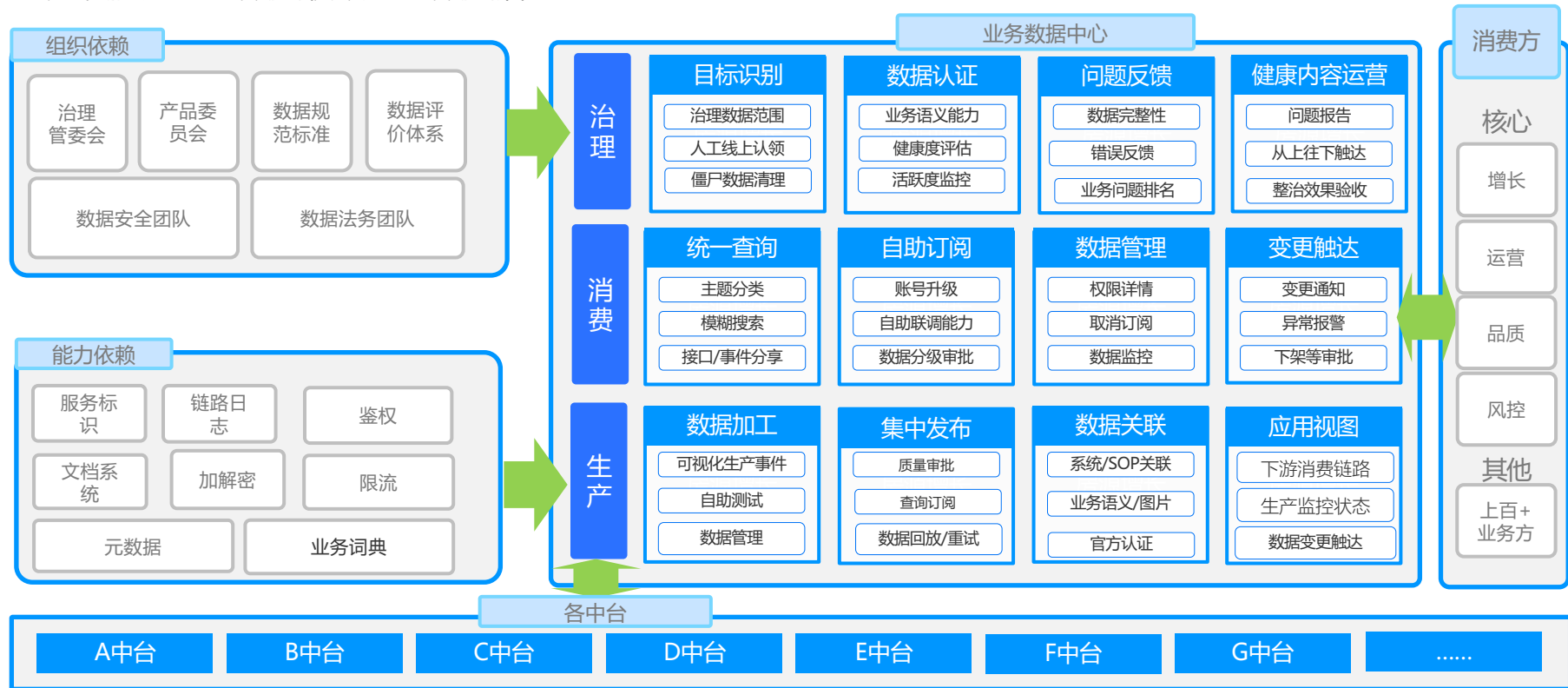
审批时间1-2天变为企微实时响应

产品框架



建设思路

建设能力——数据收敛——数据治理



05 治理项目的目标管理

目标设定

提升业务数据流通的自助化水平

能找到



业务数据需求数量满足XX%

看明白



核心维护信息完备率XX%
平台认证数据占比XX%

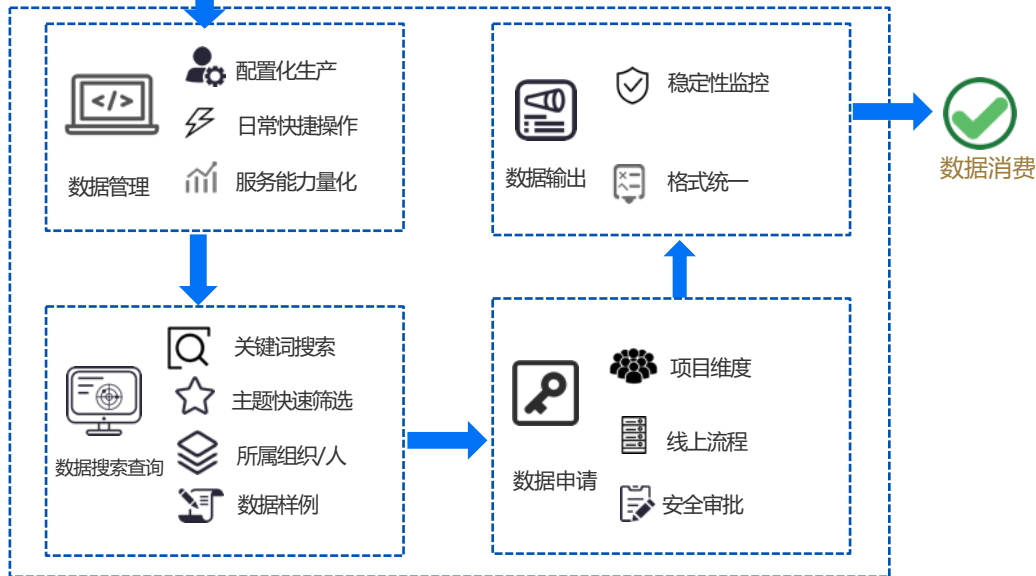
放心用



对下游影响案例量下降XX%



业务数据源



06 产品及运营落地经验

产品设计-躬身入局

- 治理团队要长在业务土壤中，从真正的需求场景中去挖掘能力建设；



- 关键点：**承接主流中台的数据需求
- 关键点：**需求全流程产品、研发、测试工作深度体验

数据收敛经验



- **难题：**在试点用户的基础上，如何接入更多的中台，让数据中心价值最大化？
- **解法：**不知道有业务数据中心存在？——触达到目标用户【公司全体产研】

受众触达 电子屏、海报机、电视，遍布各工区

找到我们 内网导航、搜索特型、企微开屏

数据收敛经验



- **难题：**主动接入方仍然少
 - 1、广告传达内容有限，缺乏主动继续了解的动力
 - 2、缺乏产品关键能力，解决用户最痛的问题
- **解法：**公司技术侧按架构定向产研深度访谈

初版能力域
一些业务方等接入

关键能力
更多业务方陆续接入

行业竞争力
对标竞品

未来迭代方向
丰富需求池

数据收敛经验

试点用户

目标用户触达

深度需求挖掘

良性牵引

- **难题**：人为定向深度访谈触达不到的中台，这部分用户应该依赖“谁”去找到他们并“约束接入”呢？
- **解法**：平台一定需要“核心价值点”作为【牵引力】去长期吸引用户

容易

存量数据不管

用户仍需对接多方混乱格式数据

存量数据
处理方案选择



难

存量数据由中心做统一转换

构建中心核心价值牵引力

成果收益

50%以上定向宣讲未触达到的中台陆续接入，数量仍在持续增加中...

【牵引力其他可能性探索】：运营、AI部门、规则引擎等等，多数据源且关联核心业务方的。

数据收敛经验

- 与核心的数据获取方建立合作机制/系统能力打通，借需求反驱动数据治理

目标：提升业务获取中台数据的自助化程度

数量全面

业务数据需求数量满足率

业务语义可读性

核心维护信息完备率
官方治理认证数据占比

质量稳定

对下游影响案例量降低



收敛效果如何保障？

确定治理目标

核心项目方绑定
承接核心数据需求



中心化梳理

系统建立强依赖
中心承接所有需求
驱动数据收敛和治理



工具建设

沉淀能力工具化
识别能力确定管理抓手

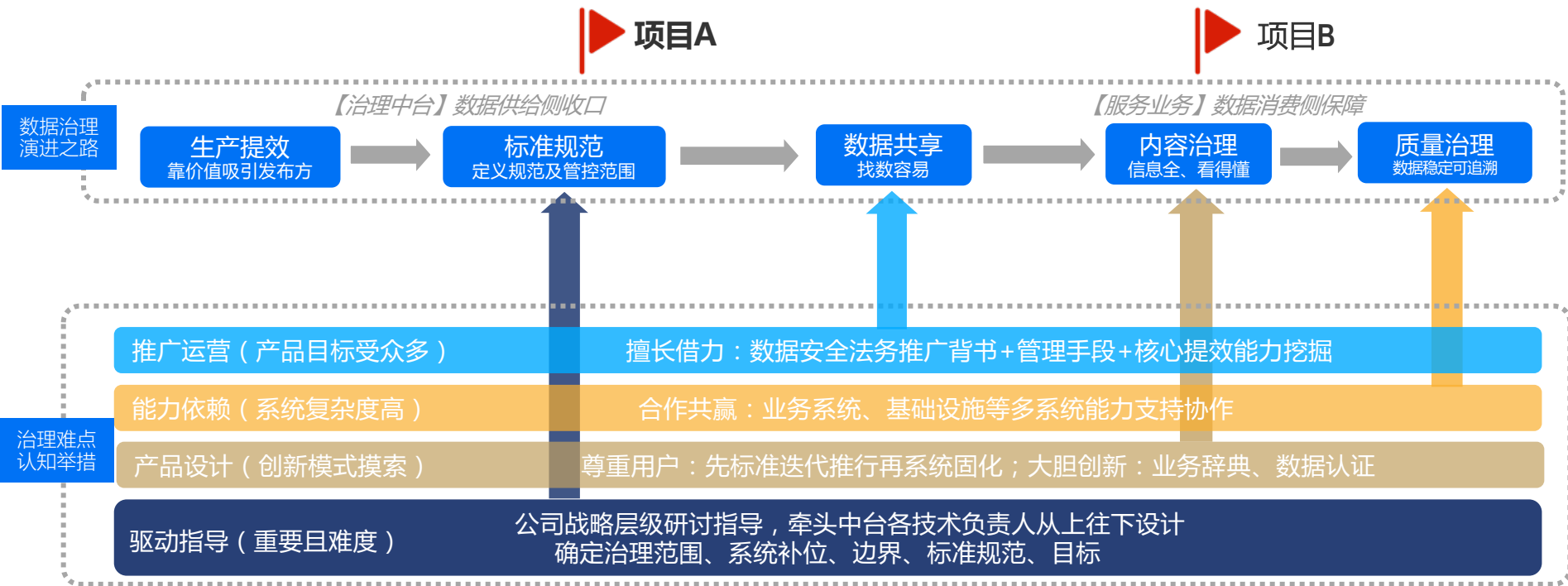


数据运营

管理手段
从上往下推动

治理难点及举措

数据治理演进过程**复杂需长期建设**，每个阶段有承载的责任和使命；**业务卡点问题**决定治理优先级



欢迎加入&学术性探讨



