

标准研究

2024 年第 6 期（总第 38 期）

证标委秘书处

2024 年 9 月 13 日

鹏华基金企业数据标准探索实践之路

【摘要】鹏华基金积极践行国家、行业标准，以数据治理为抓手，通过“标准解读、知识入库、标准制定、落地执行、自动检核、量化评分”的标准落地管理机制，从数据源头治理，强化数据标准管控，逐步消除存量非标数据，形成“系统互联、数据互通、持续改善”的良性系统生态机制，助力公司数字化转型。

关键字：数字化 标准化 标准落地 数据治理 数据要素价值

一、背景

近年来，加快数字化发展，建设数字中国已成为国家重要战略之一，数字化已然成为大势所趋。标准化是数字化的重要一环，只有建立统一的数据标准，才能持续沉淀高质量的数据资源，从而更好地发挥数据要素价值。

2023 年 11 月，证监会与国家标准委联合发布《关于加强证券期货业标准化工作的指导意见》，对证券期货业标准化发展作出了全面部署，推动金融业务规范化发展，加速资本市场标准化进程。

鹏华基金作为资本市场参与者，始终以行业标准践行者的角色积极推动行业标准化工作落地。在公司内以数据治理为抓手，对相关国家及行业标准进行深入解读，结合实际运作情况，解决标准之间冲突，形成企业内部数据标准。以数据标准为指引，构建“系统互联、数据互通”的系统生态，赋能投研、营销、运营等核心业务。

二、实施路径

鹏华基金以数据治理为牵引，全局规划数据蓝图，充分参照国家、行业等权威标准，建立企业级数据标准，对应用系统采用“控制存量、保障增量”的管理原则，避免“边污染边治理”的情况发生，从严、有序地推进数据标准落地。整体实施路径分为三个阶段：

第一阶段：成立数据治理推进组织，制定并发布数据管理制度，以主数据的标准管控为试点，通过与业务专员共同评审定标建立起主数据标准，明确数据归口部门及权责，梳理主数据全生命周期，认证主数据的权威源头，确保主数据按照标准及时、准确地维护落地。

第二阶段：梳理各业务系统中主数据的维护及使用情况，对非主数据系统中承载的主数据维护功能全面改造，统一从主数据系统获取相关信息，打通主数据系统到各业务系统的数据流通通道，避免业务系统各自“挖井取水”。

第三阶段：完成公司全域数据标准覆盖，整合内外部数据要素，统一数据实体 ID，实现数据的互联互通。在此基础上，借助大数据及 AI 能力，赋能核心业务环节，支持公司管理决策，变现数据价值。

三、实施方法

数据标准的实施过程采用“标准制定、标准执行、标准检核”三步走的方式。第一步，建立覆盖所有业务系统的全域数据标准；第二步，推动业务系统建设参照数据标准执行落标；第三步，自动检核业务系统落标情况。

（一）数据标准制定

在制定公司内部数据标准前，对已发布的国家、行业、团体标准能收尽收，逐一解读和梳理，从外部标准文档中识别出与公司业务相关的标准内容，将其纳入企业知识库管

理，构建企业外部标准知识库，并根据外部标准变化持续更新。借助大模型技术，对外部标准知识智能提炼，实现金融数据标准智能问答。

数据标准新增定标时，优先从外部标准知识库中检索相关标准，然后围绕业务、技术、管理三个视角共计 20 余项定标要素展开，以“词根”“词库”“标准”三个层面逐层梳理：通过梳理专业术语、常用词汇形成原子化的基础词根库；在词根基础上根据业务含义生成包含技术属性的词库数据项；结合数据项的共享与交互价值，补充完善业务与管理属性，最终形成完整的数据标准。

基于上述定标流程，由主数据到业务数据，由核心数据到一般数据，逐步建立起企业级的统一数据标准规范，实现投研、营销、运营三大核心业务域数据标准全面覆盖，为业务系统建设提供了完备的参照依据。

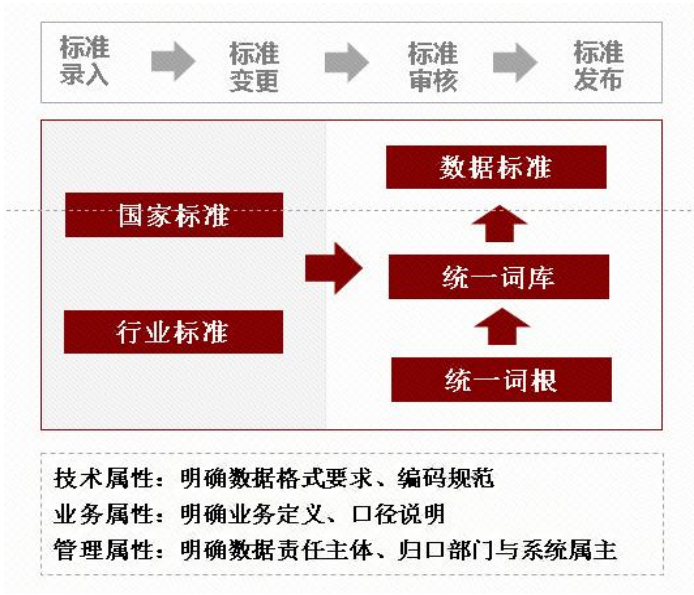


图 1：内部数据标准定标流程

（二）数据标准执行

在标准落地方面，以“系统开发、标准先行”为原则，以自研数字资产管理平台为抓手，将数据标准与开发流程深度融合，把标准执行嵌入到开发过程之中，从源头进行管控。

自研系统：在需求设计阶段，按照数据标准建立数据模型，通过标准化流程，由专人对数据模型进行统一、严谨的评审。评审内容涵盖字段类型、长度、值域等各个维度，从而确保数据标准在系统中得以准确落地，实现有效的事前控制。

外购系统：采用建立映射联接的方法，梳理系统字段与内部数据标准的对应关系，让短期无法满足内部标准的外购系统有据可依；同时，建立外购系统的数据标准评价体系，为系统选型采购提供数据治理评估依据。

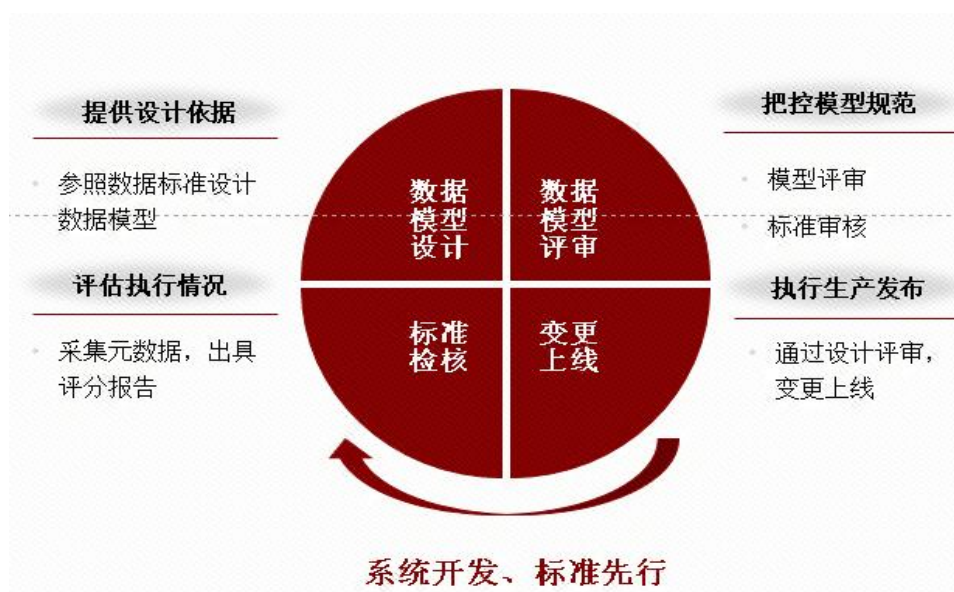


图 2：内部数据标准落标

（三）数据标准检核

数据标准检核主要依托元数据来实现。通过全面采集各业务系统元数据，对比已上线的元数据与既定数据标准间的差异，深入检核业务系统的数据标准化程度。在此基础上，建立完善的质量评估体系，实现事后自动检核，以定期发布数据质量评分报告的形式，确保数据标准的落地执行。

针对已发现问题，建立数据认责机制，将责任落实到岗位角色，做到源头防范、正本清源。根据公司《数据治理管理规定》，由数据认责主管部门对违反数据治理管理规定的组织及个人追究相关责任，数据生产者、加工者、消费者都应当承担其对应责任。

通过自研的数字资产管理平台来管理数据资产与数据责任主体归属关系，把数据问题责任落实到部门，跟踪问题整改情况，为数据质量闭环管理奠定基础。

四、亮点与关键技术

（一）数据标准全面覆盖

构建了企业级全域统一数据标准，规避数据传递与使用分歧。通过将业务对象在企业内部统一定义，消除同名不同义、同义不同名等二义性问题，保证跨系统调用的一致性。为了满足各业务条线的使用需要，在企业内部共计发布了一万六千余项数据标准项，统一数据语言。

（二）数据标准闭环管理

以自研的数字资产管理平台作为数据标准落地抓手，将数据标准管理与元数据管理、数据模型管理、数据质量管理等模块有机结合：数据模型设计所需字段完全引用数据标准，数据模型上线后通过元数据采集获取到有效数据资产，通过数据质量管理，检核数据标准的真实落标情况。对于数据标准自身的规范性与合理性，建立一套完善的标准检核机制，确保数据标准在迭代更新过程中能保持统一、规范。

（三）基于大模型的金融标准智能问答

借助本地私有化部署的大模型实现高质量的金融标准问答，将外部金融标准文档经过适当处理与训练，形成垂直领域的专有模型，对外部标准智能提炼，大幅提升公司内部标准制定效率。通过大模型智能问答，还能够帮助我们清晰识别近似标准在不同场景与范围下的差异，为内部标准制定提供了更加全面、细致的参照依据。

五、实践案例

TA 系统是基金公司的核心系统之一，主要用于基金投资人的份额登记过户。过去，我司采购了多个不同厂商的 TA 系统以满足不同类型产品的份额登记需求。由于不同厂商采用的数据标准各异，在自研 TA 系统进行替换时，为了保障历史数据的无缝衔接，必须对各厂商的数据标准进行兼容，从而导致多个自研 TA 系统之间内部标准不统一，这给系统的持续开发与运维带来了诸多不便。

为从根本上解决该问题，我司以《开放式基金业务数据交换协议》为主要标准依据，制定了统一的 TA 数据标准规范。借信创契机，所有自研 TA 对系统进行了统一数据标准改造，彻底解决不同 TA 数据标准不一致的问题。此举不仅大幅降低了开发成本，也简化了工作的复杂度，更有效降低了因标准切换而引发程序缺陷的风险。同时，运维人员亦能在统一、规范的标准框架内高效开展工作，运维数据采集脚本多 TA 系统通用，显著提升了整体运维效率。

六、问题与不足

标准冲突是我们制定内部数据标准过程中面临的突出问题，在引入外部标准时，跨业务域交叉场景最容易出现“打架”的情况。如何解决冲突问题是一大难题，除了依据不同层级标准的权威性来判断处理以外，更多的是要考虑具有兼容性的平衡方案，以同时满足不同业务域数据使用需要为前提，通过共同评审定标来解决冲突。

为了使数据标准在公司内部具有持续的生命力，我们会定期参照现行外部标准开展滞后、老化内部标准的评审和修订，解决标准缺失、矛盾交叉等问题，确保标准的先进性和权威性。

通过制定统一的数据标准，不仅可以提高数据质量，更能有效降低企业间的沟通成本，进一步促进数据的流通与价

值释放。未来随着更多公司加入数据标准化的行列，我们将迎来一个数据更加规范、应用更加广泛的数字化新世界。

（鹏华基金管理有限公司金融科技部总经理助理王良楷、
数据总监助理薛载广、数据工程师周子鹤供稿）