

【数据治理实践】第十三期：数据“智”理的工具——数据管理平台

前言

[上期文章](#)，我们分享了银行数据架构设计与规划的相关实践，本期我们将探讨数据管理平台。很多银行已经开展了数据管理平台相关的建设，帮助银行实现数据管理从手工管理到自动化甚至是智能化的升级。下面我们将从监管要求、数据管理平台的功能、数据管理平台的智能发展来进行分析介绍。

监管要求

在《银行业金融机构数据治理指引》（下称“指引”）有以下的一些要求：

第五条 银行业金融机构数据治理应当遵循以下基本原则：

（一）全覆盖原则。数据治理应当覆盖数据的全生命周期，覆盖业务经营、风险管理和内部控制流程中的全部数据，覆盖内部数据和外部数据，覆盖监管数据，覆盖所有分支机构和附属机构。

第二十条 银行业金融机构应当建立覆盖全部数据的标准化规划，遵循统一的业务规范和技术标准。数据标准应当符合国家标准化政策及监管规定，并确保被有效执行。

第二十一条 银行业金融机构应当持续完善信息系统，覆盖各项业务和管理数据。信息系统应当有完备的数据字典和维护流程，并具有可拓展性。

第三十二条 银行业金融机构应当建立数据质量监控体系，覆盖数据全生命周期，对数据质量持续监测、分析、反馈和纠正。

第三十五条 银行业金融机构应当建立数据质量整改机制，对日常监控、检查和考核评价过程中发现的问题，及时组织整改，并对整改情况跟踪评价，确保整改落实到位。

——《银行业金融机构数据治理指引》

监管机构要求银行的数据治理工作覆盖全部数据，如果单纯通过人工方式去覆盖数据的**全生命周期、覆盖流程中的全部数据、覆盖内外部数据和所有分支附属机构**，必然是一项浩大的工程。通过我们在第九期文章中提到的数据资产目录，可以将银行的所有数据资产进行编目，逐步形成行内覆盖全生命周期的，覆盖银行内外部数据的数据资产目录，有效支持银行开展各项数据管理工作。

而对于数据标准的有效执行，**不仅需要有良好的数据标准管理规划，嵌入系统开发的流程，还需要有一定的事后检查和评估方法**，例如可以利用数据管理平台，对数据字典的定义与数据标准进行比较；也可以通过数据标准的代码等定义与现有的系统进行比对，评估数据标准的执行情况。

此外对于监管要求的数据字典管理，我们在之前文章里也提及到**元数据管理工具是数据管理成功有效开展的重要条件**。数据在各个重要系统中存储、流转、加工、应用，由于高度复杂的银行业务以及庞杂的信息系统，如果纯粹依靠人工的方式对各个库表结构，ETL关系等元数据进行采集建设，那么将投入极大量的人力和物力进行维护和管理，因此一个便捷的元数据管理工具是元数据成功有效管理的重要条件。

数据质量的监控是数据管理平台重点应用的场景之一，包括对银行各系统中的数据质量问题进行监测、分析、反馈和纠正。对日常监控、检查中的问题通过系统化的方式进行跟踪管理。通过数据管理平台将数据生命周期的各个环节的质量处理情况进行记录跟踪，可以作为数据质量的考核依据支持银行开展专项管理功能。

数据管理平台的功能

一般来讲数据管理平台关键基础功能包括以下几个方面：



1 各项管理流程的支持：实现数据管理各领域中的流程常态化运行。

数据管理平台最基础的功能，是支持各项数据管理流程线上化运行，包括数据标准、数据质量、元数据、数据需求、数据治理考核、数据治理自评估等流程，将各项维护和管理工作嵌入流程，实现数据治理的常态化。

2 数据标准管理：实现对全行级数据标准的维护和落标分析。

数据标准的落地情况分析，依赖于数据标准与数据字典的映射关系。但由于生产环境中的系统数据字典更新频繁，映射关系如果仅仅依靠人工维护，费时费力。数据管理平台通常可以提供的数据标准管理功能包括：

- 数据标准维护：支持在平台上新增、修改、废止数据标准，包括批量维护。
- 数据标准版本管理：支持数据标准的历史版本查询和版本比对功能。
- 数据标准落标比对：基于系统数据字典与标准项之间的映射关系，对两者属性的自动校验，形成详细校验结果。

3 数据质量管理：实现数据质量检核定义、质量检核分析、质量检核执行，形成数据质量知识库。

由于银行业务高度复杂，一些重要数据通常在行内多个系统中产生、流转、保存，部分数据可能出现重叠或不一致的情况。为对数据进行及时有效地监测并发现数据质量问题，并予以解决，就需要数据质量管理工具作为辅助。数据管理平台通常可以提供的数据质量管理功能

包括：

- 数据质量规则知识库：通过系统中的模板自定义配置检核规则；通过系统中已建立的数据字典和数据标准的映射关系，自动根据数据标准的技术属性，生成检核规则；能自定义编写检核规则与检核脚本。
- 数据质量检核：能手工执行和自动执行检核规则调度，支持事件触发策略；可以自定义检核范围与检核周期，生成数据质量问题清单。
- 数据质量问题分发：在流程上支持数据质量问题多层级跨机构分发到指定人员；问题清单的接收机构可对全部或部分问题做反馈处理；支持问题处理机构对明细问题填写处理情况备注，并上传辅助证明材料。
- 问题数据下载：能按照制定规则导出所有的问题数据清单。
- 分析和跟踪：支持质量问题跟踪管理，可进行提交、处理和审批，能够与银行OA流程集成；问题严重程度能自动升级告警；支持问题知识库的管理，能够将数据质量核查问题归集到特定问题集，与其特定解决方案一起形成知识信息。

4 元数据管理：实现对元数据采集、维护、版本管理，展现血缘分析、影响分析与数据地图。

为衔接数据质量与数据标准等管理领域、建立企业级数据加工链路、支撑元数据的血缘分析和影响性分析、展示全行数据地图，元数据管理工具是关键和基础的一环。数据管理平台通常可以提供的元数据管理功能包括：

- 元数据采集：能与行内调度工具集成，批量自动定期或一次性采集信息系统元数据（包括数据字典）；与银行信息科技管理系统对接，当发生系统上线或变更时，自动发起元数据采集任务。
- 元数据维护和版本管理：支持元数据的增、删、改，修改和补充自动采集和解析的元数据，并支持元数据使用权限控制。
- 数据血脉分析：分析并展现字段级元数据的来源及转换关系。以当前元数据对象为中心，分析并展现后续数据加工关系中的元数据对象及之间的加工逻辑。

5 数据需求管理：提供面向业务用户的数据需求沟通渠道，跟踪数据需求、统计数据需求。

支持发起数据需求申请、审批和退回流程，数据需求可以包括取数需求、数据分析应用需求等，通过此平台进行登记和统计。

数据管理平台的“智”能发展

数据管理平台的智能化要点主要可以从以下的要点内容里面进行开展：

1 数据管理与数据应用打通

数据管理最终要为数据应用服务，数据标准的统一，数据安全的有效管理，数据质量的提升最终都要在数据应用上体现。银行业务部门困惑的数据在哪里，数据质量如何，如何获取自己想要的数据（标准的、统一的、高质量的数据），希望可以一站式解决。

2 数据字典采集自动化

数据字典变动的影响数据范围、影响部门范围分析，需要结合行内现有流程与自动化的方式进行管理。首先是自动化的采集与比对，每一次的系统变更，需要有更加智能的控制和管理。

基于OCR/RPA/NLP/机器学习技术，快速建立基准和后续动态更新。由机器自动匹配高度置信的数据项，由人工将其他模糊的匹配项进行优化调整，并纳入后期机器学习的范畴，减少人工投入，不断提高数据字典对标的准确性，保持数据动态更新。

3 数据标签智能化

对于业务部门来说，数据的标签化是最简单最直观的使用方式。例如要对某个城市的高净值人群开展营销活动，可以通过数据标签，筛选年收入一定数量以上，有出国记录等数据标签的用户进行快速的筛选出目标客户群。而另外的一种标签是针对数据库表字段进行的标签：

单一视角管理

通过金融数据模型，例如 FSLDM对数据资产进行整理分类和管理。

多视角、扁平化管理

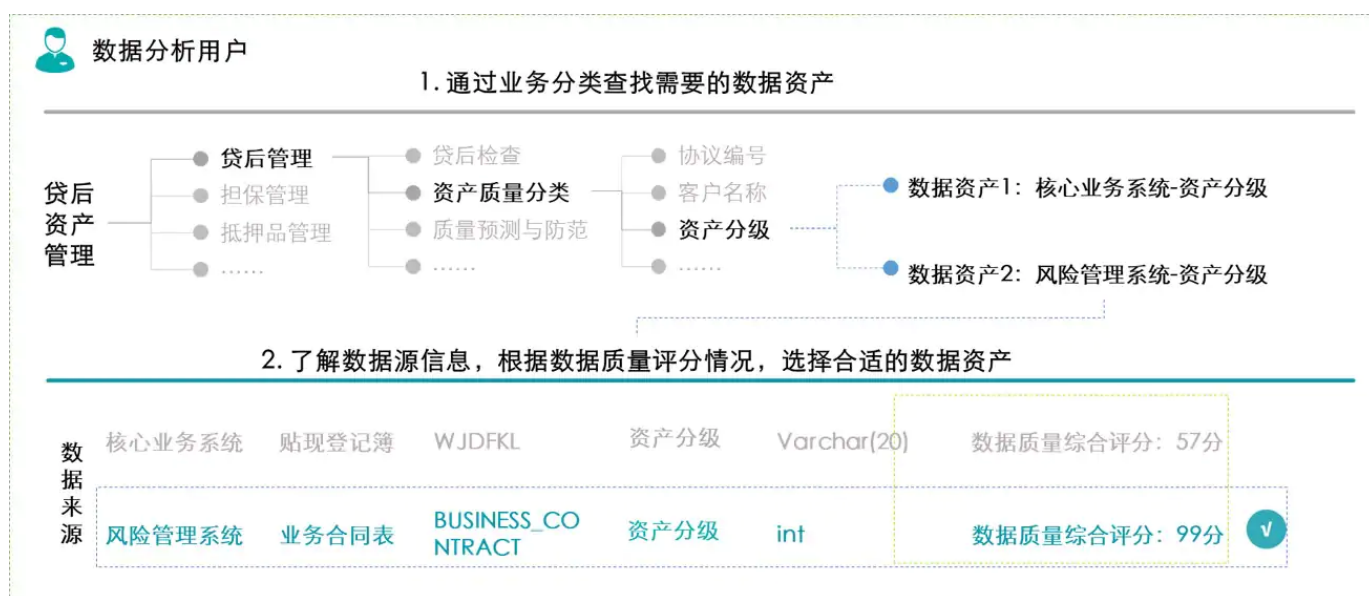
- 业务标签
- 监管标签
- 安全标签
- 系统标签
- 管理标签
- ...



- 客户标签
- 热度标签
- 自定义标签
- 扩展标签
- ...

可以基于数据字典名称进行自动化地定义设计标签，可以基于样例数据进行自动化定义标签，例如日期数据，可以基于历史的样本数据定义设计标签等，快速自动地丰富银行的数据字典标签，支持银行数据应用。

- 数据字典名字规则的
 - 如包含“姓名”
 - 如包含“账号”
 - 如包含“币种”
- 样例数据符合正则表达式规则的
 - 如样例数据满足YYYYMMDD
 - 如样例数据满足身份证编码
 - 如样例数据满足IP地址编码
- 基于人工标记的样本数据



结语

数据管理平台的自动化是数据管理的必要辅助，**单靠手工的数据管理不仅无法对全量的数据进行管理，也不能及时的体现管理的效果。**此外，数据管理平台的智能化可以大大地解放业务人员与管理人员，银行应积极地探索尝试，通过智能化的管理提高银行的数据管理水平，提升数据质量，发挥数据价值。后续我们将就数据管理的评价、审计等模块内容持续展开，敬请期待。