

基于数据模型的证券期货行业 数据治理

Securities and Futures Data Governance Based on Data Models

■ 中国证券监督管理委员会信息中心 张 野

摘 要 针对目前国际数据治理体系中模型编制方法论和成果理论两个环节缺乏具体方法论的问题，提出证券期货行业的解决方案。描述了证券期货行业数据模型，以及从行业现状出发逐步建设抽象模型和逻辑模型的具体方法，并形成了分类分层描述数据模型成果的架构。提出了对行业数据模型成果的应用及其实践情况。

关键词 证券期货行业 数据治理 数据模型 标准化

Abstract: Pointing to the problems that two parts in the international data governance system - model generating methodology and resulting theories lack specific methodologies, further understanding and solutions on these issues particularly in the securities and futures industry are put forward. A specific method based on the industry data model's current situation, and how the abstract model and logic model are gradually built are described. And the categorizing and layering structure is then formed, so as the data model achievements structure. Subsequently, application scenarios of data models are provided with their practical situations.

Keywords: securities and futures industry; data governance; data model; standardization

1 数据治理的现状行业思考

大数据技术的广泛应用使越来越多的行业和机构认识到，数据已经成为最重要的企业资产甚至是行业资产，对于数据资产的了解、规划、治理和预测成为各行业的新兴赢利点。证券期货行业数据治理是对全行业数据资产行使权力和控制的活动集合，其包括但不限于规划、监控和执行。数据治理能指导数据管理工作的执行。数据治理方法论体系涵盖了概念、框架、应用、目标要求以及成果理论。

在目前的国际数据治理体系中，各国际标准组织及数据管理组织对数据治理的概念、框架、目标要求都已有明确定义，但是在应用与成果理论两个环节，还没有深入和具体的方法论，国际数据治理方法论体系见图 1。

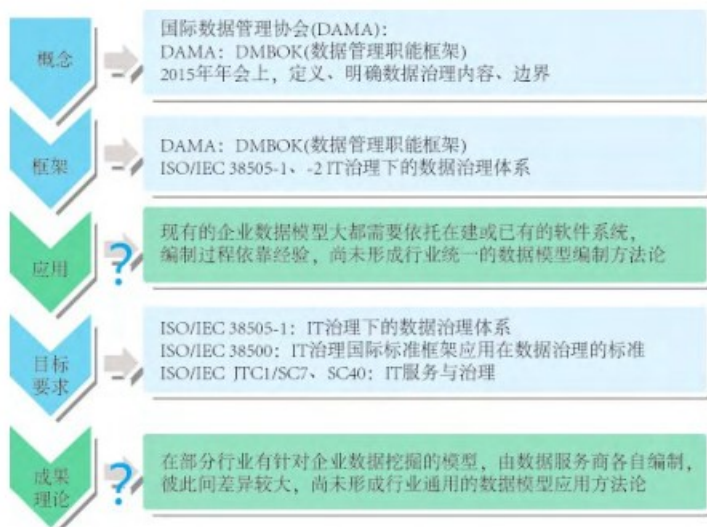


图 1 国际数据治理方法论体系

基于数据模型的证券期货行业数据治理方法论就是围绕上述欠缺的两点开展研究。

现有的证券期货行业机构多、类型广，交易方式多样，变化迅速。市场上机构间数据交互相对复杂；

机构内部应用系统多，数据交换接口多样，通用程度差。证券期货行业的数据化程度相对较高，能够准确、高效地掌握行业数据的情况，不论对监管者还是市场参与者，都是至关重要的。因此，急需一套数据模型，用于描述整个证券期货市场业务、数据情况，明确数据定义、规范数据交换、指导行业系统建设。为了解决上述问题，证券期货行业应建立行业数据治理体系，以行业数据模型为核心，包含行业数据模型建设和数据治理两大块工作。行业数据治理可以规范数据定义，服务行业标准化，其架构见图 2。

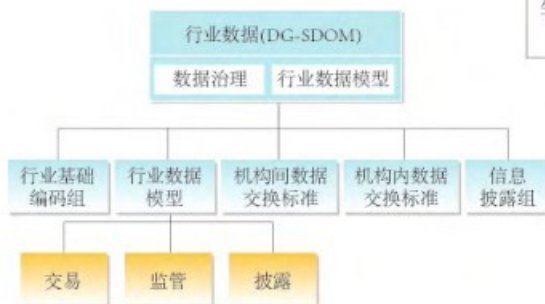


图 2 行业数据治理架构

注：“披露”指行业数据模型中的披露条线及其模型，信息披露组是规范、指导行业现行信息披露工作的专业工作组，在其工作过程中将应用“披露”条线的行业数据模型。

目前，在行业数据生成、交换与应用整个生命周期中，均面临着一些数据应用的相关问题，数据模型作为数据治理体系的核心，针对以上问题，提供相应的解决思路，具体见图 3。

2 数据治理核心——行业数据模型

证券期货行业数据模型是以证券期货行业相关

法律法规、业务规则、制度及流程等为依据，绘制行业顶层数据流图及机构内数据流图，识别行业数据的现状。进一步以“交易”、“监管”、“披露”三大业务线条为切入点，分别完成对行业中各种业务的全面遍历，最终形成一系列有关联关系的数据项和数据表。

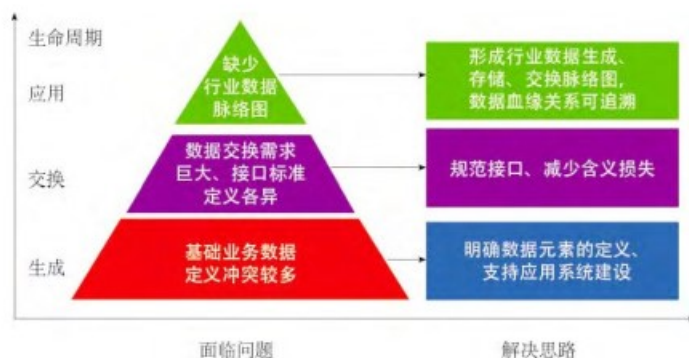


图 3 数据治理面临的问题及解决思路

行业数据模型从识别行业现状出发，根据模型应用深度的不同，分两步形成用于规范行业标准化的抽象模型和指导行业数据应用建设的逻辑模型，模型编制步骤见图 4。

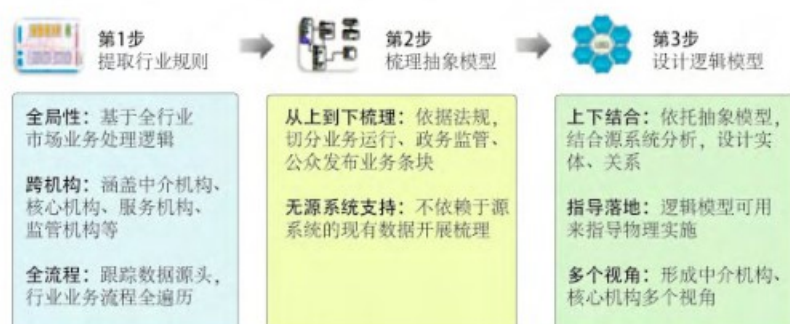


图 4 行业数据模型编制步骤

其中梳理抽象模型和逻辑模型分别形成了一套方法论。

(1) 提取行业规则

在提取行业规则时，综合考虑法律法规、业务规则和数据特征。其中，涵盖证券业务法律法规近 50 部；3 条业务主线，20 余类主体，10 余个品种，

共计业务规则 400 余项；同时，根据机构内、机构间数据交换现状，形成数据流图，提炼数据特征。

(2) 梳理抽象模型

在抽象模型梳理过程中，依据“IBR”（Identity、Behavior、Relevance）方法，以“内部 + 对外 + 监管”三个角度划分三大业务条线。在证券期货行业，这三个条线按照行业特征对应为内部业务运行类即“交易”模型、对外信息发布类即“披露”模型和监管模型。三大条线模型分别采用各自的方法进行梳理。

其中，交易模型使用“SPB”（业务（Service）、行为（Behaviour）、过程（Process））方法，即针对市场中的每个品种，识别其交易行为和过程的相互关系来抽取数据模型；监管模型使用“TM”（监管主题（Theme，譬如获取证券期货业资格）、监管方式（Method，譬如行政许可））方法，即通过监管主体和法律法规，识别监管对象和方式形成模型；披露模型使用“T4R”（模板（Template）、映射（Reflect）、提

炼（Refine）、重组（Recombine）、回归（Regress））方法，即通过披露模板逐层反向提取数据表和数据项形成模型。最终，综合三大业务条线，抽取、提炼全市场业务流程与数据的共性部分，形成统一、完备的具有通用性、稳定性和扩展性的资本市场多层次数据模型。

然后，针对梳理结果，应用元数据管理思想，将数据模型按照其自身的属性分为原子数据、复合数据，可复用数据表、语义独立数据表四个层次。同时根据应用过程中的通用性差异，形成通用基础、业务条线通用、业务个性化三个层级的“1+3+N”式模型层级，其中“1”为全市场通用的通用基础数据模型，“3”为交易、监管、披露的业务条线通用模型，“N”为各业务应用的个性化模型。上述分类、分层的组织架构，保障了数据模型的完整性，兼顾了数据模型的共性和个性。抽象模型分层描绘见图 5。

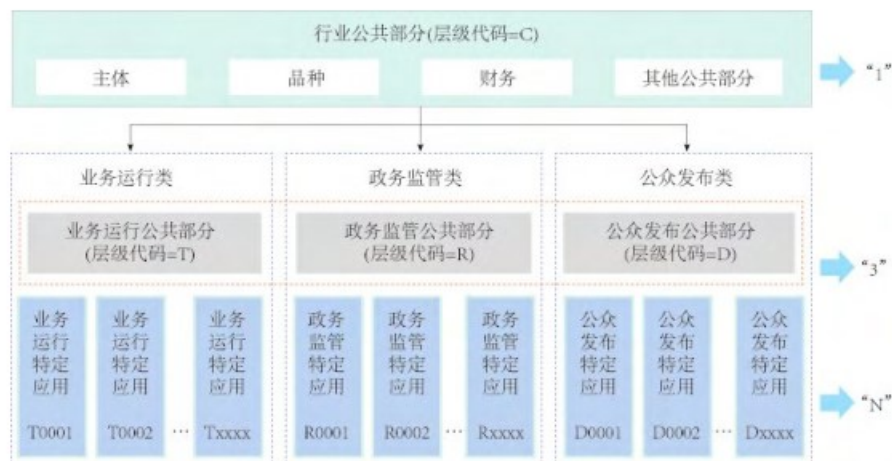


图 5 抽象模型分层描绘

(3) 设计逻辑模型

为了方便用户读懂并使用数据模型，按照行业属性代码、证券交易所、期货交易所、证券公司、期货公司、基金公司、监管机构等视角，以“1+6”的方式，依托抽象模型，设计一系列实用性比较强的表形成逻辑模型，逻辑模型架构见图 6。

逻辑模型设计时，首先依托抽象模型成果，归纳数据共性，合并、提炼划分逻辑模型主题域；其次，通过“IBR”方法，找出主题域中核心数据的特征和关系，构建主题域之间的核心关系；最后根据主题域和核心关系形成从核心到外延的逻辑模型架构。证券公司逻辑模型样例见图 7。



图6 逻辑模型架构

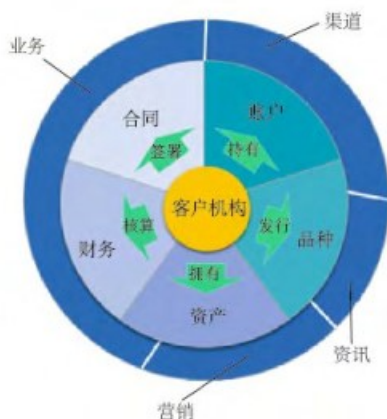


图7 证券公司逻辑模型样例

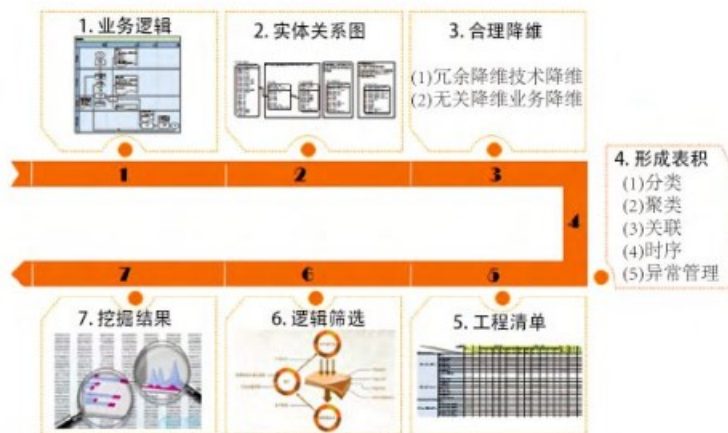


图8 大数据智能挖掘思路

3 数据模型应用

证券期货行业数据模型通用性强、覆盖面广、实用度高，是行业数据治理工作的核心。现已用于统一证券期货行业数据标准、规范行业各类标准制定、指导行业各机构应用系统建设，乃至为行业数据大集中项目提供权威数据标准。

数据模型应用领域主要体现在以下三个方面：

(1) 规范行业标准化

建立一套基于行业数据模型的行业标准编制、审核机制、关联查询方式，通过行业标准审核，规范、统一在行业中交换数据的含义和口径。

(2) 支持核心应用系统建设

运用不同视角的逻辑模型，直接指导行业系统

建设，特别是数据仓库类的数据库应用系统建设，规范机构内部甚至机构间关联的数据含义，减少数据损失。

(3) 指导智能数据挖掘方法论

建立智能数据挖掘方法论。通过数据模型中的业务流程和实体关系图，形成数据挖掘地图，开展不依赖于业务需要的工程化数据挖掘，遍历特定业务方向上的数据特征，全面提取业务间的关系列表，提高数据挖掘准确性和挖掘效率。大数据智能挖掘思路见图8。

4 行业数据治理规划与展望

为了更高效地管理行业数据资产，以行业数据模型为核心的行业数据治理工作将作为一项长期工作，从模型设计、平台建设、模型推广和应用等方面不断推进。行业数据模型将与市场创新同步发展，形成业务全覆盖的、具有世界先进水平的数据模型，并针对核心业务深度扩展，构建逐层深化的行业模型体系。今后，应大力开展模型推广和应用，推动以数据模型为原点的行业数据治理工作。

参考文献

- [1] 李鸣, 张旻旻, 蔡震宇. IT 治理标准研究 [J]. 信息技术与标准化, 2016(03): 29-33.

(收稿日期：2016-06-30)