

证券期货行业数据模型研究

更新时间：2015-09-06 14:15:39 点击量： 2703次

中国证券期货行业正处于新兴转轨时间，处于高速发展阶段，行业对标准要求多、要求高。证券期货行业标准建设起步较晚，由于缺少行业标准，各市场参与主体自行规划，基于同样业务的操作、监管、分析系统也各成体系。为确保行业内机构信息数据的一致性，证券分委会于2013年12月正式启动了证券期货行业数据模型项目，对交易、监管、披露过程进行了全面业务梳理，对业务进行分类和程序化，从整体的数据视角描述各部门、各系统、各环节之间的内在联系。

一、数据模型定义和目标

证券期货行业数据模型（以下简称“SDOM”）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，提取市场全业务流程与数据共性，形成具有通用性、稳定性和扩展性的数据模型。

行业数据模型是以“交易”、“监管”、“披露”三大业务线条为切入点，完成对行业各业务的全面遍历。通过对市场中的证券、基金、期货、债券、回购等主要交易行为的梳理形成行业交易模型；通过对已上线及待上线的信息披露项目的逆向梳理形成行业信息披露模型；通过配合中央监管信息平台建设完成监管模型，最终形成真正意义上统一、完备的资本市场多层次数据模型。

二、数据模型“1+3+N”设计架构

数据模型的设计遵循“1+3+N”的设计架构，其中：

“1”是指数据模型以品种、主体、财务三大主题形成公共部分。这些模型可以被“交易”、“监管”、“披露”中至少两条业务线条复用，“1”中的模型需要保持语法、语义一致。

“3”是指以“交易”、“监管”、“披露”三大业务线条分别梳理形成三大组数据模型，每组数据模型根据某一业务线条的特点提取出来，三大组模型彼此耦合度不高，组间个性化部分允许出现同名不同义的冲突情况，不要求在语法语义上保持一致。对于每组数据模型再分为该业务范畴内的公共部分和覆盖其下应用系统的全部数据和表。其中公共部分可在该业务范畴内共用，需要在组内保持语义、语法一致。

“N”是指在每个业务线条的大组数据模型中，覆盖该应用全部数据、表，用于指导该线条下特定应用系统的专用数据模型，该部分的数据和表为某一应用特有的，只在该应用中被使用，与业务线条中其他应用的模型不复用。每个“N”中涉及的数据、表在语法、语义上不要求一致。

“1+3+N”的设计模式表达如下：

其中： $P1 \in \{\text{主体, 品种, 财务, 其他公共部分}\}$

$$P2_i = \text{SDOM} = P1 + \sum_{i=1}^3 \left(P2_i + \sum_{j=1}^n P3_{ij} \right)$$

$P3_{ij}$ 代表每个业务线条下，特定应用系统的专用数据模型。

$$\begin{cases} \text{交易公共部分, } i=1 \\ \text{监管公共部分, } i=2 \\ \text{披露公共部分, } i=3 \end{cases}$$

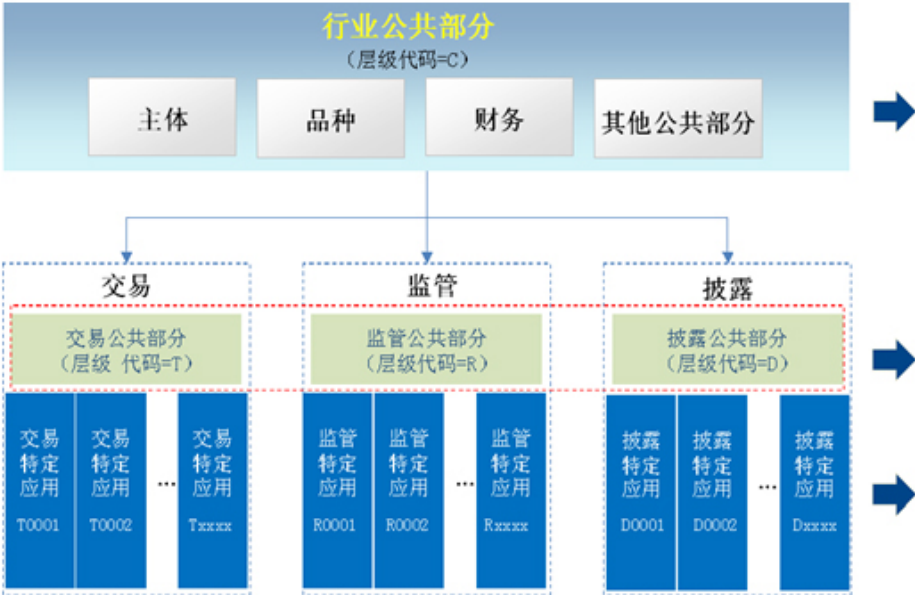


图5-1 “1+3+N”设计框架

三、各业务线条模型梳理方法
(一) 交易类模型梳理方式（SPB方法）

概述：交易类数据模型主要是采用SPB方法进行梳理，首先，梳理国内证券品种，整理证券品种分类数据。其次，遍历各证券品种业务行为，分析所有交易环节，提取共性交易过程，整理业务行为和业务过程，形成各业务品种的行为过程二维图。针对行为过程二维图中行为，结合该行为的业务流程，整理业务流程的各环节及各环节的产出，形成业务流程的甬道图；细化业务流程图中各环节产出物涉及的数据项，明确数据项定义及属性信息。数据项按主题归并提炼形成数据对象，并将数据对象及其之间的关系以ER图形式表达，形成理论数据模型。最终，通过现实系统的数据模型对理论模型进行回归验证及完善工作。

如图5-2所示：



图5-2 SPB方法

定义：交易类模型梳理从品种出发，以CFI为依据，结合国内资本市场现状梳理出能覆盖国内现有证券品种二叉树。从品种二叉树入手，对各品种分别按照行为（B: Behaviour）、过程（P: Process）进行遍历分析，从中提取出原子数据、复合数据、可复用数据表、语义独立数据表这四层模型信息及其元语定义。这种模型梳理方法即为SPB（Securities Process Behaviour）方法。

品种梳理：品种梳理工作参考CFI编码，结合国内现有证券品种，构建符合国际规范、适应国内资本市场现状的品种分类及定义。其中，品种分类的一、二级沿用CFI标准分类；三级参考CFI品种属性分类；四、五级是结合国内品种现状的自定义分类。

行为及过程梳理：本环节针对每个品种所涉及到的交易行为进行梳理，行为分解过程中，保证每个行为的独立性，行为间不存在包含关系，对每个行为进行过程拆解，最终形成针对某一特定行为及过程拆解的二维矩阵图。然后参照相关业务流程及规则定义，梳理出具体的业务流程图，便于后续对行为过程进

行详细分析以及梳理提炼数据项。

数据项梳理：参照业务流程图，对每个业务环节以“5W1H”的方法梳理和提炼数据项，并根据数据项的业务特征进行分类。

（二）监管类模型梳理方式（T/M方法）

原则：以证券期货行业相关法律法规为基石，遵循有法可依的原则，建立“证券期货行业数据模型（SDOM）-监管类数据模型”。

定义：以证券期货行业相关法律法规为依据，对监管对象（譬如会计师事务所等）、监管主题（Theme 譬如获取证券期货行业资格）、监管方式（Method 譬如行政许可）进行三维分析，梳理出对应监管流程及相关信息，从而得出“证券期货行业数据模型（SDOM）-监管类数据模型”的方法。

T/M方法的原理图如图5-3：

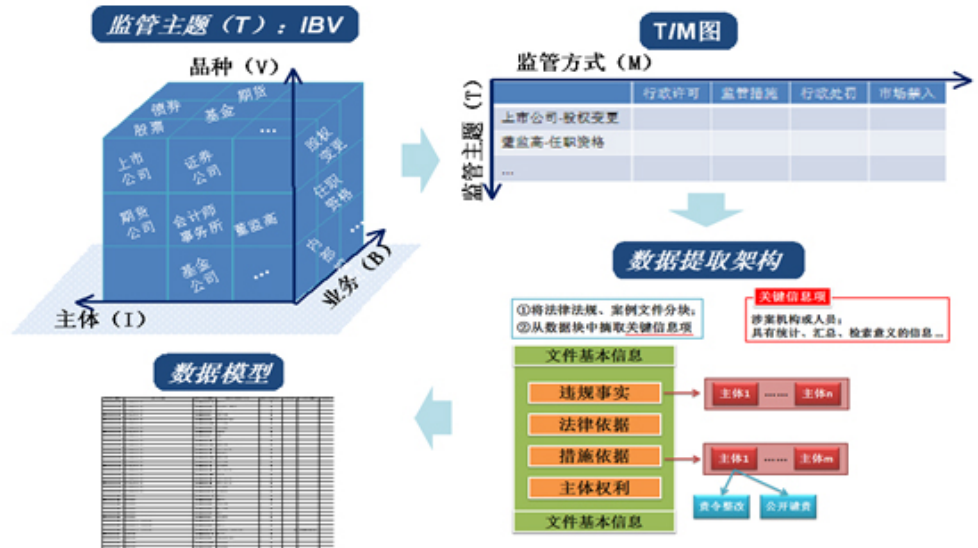


图5-3 T/M方法

步骤：T/M梳理方法用五个步骤梳理出“证券期货行业数据模型（SDOM）-监管类数据模型”：

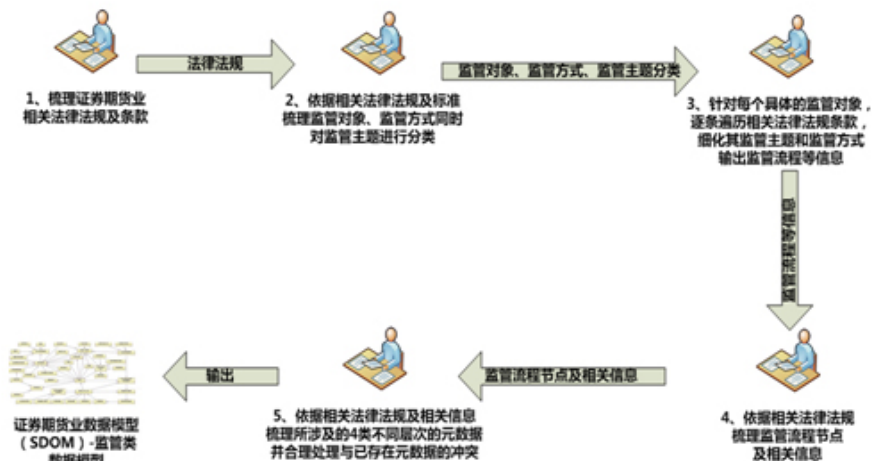


图5-4 T/M方法步骤

步骤1：梳理证券期货行业相关法律法规及条款。

步骤2：依据相关法律法规及标准梳理监管对象、监管方式同时对监管主题进行分类。

步骤3：针对每个具体监管对象，逐条遍历相关法律法规条款，细化其监管主题和监管方式，输出监管流程等信息，填写T/M表。

步骤4：依据相关法律法规梳理监管流程节点及相关信息，并填写监管流程表，同时使用流程图描述监管过程及节点。

步骤5：分析元数据，针对每个监管过程节点，依据相关法律法规及相关信息梳理所涉及的4类不同层次元数据并合理处理与已存在数据的冲突，如果需要则使用ER图描述元数据之间的关系，最终分析出“证券期货行业数据模型（SDOM）-监管类数据模型”。

（三）信息披露类模型梳理流程（T4R方法）

概述：信息披露类模型采用基于信息披露模板的模型梳理方法，包括模板（Template）、映射（Reflect）、提炼（Refine）、重组（Recombine）、回归（Regress）等五个重要环节。其中模板是信息披

露的基础，映射基于模板制定出语义独立的数据表，进一步提炼出原子、复合及可复用数据表，重组后进行逻辑推演和测试验证，经多次迭代后形成最终的披露信息数据模型。各环节的关系如图5-5所示：



图5-5 披露数据模型各环节关系图

模板：模板是信息披露内容的具体体现，构建模板，首先以统一的主体和品种分类为方向，在交易类及监管类模型梳理构建过程中，结合国内资本市场现状，已梳理并提取出交易主体和品种的分类，在此基础上，结合主体和品种运行周期，分别整理归纳和各自信息披露有关的法律法规，将梳理出的信息披露业务规则和要求进行规范后形成的文档即为信息披露模板，如图5-6所示：

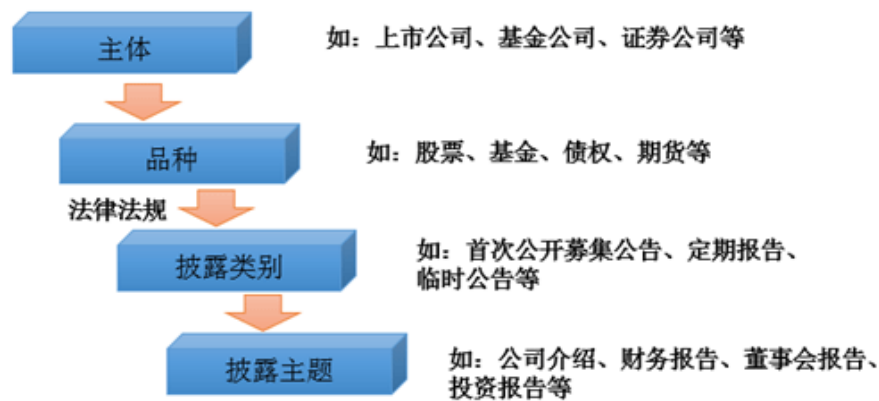


图5-6 信息披露模板图

映射：选定信息披露模板后，对模板结构进行梳理，定义出语义独立数据表，以及数据表间多层次构成关系。语义独立数据表可面向应用进行建模指导，由其他语义独立的数据表、可复用的数据表、复合数据、原子数据共同组成。

提炼：原子数据梳理工作是在每个模板内，根据披露信息的业务含义，指出不需要再分的最小数据单元。复合数据的梳理工作方法和原则与原子数据梳理类似，区别在于识别该数据是否可以通过计算、构成以及其他因素而成为一个复合数据。可复用数据表梳理工作是在相同主体或相同类别的模板范围内，对每个模板内容进行分析，对相关的语义上独立的数据表进行抽象与归纳，提炼出逻辑上可复用的数据表。

重组：从模板映射出语义独立的数据表，提炼出原子数据、复合数据、可复用数据表后，不同披露主题、产品的披露模板中仍然存在一定的共性内容，需要对比其共性内容进一步识别和提炼可复用的数据表，识别业务含义相同的原子数据和复合数据，替换数据模型中的相应元数据定义，使得同一业务含义的数据只定义一次。

回归：对于已经分析得出的信息披露数据模型，为了保证其能够满足信息披露的业务需求，适合信息披露的业务场景与流程，需要对其进行回归验证。在完成逻辑推演和测试验证后，可以基本保证信息披露数据模型比较准确、完整地表达了信息披露业务所涉及的数据，以及数据之间的关系，适合大部分信息披露业务的场景与流程，且比较适合在信息披露相关系统中实际应用。

四、研究意义及阶段性成果

证券期货行业数据模型通过应用证券期货业统一编码体系，为行业机构间、机构内部数据交换提供数据及数据表素材，用于规范和指导数据治理、机构及系统间数据交换和行业应用系统建设，最终形成“一套标准”、“一个库”、“一本书”的行业数据模型标准体系。

为开展证券期货业数据治理有关研究工作，证券分委会于2014年7月启动了DG-SDOM（基于证券期货业数据模型的数据治理）项目。该项目基于证券期货行业数据模型，从技术上描述了整个行业的业务行为，统一了各种协议、接口、编码、报文格式等技术标准。该项目在人才队伍建设、标准体系建立、数据字典扩展机制研究、机构间接口标准行业规划、标准制修订工作等方面取得了阶段性成果，并不断充实和验证了行业数据模型，逐步提高了行业数据模型的完备性和可用性。

行业数据模型是证券期货行业数据治理的核心工作，也是行业中各接口、标准制定的基础。制定行业数据模型标准用以制定中央监管信息平台建设、支持行业中机构间数据交换和全行业风险管控，为证券期货行业标准化工作奠定基础，进一步规范行业标准化管理，提升行业信息标准质量，提高行业数据治理效率。