

蒋东兴<sup>1</sup>, 高若楠<sup>2</sup>, 王浩宇<sup>2</sup>

1. 中国证券监督管理委员会信息中心, 北京 100033;

2. 中证信息技术服务有限责任公司, 北京 100033

**摘要:** 为充分发挥数据的资产价值, 监管部门高度重视大数据治理工作。通过梳理证券期货行业监管大数据治理的需求与特殊性, 对证券期货行业的大数据治理体系进行了深入研究, 包括构建证券期货行业数据模型、搭建公共数据平台、建设数据服务体系以及构建组织保障体系4个方面。通过大数据工程建设的实施, 进一步发现了证券期货行业大数据治理在工程共建共享、数据多源校核方面存在的深层次技术问题, 提出了建设超级大数据管理平台、全面保障大数据工程效果的研究思路, 为监管决策提供了更为全面、科学、客观的支持。

**关键词:** 证券期货行业; 大数据治理; 科技监管

论文引用格式:

蒋东兴, 高若楠, 王浩宇. 证券期货行业监管大数据治理方案研究[J]. 大数据, 2019, 5(3): 23-34.  
JIANG D X, GAO R N, WANG H Y. Research on supervising big data governance method for securities and futures industry[J]. Big Data Research, 2019, 5(3): 23-34.

1 引言

中国资本市场经过20多年的发展, 已具备了相当的规模体量, 基本形成了较为完备的多层次资本市场体系。近年来, 金融科技的快速发展为资本市场注入了新的活力, 大数据、人工智能等技术与金融行业的深度融合, 推动了产品形态、盈利模式的不断创新。跨行业、跨市场、跨地域的金融服务日益丰富, 借助新兴技术的力量, 市场中涌现了智能投顾、智能交易等应用, 这些应用为市场提供了更为高效、全面、智能化的业务服务, 提升了证券期货行业机构的经营管理能力, 推动了资本市场的快速发展。

信息科技的飞速发展对资本市场的监管工作提出了新的课题和挑战。监管部门既要促进金融行业创新发展, 又要守住不发生系统性金融风险的底线, 因此需要利用科技的手段及时发现潜在风险, 提高监管效能, 维护市场稳定。应用监管科技需要更多数据进行支撑, 包括传统的交易、披露、监管等内部数据, 也包括工商、司法、税务、舆情等外部数据, 并需要具备在海量规模数据下进行快速的历史数

据分析、实时数据流处理以及利用人工智能算法深入挖掘数据价值的能力，因此，监管大数据平台的建设变得尤为重要。

证券期货行业监管工作涉及中国证券监督管理委员会（以下简称证监会）机关、派出机构和会管单位等，机构数量较多，机构之间的业务关联性强，数据尚未完全在同一个数据模型上进行整合，并且存在数据敏感性强、时效性高的特点以及数据共享技术标准仍不完善等问题。海量、多源异构的数据以及监管数据的特殊性对数据的管理、存储和应用提出了新的要求。因此，证券期货行业监管科技的应用需要进一步建立符合数据科学和行业特点的大数据治理体系，以提供全面的数据治理保障，从而充分发挥数据的资产价值。

本文介绍了证监会大数据工程建设的内容，分析了证券期货行业监管大数据治理的必要性。针对资本市场监管业务关联性强、参与主体多、监管数量大、数据敏感性高等特点，对证券期货行业监管大数据治理方案进行了深入研究。最后，本文提出建立超级大数据管理平台，以全面保障大数据工程建设，从而推动证券期货行业监管科技规范、快速、健康地发展。

## 2 相关工作

数据治理目前尚未形成统一、标准的定义。国际数据管理协会（Data Management Association, DAMA）将数据治理定义为对数据资产管理行使权力和控制的活动集合（规划、监控和执行）[1]。国际数据治理研究所（The Data Governance Institute, DGI）认为，数据治理是包含与信息相关的过程的决策权及责任制的体系，根据基于共识的模型执行，描述谁在何时何种情况下采取什么样的行动、使用什么样的方法[2]。

目前，数据治理及相关标准体系的研究是国内外研究的热点之一[3]。国际标准化组织（International Organization for Standardization, ISO）也着力于数据治理国际标准的制定工作。2014年6月，ISO/IEC JTC1/SC40（IT服务管理和IT治理分技术委员会）召开第一次全会，2014年11月、2015年5月召开第二次、第三次IT治理工作组会议，我国于2015年5月在第三次工作组会议上正式提交了《数据治理白皮书》国际标准研究报告，报告中提出了数据治理模型和框架[3]。由中国国家成员体（SAC）申请立项并由中国专家作为联合编辑研制的国际标准ISO/IEC 38505-1《信息技术-IT治理-数据治理第1部分：ISO/IEC 38500在数据治理中的应用》于2017年正式发布，我国提出的数据治理理念和方法论在国际上达成共识[4]。

大数据背景下，数据科学研究的兴起以及组织业务的增长对数据治理的方法及手段均提出了新的要求。大数据治理是广义信息治理的一部分，它制定与大数据的优化、隐私、货币化相关的政策与目标[3]。张义祯[5]从社会治理的角度阐述了要顺应大数据时代的发展趋势和本质要求，树立大数据治理意识。梁芷铭[6]梳理了大数据治理的概念和理论框架，明晰了大数据能力和国家治理能力的变迁，提出通过数据挖掘、数据整合、数据分析、数据共享和数据推送等方式提升国家治理能力。常朝娣等人[7]通过分析大数据治理现状，分析了医疗健康大数据的资源特性及治理问题，基于现有大数据治理方法和模型，探索了大数据环境下的医疗健康数据治理体系和实施步骤。顾立平[8]、任亚忠[9]从不同角度论述了大数据背景下图书馆事业的数据治理问题。

近年来，证监会大力发展监管科技工作，并积极推动数据治理的探索，上海证券交易所和深圳证券交易所（以下简称沪深交易所）已开展大数据平台建设，实现了基础平台建设和部分数据及应用的迁移，并依托大数据平台探索结合业务需求的应用服务，达到提升监管效率的目的。证监会于2014年开展以数据模型为基点的行业数据治理工作，建立了行业标准化数据模型，并且研制数据模型的方法论被纳入数据治理国家标准中，推进了与证券期货相关的数据模型的国际化进程。

3 证监会大数据工程建设内容

证监会大数据工程建设旨在构建覆盖宏观监管与业务监管的智能化监管平台，整合证券期货行业的监管信息与数据资源，充分发挥科技在证券期货行业监管工作中的作用，有效提升资本市场的监管效能，防范系统性金融风险，促进市场主体健康有序发展，切实保护投资者的合法权益，为监管工作提供更为全面、科学、客观的决策支持。

大数据工程建设内容主要包括：构建逻辑上融合的监管大数据平台，设立多个灵活、智能的数据分析中心，提供多项标准、多样的专业分析服务，形成与中央监管信息平台的有效联动。证监会大数据工程总体架构如图1所示[10]。中央监管信息平台是证监会监管工作的信息化支撑平台，整合了证监会各个方面的监管应用系统，通过数据共享和流程互通，提高了监管的有效性和针对性。

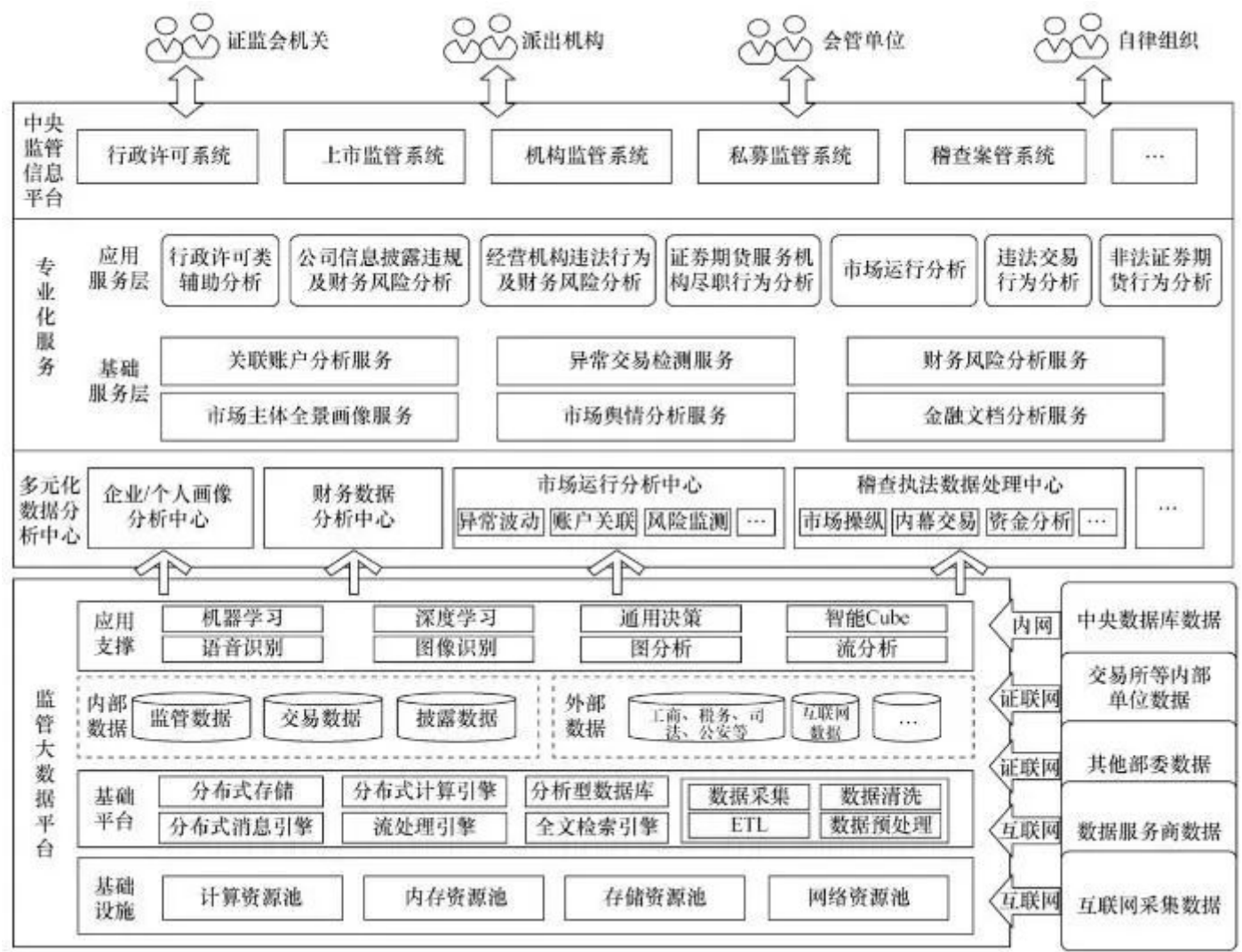


图1 证监会大数据工程总体架构

### 3.1 监管大数据平台

证监会大数据工程的核心任务是建设一个运转高效的监管大数据平台。平台承载交易数据、披露数据、监管数据和外部的各类数据资源，在逻辑上集成统一，为上层的各类数据分析和应用提供基础性的数据支撑。

监管大数据平台可分为基础设施层、基础平台层、数据服务层和应用支撑层。其中，基础设施层利用虚拟化或容器技术，实现对计算资源、内存资源、存储资源、网络资源等的统一管理，在逻辑上构建证监会专有云平台，为上层提供硬件资源保障；基础平台层利用分布式架构实现对海量数据的采集、存储、计算和管理，为数据服务层提供分布式存储管理服务；数据服务层通过采集各类内部、外部数据资源，提供全面的数据服务；应用支撑层提供深度学习、图分析等通用的算法和模型以及语音识别、图像识别等工具，对数据进行加工处理，供上层分析中心使用。平台内部各层之间相互关联，互为基础和条件，共同构成逻辑上融合的监管大数据平台。

### 3.2 多元化数据分析中心

基于监管大数据平台提供的全量数据，根据监管场景的不同，证监会大数据工程设立了多个智能分析和处理中心，为监管工作提供专业化服务，如企业/个人画像分析中心、财务数据分析中心、市场运行分析中心等。数据分析中心按监管领域划分，每个数据分析中心均可以依据职责申请使用大数据平台中的海量数据，提供一种或多种分析服务，多个数据分析中心也可对某一监管场景提供有针对性的多元化服务。

### 3.3 专业化服务

由于监管科技贯穿于监管工作的事前、事中、事后整个链条，在专业服务建设过程中，要统筹考虑底层通用功能与应用层定制化功能，构建双层服务架构，高效有序地实现监管的科技化和智能化。

双层数据分析服务架构包括基础服务层和应用服务层。基础服务层由独立的基本数据分析服务构成，每个基础数据分析服务提供单一的基本公共数据分析服务。基础服务层具体包括关联账户分析服务、异常交易检测服务、财务风险分析服务、市场主体全景画像服务、市场舆情分析服务、金融文档分析服务六大基础分析能力。应用服务层面向业务部门，被划分为多个应用分析场景，提供业务条线的定制化大数据分析服务，每个应用服务可能依赖一个或多个基础服务。应用服务层主要包括行政许可类辅助分析服务、公司信息披露违规及财务风险分析服务、市场涨跌动力分析服务、非法证券期货行为分析服务等重要业务分析方向，同时在这些业务方向之下分别细化出多个应用分析场景，解决监管科技化、智能化问题。

各分析中心将基于大数据平台分析得出的结果以功能模块或数据服务的形式提供给中央监管信息平台中的应用系统，在应用系统中进行展示及后续处理，为监管工作所用，有效提升资本市场的监管效能，辅助监管决策。

## 4 监管大数据治理方案

## 4.1 大数据治理需求

为保障证监会大数据工程的建设，需要制定相应的数据战略和管理方法。数据治理能够统筹指导、规划、控制各项数据管理活动，是厘清数据资产的定义和范畴、发挥数据资产价值的关键点。

在证券期货行业监管领域，大数据治理存在其特殊性，主要体现在以下4个方面。

- 参与机构多。证券期货监管系统包括证监会机关、派出机构和会管单位等多个单位，每一个单位都是一个独立的实体机构。其中，不少机构自身就建有大数据平台，如沪深交易所等。
- 系统业务关联性强。多个单位协同承担市场重任，一项监管业务所需的分析数据源可能涉及多家机构，并且随着跨行业、跨市场、跨地域交易量的日渐增长，产品间的联动、传导还在不断加强。
- 资本市场参与主体数量较大。我国拥有1.2亿股民、3 500多家上市公司、11 000多家新三板公司、24 000多个私募基金管理人等，参与主体数量庞大、类型复杂。
- 数据敏感性高。数据涉及交易、披露、舆情、监管等各类信息，数据的敏感性和时效性非常强，一些信息的发布可能对市场造成较大影响。

证监会大数据治理必须要考虑到上述特殊性，在全系统逻辑融合的监管大数据平台建设中，除了传统的数据标准管理、元数据管理和数据质量管理外，还必须解决多机构间的数据关联、数据血缘管理、多层次数据共享和敏感数据保护等问题。

## 4.2 监管大数据治理体系

通过调研学习和需求分析，笔者提出了证券期货行业监管大数据治理体系的初步构想，主要包括构建证券期货行业数据模型、搭建公共数据平台、建设数据服务体系、建立数据治理实施与管理组织保障体系4个方面。通过构建行业数据模型，厘清行业数据脉络，规范机构数据交换，支持行业标准化，实现对证券期货行业大数据的整体把握。通过搭建公共数据平台，整合行业内外数据，统筹加工处理，为监管系统提供数据支撑。通过建设数据服务体系，实现数据融合共享，支持多层次数据共享，提高数据利用效率。建立证监会系统逻辑融合的监管大数据平台实施与管理的组织保障体系，确保大数据治理工作的有序、有效实施。监管大数据治理体系如图2所示。

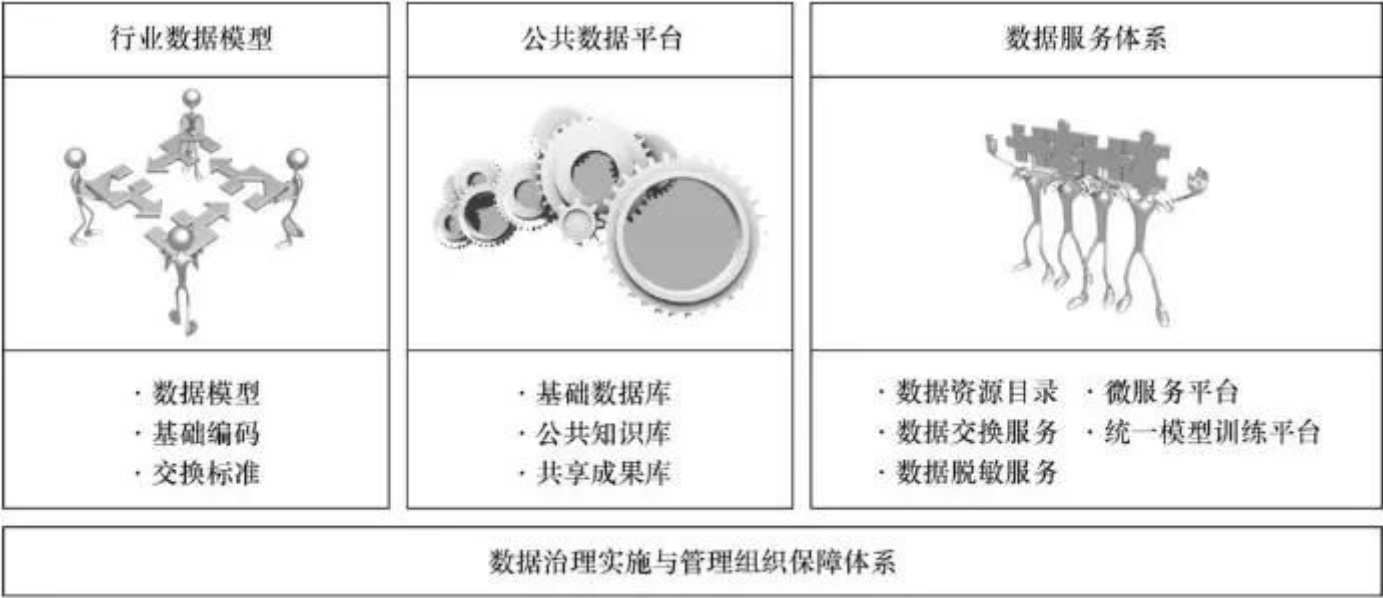


图2 监管大数据治理体系

4.3 构建行业数据模型

构建行业数据模型包括数据模型的构建、基础编码及交换标准的制定。其中，核心工作是数据建模，数据建模为行业数据语言的统一和规范化提供了重要依据。基础编码能充分打通各业务系统，实现市场理解共识，有效减少市场业务运营成本，提高行业监管效率。交换标准通过全面梳理机构内部关键业务的流程与数据交换，研究各机构内外部信息交互的共性热点问题，致力于促进各机构快速建立统一、开放、灵活的内部信息技术架构，统一机构间的信息交换模式。

4.3.1 数据模型

数据模型通过规范数据定义、定义接口规范，形成行业数据生成、存储、交换脉络图。数据模型既是数据治理的核心，也是数据价值挖掘的重要基础。证券期货行业数据化程度相对较高，机构多、类型广，交易方式多样，机构内及机构间数据交换频繁，业务发展迅速。因此，需要研究制定一套完整的行业数据模型，清晰描述整个市场的数据流向、数据定义、结构类型和关联关系等，为推动实施行业数据治理打下坚实基础。

建设证券期货行业大数据平台的一个关键是处理由行业数据来源多样性带来的固有无异构性问题。在进行数据模型设计时，必须针对数据的多源异构性进行设计，即对于同一个对象或事务的数据描述，其数据模型既要支持多个来源（包括不同来源的数据血缘记录），又要包容数据的不同形态（包括结构化、半结构化乃至非结构化数据）。在数据模型层面，证券期货行业大数据平台支持数据多源校核、数据存异、数据血缘追溯以及历史数据回溯等处理操作。

4.3.2 基础编码

随着交易场所、金融产品的增多，参与主体也越来越多。不同机构、不同场所对相关数据的编码规则不一致，导致金融产品跨机构、跨交易场所交易时难以进行，同时，也使监管机构无法及时掌握市

场上的真实数据，无法有效防范全市场的系统性金融风险。因此，亟需制定证券期货行业基础编码标准。

证监会参考相关的国际标准和国家标准，针对我国证券期货行业的具体业务和特点及行业发展趋势和监管要求，研究制定了一套统一的基础类编码标准体系和方法。编制行业基础编码的主要困难是如何确保编码能够完全覆盖业务场景和能够完全正交。场景覆盖不全或编码存在交叉都会影响编码使用效果。通过梳理所有可能的参与人、场所、活动、产品，分别形成参与人树、场所树、活动树、产品树，按照一定的逻辑方法逐一遍历这几棵树，形成多维空间中的多个节点。对每一个有意义的节点进行分析，分析在此节点上对应的参与人、场所、活动、产品是否需要编码，形成需要编码的参与人集、场景集、活动集、产品集，建立包括业务参与人、场所、活动、产品4个维度在内的多维编码体系模型，并严格遵照这个模型开展基础编码工作。

### 4.3.3 交换标准

交换标准分为机构间交换标准和机构内交换标准。为适应证券期货行业各类业务的创新与发展，降低行业整体数据的通信安全风险与应用成本，进一步规范行业机构间各类接口标准的制定与应用，提高机构间接口标准化水平，证监会制定了《证券期货业机构间接口标准规划》。同时，为解决各机构内信息系统数量众多、缺乏统一标准、数据共享难度大、系统架构相对耦合、市场技术就绪周期长等问题，证监会制定了《证券期货业机构内部接口标准规划》。目前，证监会正在组织行业专家根据规划，按照急用先行的原则，不断建立、完善数据的交换标准。

## 4.4 搭建公共数据平台

公共数据平台包括各类基础数据库、多个公共知识库和共享成果库，具有规模大、数据量大、数据使用人员广等特点。各类用户在不同的环境中可直接从公共数据平台调取基础数据、获取公共知识、共享分析成果，实现多层次数据共享，使证券期货行业监管工作更加便捷、同步、立体化。

### 4.4.1 基础数据库

基础数据库是与证券期货行业监管相关的公共基础信息库，其主要目标是解决监管业务数据库多头建设、数据无法有效流动的问题。因此，基础数据库建设必须严格遵循行业数据模型和基础编码，确保基础数据库的公共基础属性，并引领行业数据库的标准化。在构建基础数据库时，主题数据库是对各个监管业务数据库的重构，强调各个应用系统的“共建共用”。监管过程涉及众多主体，如上市公司、非上市公司、拟上市公司、证券公司、基金公司、私募机构、资产评估机构等。主体数据是主题数据库建设的基础，在立体化监管过程中，对上述主体信息的维护是一项基础且重要的工作。除主体数据外，交易、信息披露、舆情等数据也是重要的基础数据，这些数据可通过数据采集以及外部数据交换等方式获得。

### 4.4.2 公共知识库

目前，证券期货行业相关公共专业知识虽然能通过书籍、在线视频等方式获取，但其体系化、精确化程度不够。同时，传统标签体系存在搜索范围大、搜索效率低，并且已有标签体系无法满足查询需求等问题。例如，对于突发的“长生生物疫苗造假”事件，该事件发生突然、传播迅速、影响较大，若标签体系中没有预先定义“疫苗造假”标签的含义或规则，就无法对业务系统标注“疫苗造假”，也无法对该热点进行筛选、识别和检索。证券期货行业监管公共知识库初期拟建设行业标签体系、金融词库和上市公司分类库，形成证券期货行业机器可读的“图书馆”，使机器学习更有针对性，进一步提高知识传播与利用的效率。

公共知识库将打造一个机器可读、用户可编辑、动态更新的基于语义的超大知识数据库。将机器可读词典作为统一资源，对释义项进行分类，然后基于释义分析自动生成用于抽取词汇知识的模板，采用模板匹配的方法，实现词汇知识的自动抽取。公共知识库建设的关键是要解决知识自生长与知识质量的问题，这是两个具有一定冲突性的目标，前期拟通过人工整合已有多方知识的方法提高有监督学习的初始质量，后期将采用微服务平台开放公共接口，供各参与方利用公共知识的同时即时反馈评价，形成持续的强化学习机制，以此协调两个目标同时达成，不断完善公共知识库。

#### **4.4.3 共享成果库**

行业传统信息化工程以“单兵作战”为主，数据分析结果难以共享，这使得分析结果不能充分发挥作用。例如，在关联账户分析中，不同单位各自建设独立的账户库，没有与其他机构有效关联，难以形成跨行业、跨地域、跨市场的关联账户分析。证监会大数据工程从监管业务中抽象出关联账户分析、财务报表分析、实体画像、交易异常检测、舆情分析、金融文档分析六大基础分析能力，归纳出7类32种场景分析服务。通过建设统一的公共基础分析服务平台，将分析方法、分析结果汇总，形成共享成果库，监管各参与方均可利用共享成果库的内容。

### **4.5 建设数据服务体系**

数据服务体系的建设充分体现了证券期货行业监管大数据治理的意义与价值，具体包括为全系统数据共享提供资源目录和数据交换服务，为敏感数据提供数据脱敏服务，为多层次数据共享提供一套支持快速开发、部署、运维管理和持续开发集成的微服务平台以及为证监会系统各单位分析测试提供统一的模型训练平台。

#### **4.5.1 数据资源目录**

数据资源目录用来解决数据的收集、数据标签化处理及数据目录的统一展示。目录中存储与证券期货行业监管相关的数据的核心元数据，包括数据资源名称、发布时间、内容描述、数据提供方、数据标签、数据分类、共享方式、数据级别、采集频率等。在对数据资源目录进行汇总的基础上，逐步开展数据资源整合服务，提供行业数据资源的完整数据视图。

#### **4.5.2 数据交换服务**



数据交换服务是数据服务体系的重要组成部分，按照松耦合的数据共享模式进行架构设计，支持多种类型数据源的交换，为各类数据库之间的数据交换提供统一服务。考虑到各会管单位之间的相对独立性和数据的异构性，数据交换服务管理模式与一个单位内部的数据管理模式存在较大差异，各单位的信息系统不通过数据库的直接连接进行数据交换，而是连向统一的平台，由数据交换系统统一维护与各个单位的数据接口，由相关工作组统一协调数据资源，为各个单位信息系统提供统一的数据交换服务。

#### 4.5.3 数据脱敏服务

在数据交换开展之前，应依法依规对数据开展脱敏工作，以保护敏感信息的安全性。数据脱敏是指从原始环境向目标环境进行敏感数据交换时，通过一定的方法消除原始环境中数据的敏感性，并保留目标环境业务所需的数据特性或内容的数据处理过程。数据脱敏能够在数据共享时有效地防护敏感数据的外泄。数据脱敏服务应完成3项基本任务：敏感数据分级分类、敏感数据应用场景定义和数据脱敏任务实施。

敏感数据分级分类应根据数据重要性、私密性、指向性的不同，并充分考虑数据应用范围、脱敏后数据对原数据业务特性的继承（如保持原数据间的依赖关系）等因素进行划分。通过分析脱敏数据的各种应用场景，并兼顾使用场所的网络环境，把敏感数据应用场景分为5类：内部分析、系统仿真、监管协作、外部分析、开放测试。基于上述工作，数据脱敏任务实施包含确定待脱敏数据、确定数据应用场景、确定脱敏策略和算法、脱敏实施、脱敏评价5个步骤。

#### 4.5.4 微服务平台

微服务通过抽象出监管业务的逻辑确定服务的边界，每个服务只关注自己边界内的事件。微服务可独立扩展伸缩，不同的微服务可由独立的团队维护，满足了业务系统快速开发及部署的需求，增强了开发平台的支撑柔性。微服务平台支持各分析中心共建共享基础数据库、公共知识库和共享成果库，实现多方协同的多层次数据共享。

#### 4.5.5 统一模型训练平台

监管大数据分析工作涉及跨行业、跨地域、跨市场的数据，在数据分析模型训练过程中，需要汇集大量的、不同领域的综合性数据，并需要对敏感数据进行脱敏。统一模型训练平台为各数据分析中心提供模型训练所需的基础数据及一定的数据计算能力，支持技术协作，减轻各单位的负担，提高系统资源的利用率。

### 4.6 建立组织保障体系

证券期货行业监管大数据治理工作的顺利推动需要管理制度为其保驾护航。证监会建立了多层次的大数据治理组织架构，设立监管科技工作建设管理委员会、科技监管专家咨询委员会、监管科技建设与协调工作组、大数据平台规划建设小组，并分别定义了其角色与职责。其中，监管科技工作建设管理

委员全面负责大数据工程的统筹、规划、协调和推进；科技监管专家咨询委员会对方案规划、工程实施评估等事项提供决策咨询；监管科技建设与协调工作组负责统筹协调各方资源以及日常工作管理；大数据平台规划建设小组主要负责大数据平台建设和数据治理工作的实施。多层次的组织架构为证券期货行业监管大数据治理工作提供了组织管理保障。

## 5 工程实施与问题发现

### 5.1 工程实施

证监会大数据工程建设采用多方协同、共建共享的工作策略，以业务为导向，先有后优、重点突破，形成推动监管科技发展良好生态体系。数据治理工作早在2014年就开始规划和推动了，目前在行业数据模型研制、公共数据平台搭建、数据服务体系建设等方面均取得了一定的成果。

目前，行业数据模型已初见成效。抽象模型已形成了能够体现行业业务脉络、表达业务运行情况和数据特征的行业数据素材库；建立了数据模型管理平台，为行业共享数据模型成果提供服务。在基础编码方面，证监会信息中心组织证券期货业机构成立了基础编码专业工作组（WG1），致力于证券期货行业基础编码的制定、修订的统一工作，已发布3项基础编码国家标准和4项基础编码行业标准，正积极推进两项基础编码国家标准和一项基础编码行业标准的制定工作。在交换标准方面，2014年9月，全国金融标准化技术委员会证券分技术委员会（以下简称证标委）成立了11个专业工作组，启动了证券期货业标准规划工作，并于2015年7月发布了《证券期货业机构间接口标准规划》和《证券期货业机构内部接口标准规划》。

此外，为了更好地支持行政许可、行政处罚等工作中的业务分析，围绕证券期货行业监管对象，证监会分析了大数据平台中各类外部数据资源，正在建设证券期货行业各类参与主体的基础数据库。针对大数据工程六大基础分析能力建设，组织了6个技术协作组分别开展研究，现已提出基于大数据分类和异常检测的财务分析方法、基于图分析和标签化处理的资本市场实体画像方法、基于 $\Delta X \Delta T$ 模型的证券市场异常交易检测方法、基于社区发现和稳态分析的关联账户分析方法等工程方法。采用上述工程方法，已经开展了私募机构画像与风险发现、上市公司画像与风险发现等工作，开始形成对证券期货行业相关主体异常问题的及时发现和预测的动态监管机制，进一步提升防范和化解金融风险的能力。

在数据服务体系建设方面，证监会建设了证券期货监管系统数据资源目录服务平台，实现监管系统内外部数据资源目录的集中展示。目前，该目录平台已基本完成数据资源目录的收录，使得数据整体情况得到统一的展示，为数据交换服务打下了基础。在数据共享制度建设方面，目前，证监会正在研究监管科技数据共享、与数据脱敏相关的制度，以进一步明确数据共享各参与方的职责，梳理数据共享流程管理，提出数据脱敏技术要求和管理要求。

### 5.2 问题发现

在证监会大数据工程建设的实施过程中，笔者充分感受到了证券期货行业监管数据具有复杂性、海量性和动态性及多源、多关联和多维度特征，传统的针对单一机构的数据标准管理、元数据管理和数据质量管理已经无法满足其数据治理需求。行业数据模型的建立在很大程度上解决了数据末端的模型管

理，但无法展现业务梳理过程的中间产物，不便于业务人员的理解，并且由于启动较早，对行业外部的大数据资源考虑不够充分。证券期货行业数据流转复杂，机构间数据关联性强，数据源头多样化，数据血缘关系难以管理，导致数据质量难以把控。高度敏感的资本市场数据如何实现外部监管协作，更加方便、快捷、高效并且依授权地共建共享，也是当前面临的挑战性工作。

此外，数据集成要求数据共享是安全、可靠的。多源异构数据的数据集成是较为复杂的问题，通常使用者并不清楚数据集成的意义，也不知道如何对已经联系在一起的数据进行操作。数据管理者在系统设计时，必须考虑到这种情况——用户的要求可能会增加数据集成的复杂性，数据集成必须能够适应这种状况。

因此，面对与证券期货行业监管相关的“数据湖泊”群，其治理需要构建一个超级大数据管理平台，不仅要解决数据标准、元数据管理和数据质量管理问题，还要解决数据模型管理、数据源管理、数据血缘关系、多源数据校核、多层次数据共享与服务等问题。

## 6 结束语

伴随着以大数据、人工智能、云计算、区块链为代表的高新技术的快速发展，金融科技与监管科技成为推动证券期货行业实行跨越式发展的重要手段，大数据已经是证券期货行业的战略资源和重要资产。监管部门应高度重视大数据治理工作，制定相应的数据战略和管理方法，统筹指导、规划、控制各项数据管理活动，充分发挥监管数据的资产价值。研究发现，建立完备的组织架构体系是数据治理工作顺利开展的根本保障；明晰行业的大数据治理需求是数据治理工作的前提；建立完整的行业数据模型，规范行业术语，清晰描述整个市场的数据业务是数据治理工作的基础；支撑行业监管，提供完善的数据服务是数据治理工作的落脚点；特别地，根据证券期货行业大数据的特殊性，解决数据工程共建共享、数据多源校核等方面的深层次技术问题，是数据治理的关键所在。

笔者在研究与实际工作中，针对证券期货行业监管业务关联性强、参与主体多、监管数据量大且敏感性高等特殊性，提出了初步的证券期货行业监管大数据治理方案及工程实施中发现的问题，对该领域后续工作的展开有一定的指导意义。

## 作者简介

蒋东兴（1970-），男，中国证券监督管理委员会信息中心研究员、副主任，负责证监会监管科技建设工作，研究证监会监管科技规划和整体架构曾任清华大学计算机与信息管理中心主任、信息化工作办公室副主任、保密管理办公室主任，中国高等教育学会教育信息化分会理事长，教育部教育管理信息化专家组组长承担了多项国家科技攻关项目、支撑计划项目、教育部重大信息化项目，发表论文140余篇，出版书籍5部，取得专利4项，研制国家标准3项，获得北京市科学技术奖三等奖、教育部提名国家科学技术奖二等奖等省部级奖项。

高若楠（1992-），女，就职于中证信息技术服务有限责任公司，主要研究方向为数据治理。

王浩宇（1992-），男，就职于中证信息技术服务有限责任公司，主要研究方向为数据治理。

## 《大数据》期刊

《大数据（Big Data Research, BDR）》双月刊是由中华人民共和国工业和信息化部主管，人民邮电出版社主办，中国计算机学会大数据专家委员会学术指导，北京信通传媒有限责任公司出版的中文科技核心期刊。