

金融统计标准下元数据信息框架构建研究

金融统计元数据研究课题组

(中国人民银行南昌中心支行调查统计处 江西 南昌 330008)

摘要: 2008年国际金融危机后,国际金融组织和主要经济体不断总结经验,积极构建银、证、保等相互协调的金融业综合统计体系。中国同样面临相关问题,特别是随着金融业态的不断发展和创新,金融风险集聚、传染和扩散更加复杂,“一行三会”金融监管模式下,宏观调控难度加大,2015年中国股市的暴涨暴跌充分暴露了金融监管的问题。因而,在中国建立银、证、保等统一、全面、共享的金融业综合统计体系是顺势而为,也迫在眉睫。其中,元数据及标准化作为金融业综合统计体系的基础性工作应首当其冲。因此,文章从金融统计元数据的应用现状出发,结合国内外已有的金融统计标准和先进技术,构建我国金融统计元数据信息模型和数据共享平台,并采用贷款实例评估元数据信息框架在统计工作中的应用情况,最后就如何继续深入推动标准化落地和加快信息共享平台建设提出建议。

关键词: 金融统计; 元数据; 标准; 信息模型; 共享平台

中图分类号: F831.2 **文献标识码:** A **文章编号:** 2095-0098(2018)03-0031-09

一、研究背景及概念

(一) 研究背景

2008年,由美国次贷危机引发的金融危机,迅速扩散到全球演变为国际金融危机,并蔓延至经济领域,欧、美、日等发达经济体陷入全面衰退,现仍处于修复阶段。国际金融危机的影响之广、扩散之快、危害之大,是某些国家对危机的起源、扩散和传染判断出现了严重失误,其中一个重要原因是金融统计未能及时反映金融变化和危机迹象,导致防范措施滞后和偏差。危机之后,国际金融组织和主要经济体不断总结经验和教训,积极构建银、证、保等相互协调的金融业监管模式,修订的货币政策和法律框架拓宽了金融统计的覆盖范围和数据获取渠道。如英国重构了金融治理框架,设立金融行为局,负责所有金融机构的行为监管,实现了对主要金融监管机构的宏观统一监管;欧盟完成了银行业联盟立法,启动了由欧央行负责运行的单一监管机制。近两年,欧美等国又充分认识到一套全面及时的数据统计标准对于经济金融市场稳定的重要性,为消除全球系统重要金融机构之间存在的巨大信息差异,G20发布了《消除数据差异提案》,致力于统一整合全球经济金融数据统计方式。

金融信息统计缺失及信息差异等方面的问题,中国同样存在。随着我国金融业发生深刻变化,新型金融机构、金融创新不断涌现,金融市场日趋复杂,金融体系关联度大幅提高,金融风险集聚、扩散、传染更加复杂,而分业监管职能分散,效率低下,传统的金融统计分析框架无法全面反映复杂金融环境下的风险及蕴含的危机,宏观调控难度日益加大。如2015年中国股市经历的震荡,尤其是7月份的暴跌,充分暴露了我国资本市场制度的内在缺陷,监管部门对资本市场的交易行为不能实行统一监管,缺乏全面的金融业统计数据

收稿日期: 2018-03-13

基金项目: 中国人民银行重点课题“金融统计元数据研究”

课题组成员: 陈燕(1964-),女,江苏盐城人,本科,经济师,研究方向为公共管理、金融;

廖燕平(1985-),女,江西赣州人,硕士,经济师,研究方向为金融经济;

王世醒(1989-),男,安徽宣城人,硕士,经济师,研究方向为金融统计、软件工程;

周璐(1994-),女,江西景德镇人,本科,研究方向为国际金融。

信息,难以监测资本市场风险对金融体系的影响,以致资本市场风险迅速在金融体系内部扩散,甚至向实体经济传染,影响金融稳定和经济增长。

在这一背景下,中国金融改革和全面监管势在必行。2016年,“加强金融宏观审慎管理制度建设,改革并完善适应现代金融市场发展的金融监管框架,实现金融风险全覆盖”纳入了“十三五规划”。习近平总书记明确提出,要借鉴国际金融监管体制改革经验,统筹负责金融业综合统计,通过金融业全覆盖的数据收集,加强和改善金融宏观调控,维护金融稳定。因此,在中国构建银、证、保等相互协调的综合性金融体系及金融业综合统计是借鉴国际经验的重要举措,是顺应时代的需要,是金融发展的必然结果。

金融业综合统计作为一项重要的金融基础设施,受到了党中央、国务院的高度重视,有利于扩大金融统计数据的覆盖率,使金融统计数据能及时、准确、全面地反映我国金融业发展状况,为及时防范金融风险、提高金融监管有效性、维护金融稳定提供更准确、全面的信息支撑。而金融统计标准化和金融统计元数据是推动金融业综合统计的重要基础性工作,应首当其冲。

(二) 相关概念引入

所谓金融统计标准化,是对金融统计中概念、定义、分类和核算原则的标准化过程^[1],其实质就是元数据的标准化。元数据是实体数据的基本单位,包含关于数据的组织、数据域及其关系等信息^[2],而标准化是对共同使用和重复使用的条款制定最佳秩序的过程。简单来说,如我们经常填写的《个人情况登记表》,包括姓名、性别、民族、政治面貌、学历、职称等字段,这些字段代表的信息就是元数据,它们通过不同的组合规则描述了个人的不同情况,而对姓名、性别等字段信息如定义、长度、类型等进行规范的过程便是标准化。在金融统计中,元数据是实现数据共享和交换的基础,在金融业综合统计体系的构建中起着重要作用,标准化了的元数据能减少银、证、保等不同部门间的数据转换环节,有利于加快金融业综合统计建设的进程。

二、我国金融统计元数据的应用现状

(一) 我国金融统计元数据的发展现状

在国内,越来越多的学者展开了各领域的元数据的研究,如图书馆等行业的元数据已经接近成熟,其元数据标准已经建立^[3]。在金融统计行业,中国人民银行、证监会、国家统计局等机构也紧锣密鼓地开展元数据和标准化的研究工作,在各自领域中为实现数据共享和交换作出了努力。

1. 中国人民银行的标准化及元数据应用情况

自2009年来,中国人民银行积极推动并致力于标准化建设,成立了金融统计标准化工作领导小组,先后发布多项适用于金融统计信息交换与共享的行业标准。中国人民银行调查统计司根据发布的各项标准,制定了《理财、资金信托专项统计制度(试行)》和《标准化存贷款综合抽样统计监测制度(试行)》,实现了相关金融统计数据按数据元逐条报送,成功地将金融统计元数据应用于金融统计实践中。

同时,为提高数据共享和交换效率,推进金融业综合统计体系建设,中国人民银行积极开展统计数据与元数据交换的研究工作,自2010年起开始跟踪SDMX^①技术动态^[4],借鉴欧盟、美国等国外先进经验,于2014年制定并发布了我国《统计数据和元数据交换(SDMX^①)标准》^{②③④⑤},为我国确立了统计数据和元数据交换的行业标准,为金融统计数据共享提供了保障。2015年,中国人民银行与证监会联合印发《债券统计监测管理信息系统文件接口规范》,实现了中央银行统计接口规范在证券业的落实,这是标准化工作的重要举措,也是金融业综合统计工作的重大进展和突破,实现了债券统计工作在银行、证券之间的互联互通。

2. 证监会的元数据应用及标准化情况

2002年,证监会开始制定《上市公司信息披露电子化规范》,并成功将XBRL(可扩展商业报告语言)应用于上市公司2003年年报摘要,以及2004年季报、半年报摘要、年报摘要和年报全文。这里的XBRL,是XML^⑥(可扩展标记语言)在商务报告信息交换方面的应用,它将公司的业务信息分解成基本元数据,采用数

① 统计数据和元数据交换英文全称为 Statistical Data and Metadata Exchange,简称 SDMX。

② JR/T 0107.1—2014 统计数据和元数据交换(SDMX)第1部分:框架[S]。

③ JR/T 0107.2—2014 统计数据和元数据交换(SDMX)第2部分:信息模型 UML 概念设计[S]。

④ JR/T 0107.3—2014 统计数据和元数据交换(SDMX)第3部分:SDMX-ML 模式和文档[S]。

⑤ JR/T 0107.7—2014 统计数据和元数据交换(SDMX)第7部分:Web 服务用法指南[S]。

⑥ 可扩展标记语言英文全称为 Extensible Markup Language,简称 XML。

据颗粒的方式来存储和发布财务报告,以便不同的信息使用者可以通过对这些元数据的选择、组合、汇总或归并,最终形成自己所需要的财务报告信息。

2009年,证监会发布《证券期货业与银行业数据交换消息体结构和设计规则》行业标准,定义和描述了各类业务元素、数据字典及日终数据接口规范等,以推动证券期货业与银行间数据交换标准化工作。2012年,中证资本市场运行统计监测中心成立,证监会力推金融数据标准化,先后制订了《证券期货业通讯协议应用指南》、《证券期货业的数据字典》,并于2016年发布《证券期货业信息系统托管基本要求》行业标准,并将陆续组织制定多项金融行业标准。

3. 国家统计局的元数据应用^[5]

对于国家统计局来说,元数据应用最为重要的就是2011年为统一设计、规范管理企业一套表指标、分组和目录,制定的《企业一套表主要元数据代码编制规则》,涉及到统计设计管理、数据采集处理和分析评估各环节。元数据标准在企业一套表中的应用,能够规范统一指标的名称、涵义、计算方法和代码,强化统计分类标准在统计调查中的执行,并在此基础上,制定了统计生产全过程的元数据标准。这有利于实现统计元数据的规范化和标准化管理,为统计数据和元数据的高效共享和交换提供基础。

(二) 我国金融统计元数据应用的现实问题

从我国统计实践上看,虽然各监管部门对元数据和标准化的研究均稳步推进中,但在当前我国“一行三会”分业监管的金融格局下,沟通渠道不畅通、基础设施建设分散等现实难以短时间实现金融业全覆盖数据的收集,以及数据共享和交换。

1. 部门间沟通有障碍

元数据和统计数据共享和交换的实现,涉及不同部门间职责、制度、标准的协调等问题。我国金融业监管和统计分属不同部门,统计职责、制度和标准等不一致^[6],金融监管体制的制约是收集全覆盖统计数据和元数据的重大障碍,部门间缺乏有效沟通和协调将无法准确定位和描述元数据及标准,以致于在推进金融业综合统计过程中困难重重。以江西为例,虽然前期与统计局、银监局、证监局、保监局、商务厅等进行了大量沟通,但是具体落实时,银监局、证监局、保监局均表示不加入金融业综合统计联席会议,导致在金融业综合统计开展时,证券、保险企业的数据报送及时性、准确性难以把握,典当、担保等小微金融企业的数据采集更是重复沟通,常需要通过金融办或商务厅协调。

2. 基础设施建设分散

在现有金融统计框架下,我国金融统计数据分散在中国人民银行、银监会、保监会、证监会、统计局、商务部门等。虽然各部门都致力于标准化的研究和元数据的应用,但在分业监管的背景下,基础设施的建设还是比较分散,如中国人民银行发布的标准和元数据多局限于银行业,证监会主要针对证券期货业等,即便是中国人民银行和银监会发布的银行业统计制度也存在差异,如各项存贷款口径等。基础设施建设的分散,导致元数据标准的分散,特别是数据规范的不统一,这在开展金融业综合统计及编制金融业概览时,需要进行分类标准和规范的统一转换,将耗费大量的人力、物力和时间,若在金融风险积聚时无法及时获取金融业全覆盖的信息,无疑会错过防范和遏制金融风险的最佳时间。

3. 应用研究基础薄弱

目前,我国在元数据和标准化的研究尚处于起步阶段,基础仍比较薄弱,氛围不够浓烈,重视程度偏弱,研究人员也较少。参与标准化和元数据研究的仅局限于与之相关的个别部门和人员,参与者几乎都是兼职人员,他们往往需要利用完成基础工作的间隙和大量业余时间参与或从事标准化和元数据应用研究,所以研究的时间跨度较长,视野不够开阔,掌握资料信息有限,研究深度不够。从现实来看,从事元数据应用研究的人员以总行层面为主,参与国际间元数据交流的机会也较多,但省级、地市级参与人员较少,获取最新的元数据相关信息比较困难,能参与国际间信息交流的人员更是寥寥无几。而澳大利亚统计局在推进元数据和统计数据交换时,其先进的元数据管理策略之一便是动员组织所有处室都参与其中,充分吸纳各方意见和建议,这是值得我国借鉴的经验之一。

三、金融统计标准下元数据信息框架构建

通过对我国金融统计元数据应用现状的分析使我们认识到,虽然各金融监管部门都致力于标准化的研究,但由于缺乏统一标准和共享机制,阻碍了元数据信息框架的建立。因此需要探索建立我国金融统计元数

据信息模型和数据共享平台,将各部门统计数据标准化,以消除数据结构差异,实现数据共享。

(一) 元数据信息模型构建

1. 元数据信息模型构成要素

元数据信息模型是我国金融统计元数据信息框架的基础部分,它对统计数据和元数据进行建模,将数据中重要的、不易改变的且具有共性的属性(如日期、地区)组成通用模型,来描述统计数据、元数据的结构和组织方式,并对数据处理过程做了明确定义,同时它还允许对通用模型框架进行扩展来描述数据间差异较大的属性。简单地说,元数据信息模型的构成要素包含数据结构定义、元数据结构定义、统计数据集、统计元数据集和代码表五个部分。

表1 江西省存贷款月报数据

数据日期	地区	指标	币种	数据属性	数据单位	数据值
2015-12-31	江西省	各项贷款	本外币	余额	亿元	18561.09
2015-12-31	江西省	各项存款	本外币	余额	亿元	25042.97
2016-01-31	江西省	各项贷款	本外币	余额	亿元	19149.85
2016-01-31	江西省	各项存款	本外币	余额	亿元	25529.76

我们以一个省的存贷款月报数据来解释元数据信息模型的构成要素。表1中共列举了四条存贷款月报数据,以第一条为例,它描述的是2015年末江西省本外币各项贷款的余额是18561.09亿元。可以看出,这些数据是有规则有结构的,它由“日期、地区、指标、币种、数据属性、数据单位”这些属性和数据值组成。数据结构定义就是用来描述统计数据的通用结构及相关的一系列概念。如果没有数据结构定义,使用者就无法明白这个数据值“18561.09”具体代表什么。

在固定的时间段中,这种有相同数据结构定义的统计数据的集合,就是统计数据集。比如说2010-2015年各月末,江西省本外币各项贷款余额就可以组成一个小的统计数据集。统计数据集是分层分类的,从一级分类看,目前中国人民银行主要的统计数据集有社会融资规模、货币统计概览、金融机构信贷收支统计、金融市场统计、企业商品价格指数以及景气调查指数六个方面,每个大的数据集又可以细分为多个数据子集。

统计元数据集是描述统计数据的元数据的集合,与统计数据集不同,它并不关注数据本身是如何构成的,它通常描述统计数据的提供方、接收方、存储方式、存储位置、数据间关系、交换过程及数据属性(如地区)的定义、类型、长度、取值范围等信息。与统计数据集一样,统计元数据集也有对应的元数据结构定义。元数据结构定义对元数据集的必备要素做了说明,告诉我们如何组织这些元数据、元数据结构是怎样的、不同元数据的定义和功能是什么、统计数据和元数据交换如何进行等。上述例子中,存贷款统计数据中“地区”、“指标”等属性的具体取值,并不在数据结构定义中描述,而是存放在代码表中,数据生产时通过查询相关元数据信息找到对应代码表取值。

2. 元数据信息模型构建基础

元数据信息模型建立的基础有两点:一是发布统一的元数据标准,二是将理论研究和IT^①技术充分结合。SDMX^②不仅仅是适用于中国人民银行内部数据交换的标准,它更是国际上通用的统计数据和元数据交换标准,有助于建立我国的元数据信息模型。SDMX^②标准中规定了元数据集和元数据结构定义、统计数据集和数据结构定义、代码表、数据间结构关系、数据约束等,给出了各部分要素的详细建模过程,并附有UML^③使用指南和人员操作指南,这是我国建立元数据信息模型的重要基础。

建立元数据信息模型,需在加强金融统计标准化理论研究的同时,注重IT^①技术人才的培养,充分发挥人才合力的作用。统计数据和元数据建模、数据交换标准的实施以及数据共享平台的建设是理论框架转换成实际应用的重要步骤,这些工作都需借助IT^①技术才能实现。

① 信息技术英文全称为 Information Technology,简称 IT。

② 统计数据和元数据交换英文全称为 Statistical Data and Metadata Exchange,简称 SDMX。

③ 统一建模语言英文全称为 Unified Modeling Language,简称 UML。

3. 元数据信息模型数据管理方式

建立一个覆盖广泛的元数据标准,让所有金融行业的元数据均按此标准进行定义和管理,是建立金融统计元数据信息模型的前提,也是元数据信息框架的必备要素。但统一的元数据标准并不意味着数据的全部重新采集,而是通过一种数据处理机制将原始的统计数据转换和重估后,再进行交换和共享,最后集中到数据仓库中统一管理。数据仓库中的数据定期进行更新,当用户需要数据时,可直接去数据仓库查询,这样不仅可以简化数据查询流程,而且能更好地保证查询的时效性。

我国金融统计数据集中在不同部门,如“一行三会”、金融办、商务部门等,而且不同时期的统计数据口径也有所差异。若要实现不同时期、不同部门间的数据共享,有效利用分布在各业务系统中的金融统计数据,就需要将统计数据与元数据分离,借助统一的元数据标准实现统计数据标准化,然后将各系统中的统计数据集成到一个共享的数据仓库中管理。

(二) 金融统计数据共享平台建设

金融统计数据共享平台是元数据信息框架的应用部分。构建我国金融统计元数据信息模型,借助元数据标准消除不同统计数据间差异,最终是要建立一个涵盖全金融业的有层次开放式的数据共享平台,实现统一、全面、共享的金融业综合统计体系。

1. 构建数据共享平台的条件

一是统计制度和标准发布。中国人民银行提出的中国金融标准化体系,从宏观层面统一了我国金融工具统计分类,明确了我国金融工具涵盖范围,界定了各类金融工具的含义,规范了金融工具的统计编码。中国人民银行发布的SDMX^①标准提供了统计数据及元数据交换和共享的标准化格式,其作用在于规范我国金融统计标准体系的内部处理和对外发布,促进金融统计的互联互通、信息共享和业务协同,提高信息共享的效率,满足金融业综合统计的需要。

二是局部协调框架建立。2013年8月15日,国务院批复同意建立由中国人民银行牵头的金融监管协调部际联席会议制度。联席会议办公室设在中国人民银行,承担金融监管协调日常工作。实现金融信息共享和金融业综合统计体系的协调是联席会议制度的重要工作之一。

三是金融业综合统计工作推进。近年来,依据国家规划和国务院的总体要求,中国人民银行积极鼓励各省先行开展金融业综合统计试点工作,并将互联网金融统计纳入金融业综合统计体系。2016年4月,中国人民银行和银监会、证监会、保监会决定在天津、广东、浙江、安徽四省市开展金融业综合统计试点工作,至此我国金融业综合统计工作进入了新的发展阶段。

2. 数据共享平台的组织结构

目前,我国缺乏统一的金融业综合统计体系,这为制定统一的金融统计元数据标准带来了困难,也决定了我国金融统计数据的共享平台需要分步、分层打造。

一是分步分层次建立。在目前金融体系框架下,实现各类统计数据共享可分为三个层次:一是实现中国人民银行内设部门之间的数据共享,如调查统计部门和货币信贷部门,这个层次较容易实现;二是实现“一行三会”,即中国人民银行与银监会、证监会、保监会的数据共享;三是实现“一行三会”与政府部门(如金融办、统计局)间的数据共享。从国际看,还要逐步实现我国金融统计数据与国外金融统计数据的交换和共享。

二是推进元数据标准统一。在现有金融统计框架下,若要实现各金融监管部门间海量统计数据的完全共享,为金融业综合统计铺平道路,必须依靠高效的计算机软件平台,而计算机只能识别规则性强的标准语言,这就要求我们必须先做到以下两点:一是建立覆盖全金融业的统一元数据标准;二是实现统一元数据标准与各部门自有元数据标准的对接,并消除部门间数据差异。在建立统一元数据标准过程中,应依照国家相关编码规则,结合金融统计标准化和金融业综合统计,增加二级类别和核心指标,尽可能地采用国内国际通用的标准,以满足不同部门间金融统计数据的互换。如“三证合一”登记制度将工商营业执照、组织机构代码证、税务登记证改为“一照一码”登记模式,统一社会信用代码可能将会替代组织机构代码成为国内更为通用的标准。同时,各金融监管部门还要研究如何将自身元数据标准与金融行业统一标准对应起来,这样用自身标准采集的数据才能方便快捷地共享给其他部门,如中国人民银行的金融机构编码可以与统一社会信

^① 统计数据和元数据交换英文全称为 Statistical Data and Metadata Exchange,简称 SDMX。

用代码建立对照关系,使用统一社会信用代码的部门就能快速地识别中国人民银行的数据。

三是基于先进信息技术建设。XML^① 由其本身的诸多优点而得以在国际间广泛应用,以 XML^① 技术为基础的统计数据和元数据交换已逐渐成为数据共享的主流方式^[2]。数据共享平台正是利用 XML^① 交换技术,通过数据标准化、数据表达传输、数据存储、数据交换发布、数据检索等实现统计数据的共享。XML^① 文档以其结构优势成为数据存储和传输的最佳选择,而文档结构定义(DTD^②)和可扩展语言架构(XML Schema)可以描述数据集的结构信息,故在建立数据共享平台时,可利用文档结构定义和可扩展语言架构来表达元数据信息,用 XML^① 文档表达统计数据。

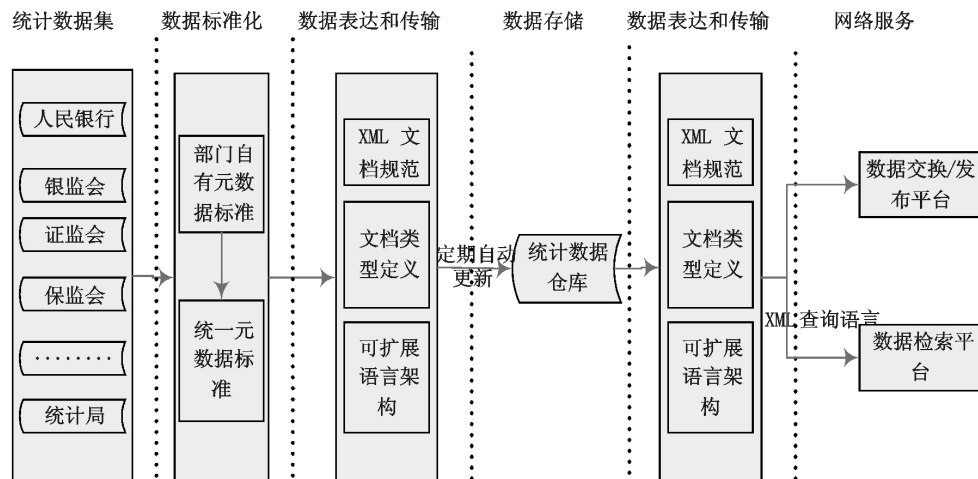


图1 金融统计数据共享平台示意图

基于 XML^① 的金融统计数据共享平台如图1所示,主要包含以下五个重要部分:

(1) 数据标准化: 首先是根据各部门自身的元数据标准,将统计数据库中的数据集合进行标准化处理,对其属性值进行重新编码。然后根据自身元数据标准和金融业统一元数据标准的对应关系,对数据进一步处理,使数据成为行业内通用的共享数据。

(2) 元数据表达和传输: 将标准化后的统计数据存储到 XML^① 文档,并将一些必要的元数据信息存储到文档结构定义和可扩展语言架构文件中,然后一并传输至数据仓库。

(3) 数据存储: 解析文档类型定义和可扩展语言框架文件,然后从 XML^① 数据文件中提取数据和元数据到数据仓库中,进行集中管理。

(4) 数据交换和发布: 将数据仓库中标准化的数据转换成可交换的 XML^① 规范文件,并将一些必要的元数据信息附加到文档类型定义或可扩展语言框架文件中,再通过数据发布/交换平台,对外发布或与其他部门进行交换。收到 XML^① 数据文件的发布平台或部门,必须再根据相关的元数据标准,将数据文件解析成其需要的结果。

(5) 数据检索: 用户可以在统计数据仓库直接查询其他部门的数据,但需要使用 XML^① 查询语言,并经过 XML^① 相关技术的处理,使用户输入的非标准化查询条件标准化,并将返回的标准化数据转换成用户可识别的信息。

四、元数据信息框架在统计工作中应用的评估

本文研究的元数据信息框架由统一元数据标准、元数据信息模型和数据共享平台三部分组成,为评估元数据信息框架在我国金融统计工作中运用的可行性,本文选取金融统计标准化中贷款元数据来模拟数据的共享和交换过程。

(一) 数据标准化处理

为方便演示,我们参照金融统计报表模拟生成一组贷款元数据指标(如表2),这组数据较接近中国人民

① 可扩展标记语言英文全称为 Extensible Markup Language,简称 XML。

② 文档类型定义英文全称为 Document Type Definition,简称 DTD。

银行《贷款统计分类及编码》^①中的要求,直接进行一次标准化即可(实际情况下金融机构的贷款数据比这个更为详细和复杂)。

表2 金融机构贷款原始数据表

数据日期	金融机构	借款人代码	企业出资人经济成分	企业规模	贷款产品类别	贷款实际投向	贷款发放日期	贷款币种	贷款借据余额	利率水平	贷款质量
2015-01-31	某农信社	101	私人相对控股	微型	经营贷款	农业	2014-08-22	人民币	36000.00	9.72	正常类贷款
2015-01-31	某农信社	102	私人相对控股	微型	经营贷款	金属制品业	2013-11-13	人民币	10000.00	6.93	正常类贷款
2015-01-31	某农信社	103	私人相对控股	微型	经营贷款	农业	2014-10-31	人民币	40000.00	9.72	正常类贷款

《贷款统计分类及编码》中详细规定了贷款数据元术语和定义,并对属性及属性描述有了详细的元数据定义,以此为参照将上述原始贷款数据标准化。数据元表示规范主要有中文名称、内部标识符、定义和数据表示、值域、计量单位、备注。如“金融机构编码”的属性及属性描述为:

表3 贷款数据元属性及属性描述

属性	属性描述
中文名称	金融机构编码
内部标识符	1010
定义	数据报送金融机构的唯一标准编码
数据表示	e14
值域	中国人民银行按照《金融机构编码规范》(银发【2009】363号)编发的代码
计量单位	
备注	

同时,编码规范中还附了相应的代码表,用于解释贷款元数据中属性定义。经过处理后,数据如表4所示,全部为标准化后的数据编码或值。比如表1中第一条贷款记录中“企业出资人经济成分”“企业规模”“贷款实际投向”原数据为“私人相对控股”“微型”“农业”,经过标准化后数据为“B0102”“CS04”“A01”。

表4 经过标准化处理的贷款元数据

数据日期	金融机构	借款人代码	企业出资人经济成分	企业规模	贷款产品类别	贷款实际投向	贷款发放日期	贷款币种	贷款借据余额	利率水平	贷款质量
2015-01-31	C1000036000010	101	B0102	CS04	F022	A01	2014-08-22	CNY	36000.00	9.72	FQ01
2015-01-31	C1000036000010	102	B0102	CS04	F022	C33	2013-11-13	CNY	10000.00	6.93	FQ01
2015-01-31	C1000036000010	103	B0102	CS04	F022	A01	2014-10-31	CNY	40000.00	9.72	FQ01

(二) 数据共享与交换

共享过程中,XML^②文件按标准的格式规范生成,内容中大致可分为两个部分:一是头结点,负责存储共享属性;二是数据列表,存储具体数据记录。头结点存储共有属性,可减轻数据存储和读取的冗余,如数据日期和金融机构编码。相关元数据信息(如数据类型、编码标准等)则放到类型定义文件和语言架构文件中。以上三条贷款数据经过XML^②表达和转换后,就可传输到统计数据仓库中存储。

标准化的贷款数据存储后,就可以与外部机构进行交换了。若是数据交换的双方拥有一套完整的元数据标准,则可直接通过此标准转换,形成可用数据,如中国人民银行内设部门之间的数据转换;若是双方数据并未使用同一套元数据标准,则需要借助元数据标准对照关系进行转换,通过数据转化、合并,形成所需要的结果。

上述评估结果表明,本文提出的金融统计元数据信息框架能够应用到统计工作中,元数据信息模型和数据共享平台也是可行的。

① JR/T 0064.2-2011,金融工具统计计值 第2部分:贷款[S].

② 可扩展标记语言英文全称为 Extensible Markup Language,简称 XML。

五、结论及相关建议

综上所述,我们尝试在金融统计元数据的理论研究和实际应用上进行探究,从金融统计元数据的应用现状出发,建立我国金融统计元数据信息模型及我国金融统计数据共享平台,并通过对元数据框架在统计工作中的应用情况进行评估,得出的结论是构建的元数据信息模型及数据共享平台总体是可行的,也是符合实际的。2016年初,陈雨露副行长在全国人民银行调查统计工作电视电话会议上强调,加快构建“统一、全面、共享”的金融统计体系以及金融统计标准化建设。要实现全国范围内金融统计数据的共享,前景很美好,任务仍艰巨。

(一) 扩大金融统计标准适用范围,提出消除数据差异的方案

前已述及,在现有金融统计框架下,我国金融统计数据分散较分散,数据的共享存在一定障碍,不利于宏观审慎监管作用的发挥。比如,中国人民银行和银监会统计了银行投向证券资产管理的表外资产,证监会也统计了证券行业的资管数据,但由于标准均不统一,数据无法实现关联和穿透性分析,无法服务于金融业综合统计和系统风险防范。因此,要加快建立覆盖全金融业的综合统计指标体系,扩大金融统计标准适用范围,形成覆盖范围较广的元数据标准。

除要扩大金融统计标准适用范围外,还要消除不同部门间的数据差异。从国际情况看,欧美等国充分意识到一套全面及时的数据统计标准对于经济金融市场稳定的重要性后,发布了《消除数据差异提案》,提出了统一整合全球经济金融数据统计方式的20条建议。就现实情况而言,我国差异数据按统一标准重新采集基本是不可能的,所以急需研究出消除数据差异的方案,使标准和口径不同的数据可以共享。

(二) 建立部门间长效合作机制,推动金融业综合统计立法

从全国情况看,金融业综合统计推进过程中阻力较多、难度较大,仅依靠中国人民银行推动是远远不够的。金融业综合统计应得到政府的支持,由国务院组织协调各金融监管部门以及其他相关部门,如银监会、保监会、证监会、国家统计局、商务部等,探讨金融业综合统计报送机制以及数据指标范围、报送时间和报送方式等,形成长效的合作机制共同推进综合统计体系建立。

金融业综合统计覆盖银行、保险、证券、典当、担保、第三方支付和住房公积金等多个领域,但由于中国人民银行统计部门对非银行业机构缺乏有效抓手,数据报送的及时性和准确性有时难以保证。所以,推动金融业综合统计应尽快完善、修订现有法律法规,修订《中国人民银行法》,出台《金融业综合统计管理条例》,明确赋予中国人民银行负责金融业综合统计工作职责,在法律层面强化金融业开展综合统计的权威性,赋予统计部门相应的权利和义务,有效促进金融业综合统计的推进步伐。

(三) 整合系统内人才资源,助推元数据应用研究

金融统计元数据应用是一项复杂的系统工程,涉及多个部门和领域,因此人才是关键,理论是保障。一是整合人才资源。通过整合人力资源,建立人才库,不仅能在标准化应用理论研究、资料编译和应用探索上协同攻关,而且能节省内外沟通、宣传、解释等工作成本,达到事半功倍的成效,充分发挥人才合力的作用。二是注重IT^①技术开发人才的培养。金融统计元数据信息模型建立及数据共享交换,最终都需要结合IT^①技术得以实现。具体的建模、数据交换标准的实施以及数据共享和交换平台的建设是理论框架转换成实际应用的重要步骤,这些工作都需要专业的IT^①技术开发人员才能实现。三是组织推动国内金融统计标准化专题研究,依托金融学会、金融统计学会等积极开展学术交流活动,将专题研究纳入重点课题予以支持,通过多种形式营造金融统计标准化研究的氛围。特别是通过研究和交流活动发现和培养优秀人才。

参考文献:

- [1] 杜金富,盛松成,阮健弘. 金融统计标准及诠释[M]. 北京: 中国金融出版社, 2012.
- [2] 曹蓟光,王申康. 元数据管理策略的比较研究[J]. 计算机应用, 2001(2): 3-5.

^① 信息技术英文全称为 Information Technology,简称 IT。

- [3] 谭娜. 基于 XML 的空间信息元数据管理系统 [D]. 郑州: 河南大学 2004.
- [4] 张艳, 邓琳莹. 统计数据和元数据交换 (SDMX) 标准解析及应用研究 [J]. 数字通信世界 2014(4): 40 - 43.
- [5] 上海市统计局课题组, 顾敏雁. 建设上海统计元数据研究 [J]. 统计科学与实践 2014(5): 36 - 38.
- [6] 胡帆. 中国统计元数据的构成及初步诠释 [J]. 中国统计 2008(11): 6 - 8.

Research on the Construction of Metadata Information Framework under Financial Statistical Standard

Financial Statistics Metadata Research Group

(Financial Survey and Statistics Department, the People's Bank of China
Nanchang Central Branch, Nanchang, Jiangxi 330008, China)

Abstract: After the 2008 international financial crisis, the international financial organizations and major economies have continuously summed up their experiences; and actively built a financial industry comprehensive statistical system that harmonizes banks, securities and insurance. China also faces relevant problems, especially with the continuous development and innovation of the financial industry, the financial risk agglomeration, infection and diffusion are more complicated. Under the financial supervision model of "one bank and three commissions", macro-control is more difficult. In 2015, the sharp up and down of stock market in China fully exposed the problems of financial regulation. Thus, the establishment of a unified, comprehensive and shared financial industry comprehensive statistical system for banking, securities and insurance is in the trend and is imminent. Among them, metadata and standardization should bear the brunt of the basic work of the comprehensive statistical system of finance. Therefore, starting from the application status of financial statistics metadata, combining with the existing financial statistical standards and advanced technology at home and abroad, this paper build the financial statistics data information model and data sharing. The application of metadata information framework in the statistical work is evaluated by the loan instance. Finally, suggestions on how to further promote standardization and accelerate the construction of information sharing platform are put forward.

Key words: financial statistics; metadata; standard; information model; sharing platform

(责任编辑: 沈 五)