

# 《资本市场产业链图谱业务标准及数据标准研究》课题结题报告

**摘要：**本课题针对资本市场产业链图谱存在的标准体系缺失、数据粒度及数据质量等差异较大等问题开展业务标准及数据标准研究，建设完成一套适合行业数字化发展的标准体系，并基本完成申万全行业分类产业链图谱数据建设，有效验证标准体系相关成果，并开发了一个产业链图谱平台，支持数据运维及图谱应用。未来，课题组将持续完善行业产业链图谱，积极探索产业链图谱应用场景，推动数字化时代下的科技监管、风险监测、上市推广等业务发展。

**关键词：**产业链图谱；业务标准；数据标准；平台建设

《资本市场产业链图谱业务标准及数据标准研究》课题立项后，课题组按照任务书约定，投入了大量资源，开展行业标准课题研究及数据平台建设，目前已完成课题任务书约定的交付成果。课题相关情况总结报告如下。

## 一、课题背景

随着我国经济快速发展，国民经济行业及产品细分粒度越来越细，各行各业之间的行业及产品上下游关系、供应链关系等高度关联且错综复杂。行业产业链图谱旨在通过对国民经济进行行业及产品细分，梳理各行业之间的上下游层级

关系、产品供应链条关系，构建一套涵盖多个行业、涉及大量公司及海量产品的产业关系图谱，从而刻画各行各业的产业组成及关联结构。

行业产业链图谱是监管机构及市场机构开展行业产业分析研究的重要支持工具，有助于推动数字化时代下的科技监管、风险监测、上市推广等业务发展。然而，产业链图谱构建存在以下难点：第一，产业链涉及的知识信息量巨大。以申万 31 个一级行业为例，可衍生出超过 300 个三级子行业，加之不同行业产品细分、上下游关系众多，导致产业链数据较为复杂。第二，第三方机构产业链数据差异较大，各有侧重，产业覆盖范围有限，在数据质量、数据粒度、数据标准等方面存在较大差异。第三，不同于工商类图谱依靠已有的结构化或半结构化信息，数据的获取较为简单，产业链图谱依赖于行业分析师/研究员等专业人员的投入。当前，如何构建覆盖全行业、高质量的产业链图谱是众多市场机构共同关心的一项重大研究课题。

在此背景下，联合行业相关机构，汇聚行业力量，共建共享，构建服务行业的产业链图谱，提升行业的科技智能化应用水平，显得尤为重要。一是发挥行业机构的资源力量，扬各自产业研究专长，创集体协作的合力优势，提高产业链的构建效率；二是以服务行业公共需求为导向，基于行业底层公共的应用需求，开展产业链共建，市场各机构可基于共

建的产业链开发各类应用，实现智能化的应用效果；三是构建产业链行业应用新生态，作为共生、共融、共创、共赢的资本市场数字化生态圈的组成部分，助力行业数字化转型，推动行业共建共享创新研究。

然而，行业共建产业链图谱是一项复杂的系统性工程，离不开统一的标准体系。当前，产业链图谱缺乏相关统一的标准体系，导致产业链数据建设、不同来源产业链数据融合、开发应用、评价评估等方面存在诸多困难与挑战。

本课题旨在面向行业自身需求，以深交所作为牵头单位，联合行业内相关机构，基于行业底层公共的应用需求，设计适合行业数字化发展的产业链图谱标准，提升产业链图谱的覆盖面、科学性和准确性，为产业链图谱的构建和应用提供支撑，推进资本市场数字化、智能化创新应用。

## **二、业务标准**

课题组调研了证券公司、基金公司、相关数据供应商的产业链图谱建设情况、应用需求、应用场景，根据行业共性需求，构建了一套业务标准，包括行业分类体系、产业划分颗粒度、上下游关系类型、公司范围、公司所属行业等方面。

### **（一）产业链的行业分类**

基础行业分类是产业链建设的基础，考虑到申万行业分类相较国标行业分类粒度更细，对产业链图谱的研究能起到更好的支撑作用，故建议参考申万行业分类标准 2021 版来

对国民经济进行基本的行业划分。

## （二）产业链的划分依据

在基于申万行业分类对国民经济进行行业划分之后，接下来，建议基于行业细分、生产设备、销售渠道、辅助设备、生产原料、主体、辅助材料、技术服务这 8 种关联关系，将具有行业上下游关系、产品供应链关系的相关联行业及产品串联起来，形成某个具体行业产业链的架构雏形。

表 1 关系类别及含义

关系类别	含义
生产设备	上游生产制造时所需要的工业机械、电气设备等
销售渠道	购买原材料或者销售成品的环节
辅助设备	产品后续正常运转所需的辅助设备
生产原料	生产制造时所需要的原材料
主体	生产流程产出的成品、待销售的产品、服务
辅助材料	产品后续正常运转所需的辅助材料
技术服务	生产时所需要的专业技术、商业服务或为相关产品提供的后续服务
细分	行业与子行业或产品之间的大小类关系

## （三）产业链的划分颗粒度

本课题采用申万 3 级+N 级的颗粒度来逐步细分构建。首先基于申万 3 级行业分类构建产业链的顶层架构，接下来依据各个行业产业链的实际情况，参照申万行业分类，经过

N 级的不断细分，最终由行业细分不断往下扩展，直至细化到可直接被人使用和消费，独立发挥产品价值以满足人的某种需求的大类产品组（不区分生产厂家），则划分终止。例如，“螺丝”通常无法独立为人所用，常作为零部件和其它物件结合方能发挥使用价值，故“螺丝”不属于大类产品组，应作为某个大类产品组的零部件；而含有“螺丝”的“玩具汽车”可直接被人使用，“玩具汽车”即属于大类产品组。

以电力行业产业链为例，表 2 简要阐明了电力行业产业链划分的颗粒度。可以看到电力行业产业链采取了申万 3 级+1 级的模式来逐步细化构建，最后一层是细化到了产品/服务层级（此处为了举例方便，仅细化到第 4 级，实际中第 4 级还可进一步往下细分）。

表 2 电力行业产业链划分示例

申万 1 级行业	申万 2 级行业	申万 3 级行业	产品/服务名称
公用事业	电力	火电	电力工程
公用事业	电力	火电	火电机组及相关设备
公用事业	电力	火电	核电机组及相关设备
公用事业	电力	热电	电站管道
公用事业	电力	水电	水利工程
公用事业	电力	水电	水电站自动化
公用事业	电力	新能源发电	发电机
公用事业	电力	新能源发电	风电机械

公用事业	电力	新能源发电	生物质燃料
公用事业	电力	新能源发电	太阳能电站
公用事业	电力	新能源发电	余热余压节能工程
...	...	...	...

#### （四）产业链的公司范围

产业链图谱具体到一个个细分行业或产品，重点关注在该行业内或生产该公司的公司中注册资本排在前 20% 的头部企业，头部企业中通常应包含上市公司及非上市公司。无论是上市公司还是非上市公司，若能找到相关可靠及可信数据支撑的，即可纳入产业链图谱中；对于无法找到可靠及可信数据支撑的非上市公司，则暂不纳入产业链图谱中。对于纳入到图谱中的公司，重点关注其主营产品。境外公司由于数据限制等原因暂不在研究范围之内。

#### （五）产业链的公司所属行业

在具体构建产业链的过程中，对于如何判断及确定某个公司的所属行业，建议采取下述策略进行判定：

**Step1** 主要基于结构化公开信息，如财报、行研报告、行业网站、政府官网、协会网站、黄页等，借助信息抽取等技术手段，从中提取已界定清晰的公司所属行业信息、主营产品信息，从而直接判定公司所属行业。

**Step2** 对于对重点上市公司及产品所属行业，建议在第一步的基础上，借鉴行业研究员领域知识及现有产业链知识，

进行进一步的人工梳理及校核，以确保重点上市公司所属行业分类的准确性。

若找不到结构化公开信息用于判断所属行业，后续也可考虑利用非结构化公开信息，如基于公司简介、经营范围等信息，借助模型算法等技术手段进行行业分类，从而间接判定公司所属行业。

### **三、数据标准**

除了业务标准以外，数据标准同样至关重要。课题组结合业务标准以及相关调研分析，构建了一套数据标准，包括数据模型、数据源采集范围、质量评价、编码规则等。

#### **（一）数据模型**

数据模型（图 1）引入了本体概念，对领域知识进行形式化描述，由概念、关系、属性及属性值四部分组成，其中概念之间有边连接，但属性只与概念连接，属性之间不存在边连接。对照此模型，可将行业、公司、产品、上下游关系等映射到相应的节点和边，从而描绘出行业的产业结构及组成。此外，课题组重点开展底层的、公共的、基础性的产业链图谱标准构建工作，聚焦产业关系，而非公司或某类具体应用。

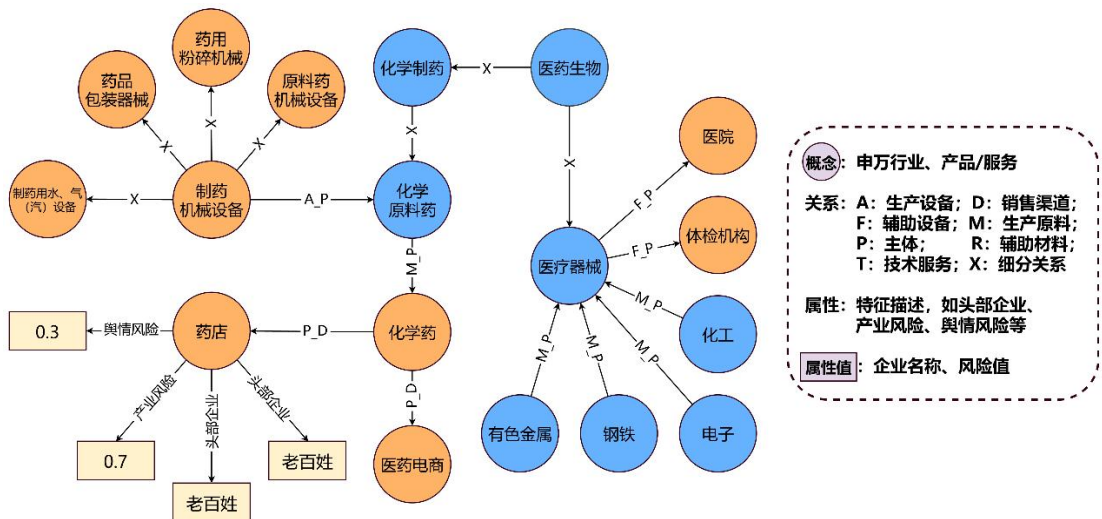


图 1 数据模型

### (1) 概念设计

本体强调的是对抽象概念的表示，而不关注具体的个体信息。与实体相比，概念是对具有相同属性事物的概括和抽象。由于产业链图谱聚焦于产业之间的关联，因此本标准重点考虑概念之间的关系。例如，相比于“格力空调”、“美的空调”等实体，“空调”是对实体的抽象，产业链图谱中的产品不会对“空调”进行品牌的细分。在概念设计方面，采取自顶向下的方法，从最顶层概念构建，逐步向下细化，形成结构良好的分类层次。

### (2) 关系设计

产业链图谱不同概念之间存在上下游关系，课题组将关系分为 8 类（详见业务标准），且后续可根据实际情况进行补充。对于头节点 A，指向尾节点 B，关系为 C，则意味着 A 为 B 提供 C。由于上游相对于下游存在关系，下游相对于上游同样存在关系，因此本课题将关系设计为双向关系。



### （3）属性设计

与公司图谱着重刻画公司之间的关联不同，产业链图谱更注重刻画产业之间的关联。因此，可将公司作为概念的属性，公司之间可通过行业、产品进行连接，从而忽略股权、债券、合作等关系。根据不同的应用场景，属性可分为头部企业、上市公司、非上市公司、舆情、产能等，后续可根据具体业务需求进行扩展。

根据数据模型，数据字典包含节点表、关系表和属性表。此外，数据的时间维度有助于全面反映企业的经营情况，更有效地从产业视角分析企业，因此在属性表中加入了时间信息。

### （二）数据源采集标准

产业链图谱的公司类型可以参考业务标准，数据维度包括企业名称、行业及产品/服务信息等。随着产业链图谱建设的不断推进，再将更多的公司囊括到图谱中。对于混业经营的公司，可参考业务标准中的行业划分依据，将企业划分到多个不同的行业。

产业链构建过程所需数据来源可参考如下范围：

- ✓ 金融数据服务终端：如万得、同花顺、东方财富等；
- ✓ 第三方数据公司或咨询公司：聚源、数库等；
- ✓ 各种官网：国家部委官网、行业协会官网、公司官网等；

- ✓ 其它相关平台或网站：黄页、财经网站、资讯平台、微博、自媒体等；
- ✓ 从实地调研或参加行业交流活动中获得；
- ✓ 从公开披露的年报、研报、财报等中获得。

### **（三）数据质量标准**

数据质量是产业链图谱得以有效应用的前提，应确保数据的完整性、准确性和合理性。具体表现如下：

- ✓ 完整性：覆盖 31 个申万一级行业分类，包含企业主要经营产品；
- ✓ 准确性：符合数据标准中的类型定义，取值合理，并能真实准确反映出各个产业之间的关系；
- ✓ 合理性：满足业务及行业分析需求，在具体业务场景中取得较好效果。

### **（四）编码标准**

为了便于产业链图谱建设，需要对行业、产品/服务进行编码，第一、二、三层行业代码与申万行业分类标准 2021 版行业分类代码保持一致，为 6 位代码；第四层代码在第三层代码的基础上添加两位，为 8 位代码。如有更多层级，依此类推，每一层都添加 2 位代码。基于公司主营业务表等数据源，对每个公司主营业务所属行业进行细分和归类，形成归类后的叶子节点。

例如，可以把申万三级的房地产开发往下细分成产业地

产、地产服务、商业地产、土地开发和住宅地产，住宅地产又可以分为酒店公寓、普通住宅和社区开发与运营。普通住宅是最细分的层级，其编码如下所示：

房地产（430000）->房地产开发（430100）->房地产开发（430101）->住宅地产（43010104）->普通住宅（4301010401）

#### **四、产业链图谱标准的成果应用**

依托产业链图谱标准成果，课题组基于申万行业分类开展了产业链图谱数据建设，并开发了产业链图谱工具平台，支撑产业链图谱的创新应用，进一步验证了标准研究成果的有效性。

##### **（一）基本完成申万全行业分类产业链图谱数据建设，标准体系得到有效验证**

为了验证课题组设计的标准的有效性，我们根据制定的业务标准及数据标准，开展产业链图谱数据建设。具体而言，由课题组各参与单位根据自身优势，选择擅长的行业建设数据，课题组针对各单位提交的数据，开展多轮次人工校核，分析问题类型，组织相关会议沟通协作，通过整合过程中遇到的相关问题，对数据标准及业务标准进行修订和完善。

目前，产业链图谱数据涉及的行业类型广泛，覆盖了农林牧渔、基础化工、钢铁、有色金属、电子、汽车、家用电器等申万行业分类下的 31 个行业。课题组已经完成了分行业建设任务，正在积极推进全量数据整合工作。

此外，产业链图谱数据涵盖国民经济所有产业分类，采用申万产业分类标准，细粒度刻画产业分类，最高达到 9 级细分（图 2）。



图 2 产业分类

## （二）开发了一个产业链图谱平台，支持数据运维及图谱应用

为更好地评估标准体系的有效性和数据建设的质量，课题组已完成产业链图谱平台开发，支持数据迭代维护及查询分析功能模块。数据运维方面，针对数据模型中的节点、关系、属性，分别开发了批量处理、新增、编辑、删除等维护操作，并通过维护日志详细记录操作类型、用户管理、变更信息、变更日期等详情。权限设置方面，充分考虑到了不同机构、不同用户、不同角色以及不同权限。在产业链图谱后续运营中，基于积分管理模块，对于使用服务时进行相应的

积分扣减，在对数据建设及维护作出贡献时会给予一定积分。

查询分析方面，包括分类查询、路径搜索、图谱查询、公司查询以及产业地图。以路径搜索为例，支持公司、产业、产品之间的关联路径探查分析，实现产业关联关系分析。例如，对于给定的两个产品，可以探查出产品之间的关联路径（图3）。课题组结合开发的产业链平台开展数据及关系验证工作，提高数据建设的准确率和召回率。

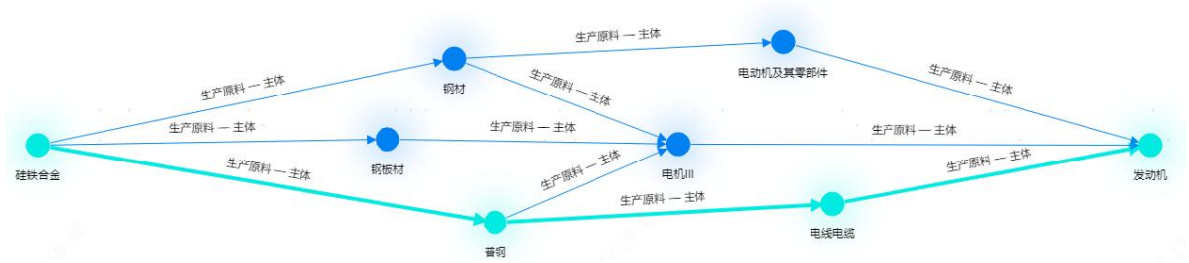


图3 关联路径探查分析

图4-图7分别是图谱查询、产业查询分析、产业地图、数据运维示例。

中文名	农林牧渔
分类级别	8层级
分类数	532
上下游关系数	4106
关联公司数	257

图4 图谱查询



此外，鉴于行业云在应用访问、资源共享、安全与效率方面具有优势，课题组将平台部署在深证通行业云，支持后续对外提供服务。

## 五、未来展望

经过课题各参与单位共同的努力，课题组在标准体系、数据建设、平台开发等方面取得了一定的成绩。下一步，课题组将做好如下几方面的工作。

一是在实践中不断完善标准体系。产业链图谱标准不是一成不变的，而是会根据业务需求及应用实践情况不断变化。课题组将持续完善标准体系，保证标准的科学性、精准性。

二是持续完善行业产业链图谱数据。课题组将依托标准成果不断完善数据，通过多渠道丰富公司主营产品信息，并借助研究所和行业研究员等专业力量优化数据质量。

三是积极探索产业链图谱应用场景。与工商、舆情、财务、业务等数据整合打通，最大程度发挥数据的价值。将产业链应用于企业研究、风险预警、智能资讯等场景，提高业务创新能力。

课题负责人：曾海泉 深圳证券交易所 副总监

课题成员：杨振新 深圳证券交易所 助理经理

余信彪 深圳证券信息有限公司 副总监

赵剑 深圳证券信息有限公司 副总监