

# 数据产品及其流通监管体系研究

熊贇<sup>1,2</sup>, 朱扬勇<sup>1,2,3</sup>

1. 复旦大学计算机科学技术学院, 上海 200438;
2. 上海市数据科学重点实验室, 上海 200438;
3. 上海数据研究院, 上海 200003

## 摘要

有序推进数据流通和有效监管数据市场的统筹部署是一项重要课题。针对数据流通标的的辨识度差、数据负外部性处理难等问题,从计量、可追溯、安全和监管等角度出发,界定数据流通标的的形态,提出可辨识的盒装数据产品流通标的;基于盒装数据产品,明确数据产品流通监管体系涵盖的内容,并提出可实操的数据产品监管体系。该体系在设计、生产、流通各阶段融入监管指标,实现有效识别和管理流通风险。

## 关键词

数据流通; 数据产品; 流通监管; 数据市场

中图分类号: TP399

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025042

## *Research on data product and their circulation regulatory framework*

XIONG Yun<sup>1,2</sup>, ZHU Yangyong<sup>1,2,3</sup>

1. School of Computer Science, Fudan University, Shanghai 200438, China
2. Shanghai Key Laboratory of Data Science, Shanghai 200438, China
3. Shanghai Data Research Institute, Shanghai 200003, China

## *Abstract*

The orderly advancement of data circulation and the effective regulation of the data market through coordinated deployment are the important issue. Addressing challenges such as the low identifiability of data circulation targets and the difficulties in managing the negative externalities of data, this paper defined the forms of data circulation targets from the perspectives of measurement, transmission, and security, proposing identifiable boxed data products as circulation targets. Furthermore, based on these boxed data products, the contents involved in the regulatory system for data product circulation were clarified and a practical regulatory framework for data products was introduced. This framework integrates regulatory indicators at each stage of design, production, and circulation, enabling rapid and effective identification and management of circulation risks.

## *Key words*

data circulation, data product, circulation regulatory, data market

## 0 引言

数据是一种关键生产要素，促进数据流通、释放数据价值是构筑国家竞争新优势的战略性举措。近年来，我国致力于部署行业、地方、区域性的数据中心和市场化运营的数据交易所（中心）建设，以支撑数据的流通。探索数据市场制度设计和规范数据市场制度运行的挑战在于明确哪些数据可流通，以及如何构建有效的监管体系，且其直接关系到数据市场的可持续健康发展。然而，数据流通标的辨识度差、数据负外部性处理难，导致切实可操作的数据产品流通监管体系尚不明确。因此，设计既能够便利流通又便于监管实操的数据产品，形成既能够快速有效识别和管理流通风险，又具有可操作性的监管方法体系，成为数据流通领域研究的重点难点。

本文提出一种贯穿数据产品设计和流通监管的数据产品监督管理体系，其涵盖监管方法、确立监管角色、明确执行流程和处置措施。该体系以数据产品标准建设为基础，在设计、生产、流通各阶段融入监管指标，实现对数据外部性、数据流通标的可标辨识度及数据产品质量等的把控，大幅提升流通监管的实操性，为数据产品高效流通奠定坚实基础。

## 1 研究现状

### 1.1 数据产品设计

我国在数据要素市场建设方面，政策布局逐渐细化，数据产品的研究和实践正在快速发展。然而，当前数据产品的界定<sup>[1-8]</sup>尚未形成统一的标准。Pei<sup>[6]</sup>将数

据产品定义为从数据集派生出的用于产品和服务的数据集。朱扬勇等<sup>[7]</sup>指出数据再生产、再再生产的结果都可被看作数据产品。叶雅珍等<sup>[8]</sup>指出作为市场流通产品，数据产品应满足一定的标准规范要求，可以独立使用、流通和监管。黄丽华等<sup>[9]</sup>从交易成本理论、市场框架和交易模式等视角分析了数据产品流通过程。然而，数据流通标的没有统一的计量单元，导致数据流通标的可辨识度差和监管难。关于数据流通标的基本形态，学术界和产业界目前提出“盒装数据”“数据元件”“数据件”等概念。其中，熊贇等<sup>[10]</sup>从技术形态上设计了数据盒模型，该模型具有数据产品形态雏形；叶雅珍等<sup>[8]</sup>设计了基于数据盒模型的数据产品，形成可计量的数据产品形态；陆志鹏<sup>[11]</sup>提出连接数据供需两端的“中间态”的数据元件概念，将数据元件定义为根据需要由若干相关字段形成的数据集，或由数据的关联字段通过建模形成的数据特征。孙凝晖等<sup>[12]</sup>提出数据件概念，并指出数据件在价值评估、动态支持和能力体系上有待进一步研究。

综上所述，盒装数据是面向数据产品形态而设计的；数据元件是面向数据产品生产的中间品而设计的，数据开发主体获取数据元件后对其进一步开发利用，打造数据产品；数据件是面向底层的数据形态。上述3种概念为探索可监管、可计量的数据产品形态奠定了基础。但为了进一步确保数据高效流通，人们亟须在此基础上发展一套既标准化又易于支持流通监管的数据产品设计方法。

### 1.2 数据盒模型设计与实现

围绕如何既开放共享数据又保持数据

价值的稀缺性，朱扬勇等<sup>[13]</sup>突破数据自治封闭的存储建模局限，提出将数据盒模型作为开放共享数据的组织模型，通过将数据装载到数据盒中，赋予数据流通和价值实现的特性。熊贇等<sup>[10]</sup>提出数据盒实现相关技术，包括数据盒建模、数据权益保护、数据使用行为管控等，并给出包含数据盒创建、灌装、部署、加载使用等的的数据盒运作系统的实施步骤。吴毅坚等<sup>[14]</sup>研制以数据盒为基本数据组织和运算的开放共享数据开发和运行系统，实现数据盒的获取、开发、调试以及部署和运行。陈德华等<sup>[15]</sup>提出数据盒应用系统，并在医疗领域的数据开放共享应用上进行验证。上述研究聚焦提高数据开放流通需求，为基于数据盒模型的数据产品形态设计奠定了可实施的技术基础。但上述研究尚未充分考虑开放流通过程中的数据产品监管需求。面向数据产品流通监管的挑战，人们需要进一步探索基于数据盒的数据产品及其流通监管体系的设计方法。

### 1.3 数据产品流通监管

数据质量作为数据产品的核心属性，直接影响数据产品流通监管体系的设计维度与实施效能。数据产品流通监管已经在完整性、准确性和时效性等维度提出了要求，构成了数据质量控制的标准。《关于构建数据基础制度更好发挥数据要素作用的意见》（以下简称“数据二十条”）多次提及数据质量相关内容，各地区也相继发布相关政策文件。ISO 8000 数据质量系列国际标准描述了数据质量和认证机制，基于ISO 8000 系列标准，中国标准化研究院制定了5项GB/T 42381《数据质量》国家标准，提供了管理、测量和提高数据质量的方法。上海数据交易所发布了《数据交易

第3部分：数据产品质量评估规范》（征求意见稿）。Huang等<sup>[16]</sup>梳理了国内外数据质量评估现状与数据质量检核常用手段。王宁静等<sup>[17]</sup>提出金融科技数据产品质量评估模型。Zhao等<sup>[18]</sup>提出针对视觉语言任务大规模基准数据集的质量评估方法。郭鑫等<sup>[19]</sup>提出用于政府开放数据的质量评估方法。

然而，随着数据要素市场的发展，数据的质量需求从“自用需求”转变为“他用需求”“监管需求”。数据产品监管应当包含数据本身的质量以及数据作为产品的质量监管。因此，满足“监管需求”的数据质量维度成为研究的新挑战<sup>[20]</sup>。数据的外部性增加了数据产品监管的复杂性。同时，数据多方融合使数据产品监管的目标变得复杂且不明确。这些因素导致目前缺乏可操作性的监管方法。

## 2 面向监管需求的数据产品设计

结合数据产品流通可辨识度、便利性、时效性和可监管等要求，本节介绍兼顾流通效率及其监管需求的标准化数据产品设计方法。

### 2.1 数据产品形态

数据的经济活动是关于数据产品的生产、分配、交换或消费等的活动。数据生产将现实事物信息化，从而在网络空间形成数据；或者直接通过计算机等输入设备将数据输入网络空间。数据生产得到的是数据资源或者数据的初级产品，可以直接供最终用户使用，也可以作为原材料（初级产品）进行再生产，即根据已有的数据，运用数据技术，形成更高级别的数据产品。

数据再生产包括数据汇集、数据清洁、数据可视化、数据分析、人工智能等<sup>[7]</sup>。数据产品的原材料、半成品、成品都是数据，并且数据的再生产、再再生产的结果也还是数据，这些都可以被看作数据产品<sup>[7]</sup>。因此，面向流通监管的数据产品设计涵盖原生数据、数据再生产以及数据再再生产之后的结果。

数据产品流通要求涵盖产品质量、生产资质、产品标识、安全标注和合规流通等方面。根据上述要求，借鉴具有标准形态的图书产品，本文从计量、可追溯、安全和监管的角度出发，界定流通标的形态，提出符合市场需求的、可计量的、标准化的盒装数据产品形态，为监管提供依据。在“内部组织”加“外部形态”相结合的数据盒形态基础上<sup>[8,10]</sup>强化监管组件，数据产品形态相关维度如图1所示。具体地，本文从计量的角度，定义数据盒的类型、结构和规格等；从需求的角度，结合数据存储技术，定义多模态盒内数据内容组织方式，支持定制化设计；从可追溯的角度，定义数据盒标识系统及其追溯机制；从安全的角度，定义数据安全保护策

略和使用行为控制策略；从监管的角度，定义盒外包装以说明生产技术、数据内容质量和数据使用规范等，包括盒装数据产品的生产资质证书、产品登记证书、合规证书、质量证书、产品说明书和外部性说明书等盒外包装规范。

## 2.2 数据产品设计

数据产品设计包括数据产品设计、数据产品规范设计、数据产品安全设计和数据产品标准设计等主要部分，数据产品设计的技术路线如图2所示。

具体地，数据产品设计包括如下内容

- 数据产品设计面向需求的数据集划分策略，针对原生态数据以及数据再生产形成的处理后数据对象，通过多粒度的跨模态多源数据融合建模方法，将数据组织为不同规模、尺度、模态和内容的定制化数据盒产品，形成可定制的数据产品；针对数据再生产形成的模型等衍生类对象，设置模型的规模、架构和约束等，形成数据产品。

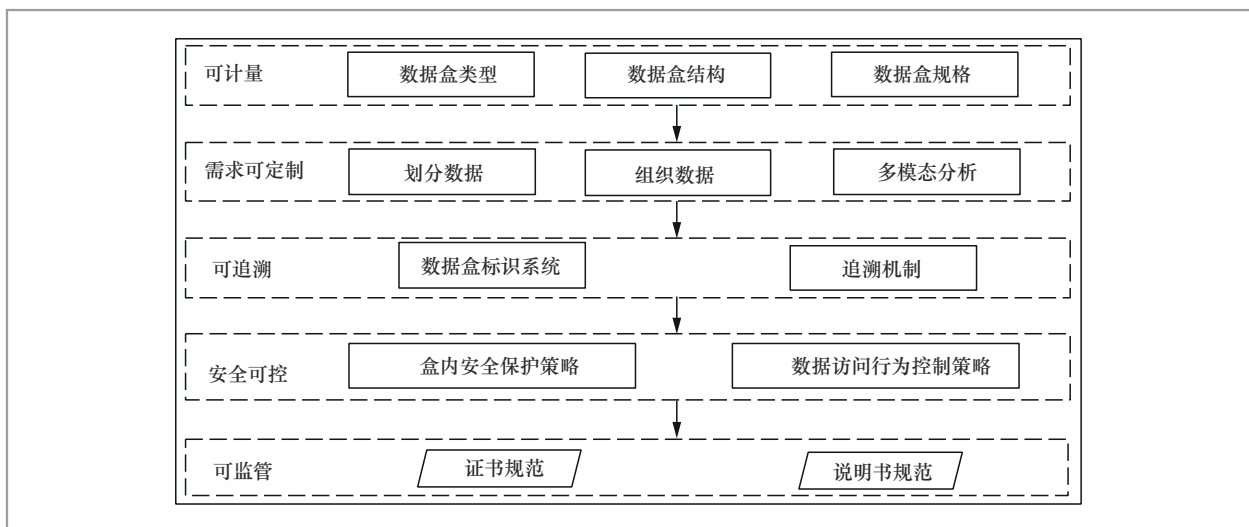


图1 数据产品形态相关维度

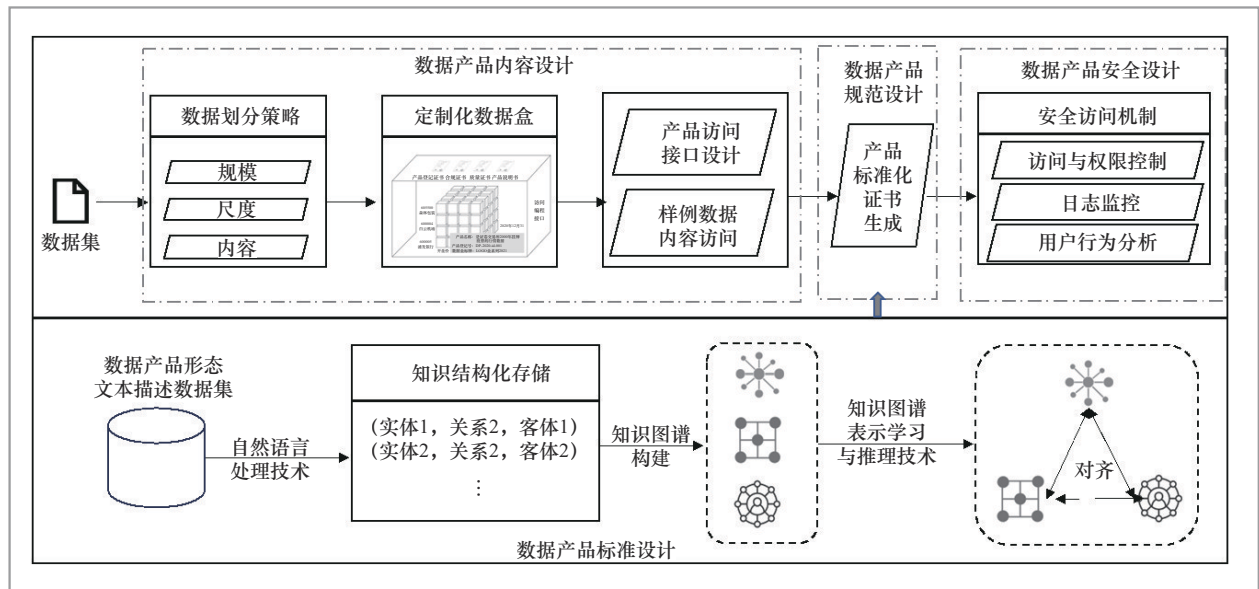


图2 数据产品设计的技術路线

• 数据产品内容设计高效可用的数据产品访问接口，在数据不可见的情况下支持对数据产品内部数据或再生产形成的模型的使用。

• 数据产品内容设计可机读的盒装样例数据内容抽取方法，形成可抽样检测的数据产品。

数据产品规范设计采用基于人工智能内容生成的方法实现数据产品标准化证书生成。

数据产品安全设计包括设计盒装数据产品安全访问机制，在数据盒中封装数据防泄露和安全计算机制，采用访问行为监控技术与权限控制技术实现数据产品的安全访问与使用，通过对访问日志和用户行为模式的定期分析进行安全审计和监控。

数据产品标准设计包括制定数据产品生产技术规范、产品内容质量规范和产品使用规范等；针对跨平台、跨机构流通导致的数据产品证书及其规范描述冲突问题，构建规范术语图谱，采用深度学习模型实

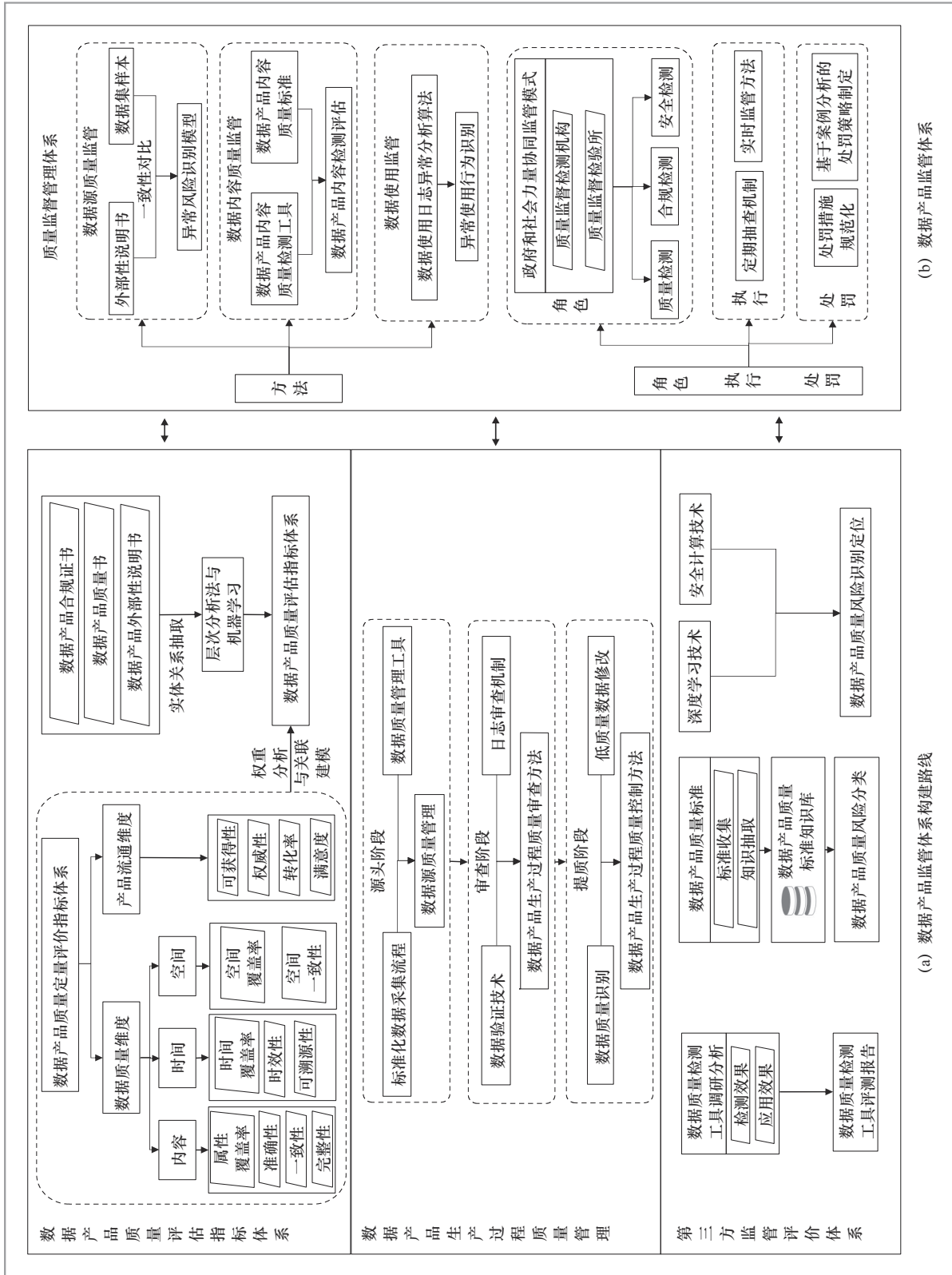
现规范的一致化，构建数据产品标准库，为形成数据产品标准提供支持。

### 3 数据产品流通监管体系

依据盒装数据产品形态，本文明确数据产品流通监管体系涵盖的内容，梳理面向监管需求的数据质量评价维度，给出数据产品生产过程质量管理方法，提出数据产品流通监管体系及其构建路线，如图3所示。

#### 3.1 数据产品流通监管内容

数据流通涉及数据采集、存储、处理和交易等多环节。数据产品是数据生产与数据再生产、数据再再生产的结果，因此，数据产品流通监管覆盖数据产品生产和流通环节，且每个环节都有相应的监管方法和策略。数据产品流通监管是从数据产品本身、数据产品生产过程到数据产品流通过程的全链条监管，具体包括数据产品质



(a) 数据产品监管体系构建路线

(b) 数据产品监管体系

图3 数据产品流通监管体系及其构建路线

量评价、数据产品生产过程质量管理、数据产品权属及其证书管理、数据产品流通成本和价格管理、数据产品流通过程和场所管理、第三方监管评价体系等方面。

### 3.2 数据产品质量评价指标体系

数据产品本身的监管主要指数据产品质量的监管，其涉及数据产品的内容质量和满足需求的质量。因此，构建数据产品质量评价指标体系应考虑数据产品的双重属性，即“数据”和“产品”。在“数据”方面，数据产品形态为数据盒，流通的数据或再生产形成的模型等衍生物对象被封装在数据盒中，其中，面向原生态数据或数据再生产形成的数据的质量评价指标维度涉及数据内容维度、时间维度、空间维度。

- 数据内容维度：属性的覆盖度、准确性、一致性和完整性等。
- 时间维度：时间覆盖率、时效性和可溯源性等。
- 空间维度：空间覆盖率和空间一致性指标等。

而面向数据再生产形成的模型等衍生物对象，质量评价指标主要考虑模型的描述（规模、架构和约束等）以及如代码等相关支持内容。在“产品”方面，面向数据产品流通需求，质量评价指标主要考虑与数据产品来源合规相关的可获得性、权威性，以及转化率和欢迎度等维度指标。

明确上述维度指标后，本文依据数据产品合规证书、质量证书和外部性说明书等，采用自然语言处理和实体关系抽取等技术，提取上述维度指标，结合机器学习方法对提取的指标进行建模，识别其权重及其之间的关系，形成数据产品质量评价

指标体系。具体地，本文围绕内容、时间和空间质量维度，结合产品流通维度，采用实体关系抽取技术提取盒装数据产品的合规证书、质量证书和外部性说明书等中的质量指标，依据数据质量国际/国家数据质量相关标准，构建涵盖数据内容的有效性、完整性、时效性、数据源的合规性、数据访问的安全性和经济性等指标的多层级、多维度的数据产品质量评价指标体系。

### 3.3 数据产品生产过程质量管理

数据产品生产过程涵盖生产来源、过程和出厂等环节，因此，本文设计“源头、审查、提质”三阶段数据产品生产过程质量管理策略。在源头阶段，本文设计标准化数据采集流程，结合数据质量管理工具，管理控制数据源的质量。在审查阶段，本文运用数据验证技术，配合定期的日志审查机制，管理控制数据产品生产过程的质量。在提质阶段，本文针对形态多样的数据产品，包括图像、文本、表数据等，开发基于多模态深度学习的数据产品质量控制方法。此外，本文构建模型训练数据库，利用多模态大模型识别低质量数据；通过修改或删除低质量数据样本降低数据产品质量风险，确保出厂的数据产品质量达到产品设计的质量标准。

### 3.4 数据产品权属及其证书管理

数据经过生产或再生产形成数据产品后进入流通环节，需要对数据产品权属进行监管。数据盒中封装数据盒标识及其追溯机制，同时封装盒装数据的使用行为控制策略，其将数据产品的权属进行了封装和保护，即采用数据盒标识与数据使用者进行映射，并将其映射记录通过存储上链的方法实现权属追溯，而行为控制策略能

够明确盒装数据的使用权属。上述对数据产品权属监管提供了方法支持。

数据盒证书用于说明生产技术、数据内容质量和数据使用规范等,包括盒装数据产品的生产资质证书、产品登记证书、合规证书、质量证书、产品说明书和外部性说明书等,这些证书直接体现数据产品质量。因此,对数据产品流通监管有必要对上述证书进行管理,包括证书内容本身、证书存储、上链、版本分析和更新维护等方面的监管。

### 3.5 数据产品流通成本和价格管理

数据盒是一种可计量的数据存储模型,为数据产品流通过程中的数据产品定价提供了可计量的基础和依据。本文通过对数据产品流通历史数据进行分析,结合经济学理论和价格形成机制,形成动态数据产品流通定价模型,实现数据产品流通成本计算和价格管理。

### 3.6 数据产品流通过程和场所管理

本文根据数据产品内容和需求场景,对公共数据产品流通和社会数据产品流通分别进行流通过程和场所管理方法设计;分析数据产品流通各环节角色,以场景需求为驱动构建数据产品供应链,通过智能推荐和优化供应链环节参与者,并且设计分配模型,通过数据产品相关主体收益分配和激励机制强化数据产品流通过程管理。

### 3.7 第三方监管评价体系

本文考虑到数据产品流通环节的复杂性、技术专业性以及多方主体利益的平衡需求,认为数据产品流通监管有必要引入第三方监管机制,实现政府管规则、市场

提效率、第三方保公平的协同治理。具体地,本文利用知识抽取等技术构建数据产品质量标准知识库,为第三方质量检测评价机构准备标准化指标库基础,确立第三方数据产品质量评测机构的人员、技术等资格标准;分析数据质量评测检测方法与工具,形成评估分析报告,确立第三方质量评测机构的评测工具和方法的资格标准。此外,本文采用深度学习和安全计算技术,实现数据产品内部质量风险识别和定位,提升第三方数据产品质量检测能力;建立第三方数据产品质量评测机构的评价体系,以提高数据产品质量评测效能。

### 3.8 数据产品流通监管体系设计

数据产品流通监管并非限制数据价值释放,而是充分释放数据价值。在体系设计上实现监管可控与高效利用的平衡,涉及在监管方法、监管角色、执行和处理等多层次上的设计和优化,为监管提供可操作、可量化评估的方法手段。本文围绕数据产品流通监管涵盖的内容,从监管方法、监管角色、监管执行和处理3个方面对数据产品流通监管体系进行设计。

#### (1) 监管方法

本文从数据内容质量监管的角度,基于产品质量证书,依据数据内容的有效性、完整性等质量指标,对内容质量进行检测评估;从数据源监管的角度,基于产品合规证书和外部性说明书,依据数据源的合规性等质量指标,对产品合规性进行检测评估;从数据使用监管的角度,基于产品说明书,依据数据访问的安全性和经济性等指标,对产品可访问的要求进行检测评估。本文综合上述方法,设计综合的数据产品质量检测工具。

### (2) 监管角色

本文建立政府和社会力量协同监管模式, 设立国家数据产品质量监督检验检测机构, 发挥行业协会、联盟等社会性组织的作用, 建立数据产品质量监督检验所(站)。

### (3) 监管执行和处理

本文构建数据产品质量监督方案数据库, 通过定期抽查和评估以及及时发现问题并采取相应措施, 形成事前、事中、事后的数据流通监管运行机制。

## 4 结束语

本文提出了一种可辨识、可计量、可监管的盒装数据产品形态, 明确了数据产品流通监管体系内容和范围, 并基于此提出政府和社会力量协同监管模式, 形成“方法-角色-执行-处理”全流程的、可实操的数据产品流通监管体系。其将为数据流通监管工具及平台的研制、规则规范的制定和落实提供依据。

## 参考文献:

- [1] LOUKIDES M K. What is data science? [M]. Beijing: O'Reilly Media, 2011.
- [2] HAZEN B T, BOONE C A, EZELL J D, et al. Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications[J]. International Journal of Production Economics, 2014, 154: 72-80.
- [3] HUANG G Y, HE J, CHI C H, et al. A data as a product model for future consumption of big stream data in clouds[C]//Proceedings of the 2015 IEEE International Conference on Services Computing. Piscataway: IEEE Press, 2015: 256-263.
- [4] BENGFORT B, KIM J. Data analytics with Hadoop: an introduction for data scientists[M]. Beijing: O'Reilly Media, 2016.
- [5] CAO L B. Data science[J]. Communications of the ACM, 2017, 60(8): 59-68.
- [6] PEI J. Data pricing: from economics to data science[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 3553-3554.
- [7] 朱扬勇, 熊贇. 数据的经济活动及其所需要的权利[J]. 大数据, 2020, 6(6): 140-150.  
ZHU Y Y, XIONG Y. The required authorization to the data-centric economic activities[J]. Big Data Research, 2020, 6(6): 140-150.
- [8] 叶雅珍, 朱扬勇. 盒装数据: 一种基于数据盒的数据产品形态[J]. 大数据, 2022, 8(3): 15-25.  
YE Y Z, ZHU Y Y. BoxedData: a data product form based on databox[J]. Big Data Research, 2022, 8(3): 15-25.
- [9] 黄丽华, 窦一凡, 郭梦珂, 等. 数据流通市场中数据产品的特性及其交易模式[J]. 大数据, 2022, 8(3): 3-14.  
HUANG L H, DOU Y F, GUO M K, et al. Features and transaction modes of data products in data markets[J]. Big Data Research, 2022, 8(3): 3-14.
- [10] 熊贇, 朱扬勇. 面向数据自治开放的数据盒模型[J]. 大数据, 2018, 4(2): 21-30.  
XIONG Y, ZHU Y Y. Data box: a novel data model for self-governing openness of data[J]. Big Data Research, 2018, 4(2): 21-30.
- [11] 陆志鹏. 创新数据治理路径 激活数据要素潜能[J]. 经济, 2021(6): 117-119.  
LU Z P. Innovating data governance path and activating the potential of data elements[J]. Economy, 2021(6): 117-119.
- [12] 孙凝晖, 郭嘉丰. 数据件: 一种数据要素标准化抽象[J]. 中国计算机学会通讯, 2024, 20(20): 48-54.  
SUN N H, GUO J F. Data item: a standardized abstraction of data elements[J].

- Communications of CCF, 2024, 20(20): 48-54.
- [13] 朱扬勇, 熊贇, 廖志成, 等. 数据自治开放模式[J]. 大数据, 2018, 4(2): 3-13.  
ZHU Y Y, XIONG Y, LIAO Z C, et al. Self-governing openness of data[J]. Big Data Research, 2018, 4(2): 3-13.
- [14] 吴毅坚, 陈士壮, 葛佳丽, 等. 数据自治开放的软件开发和运行环境[J]. 大数据, 2018, 4(2): 31-41.  
WU Y J, CHEN S Z, GE J L, et al. Software development and runtime environment for self-governing openness of data[J]. Big Data Research, 2018, 4(2): 31-41.
- [15] 陈德华, 潘乔, 王梅, 等. 数据自治开放应用平台设计与实践[J]. 大数据, 2018, 4(2): 63-71.  
CHEN D H, PAN Q, WANG M, et al. Platform design and practice for self-governing openness of data application[J]. Big Data Research, 2018, 4(2): 63-71.
- [16] HUANG Q Q, ZHAO Z, LIU Z Y. Comprehensive management system and technical framework of data quality in the data circulation transaction scenario[J]. Data Analysis and Knowledge Discovery, 2022, 6(1): 22-34.
- [17] 王宁静, 郭树行. 面向数据要素可信流通的金融科技数据产品评估研究[J]. 互联网周刊, 2022(20): 22-25.  
WANG N J, GUO S X. Research on financial technology data product evaluation for trusted circulation of data elements[J]. China Internet Week, 2022(20): 22-25.
- [18] ZHAO R B, XIE Z W, ZHUANG Y P, et al. Automated quality evaluation of large-scale benchmark datasets for vision-language tasks[J]. International Journal of Neural Systems, 2024, 34(3): 2450009.
- [19] 郭鑫, 聂磊, 王继民, 等. 面向数据检索的我国政府开放数据质量评估方法研究[J]. 数据分析与知识发现, 2024: 2024.0226.  
GUO X, NIE L, WANG J M, et al. Research on China's government open data quality assessment methods for data retrieval[J]. Data Analysis and Knowledge Discovery, 2024: 2024.0226.
- [20] 蔡莉, 朱扬勇. 从数据质量到数据产品质量[J]. 大数据, 2022, 8(3): 26-39.  
CAI L, ZHU Y Y. From data quality to data products quality[J]. Big Data Research, 2022, 8(3): 26-39.

#### 作者简介



熊贇 (1980-), 女, 博士, 复旦大学计算机科学技术学院教授, 上海市数据科学重点实验室副主任, 主要研究方向为数据科学和数字经济。



朱扬勇 (1963-), 男, 博士, 复旦大学计算机科学技术学院教授, 复旦大学数据产业研究中心副主任, 上海数据研究院学术副院长, 主要研究方向为数据科学和数字经济。

收稿日期: 2025-02-25

通信作者: 朱扬勇, yyzhu@fudan.edu.cn