

数据领域标准术语冲突现状分析

黄科满¹, 张鼎镇¹, 朱尧¹, 刘艳芳^{2,3}, 李洋^{2,3}, 杜小勇^{1,4}

1. 中国人民大学, 北京 100872;
2. 中国民航信息网络股份有限公司, 北京 101318;
3. 民航旅客服务智能化技术应用重点实验室, 北京 101318;
4. 数据工程与知识工程教育部重点实验室, 北京 100086

摘要

当前我国数据领域标准化工作已进入体系化推进的新阶段, 术语冲突问题凸显。为了系统识别与治理术语冲突, 从术语定义出发, 构建包含10种冲突类型的术语冲突系统化分类体系与检测方法, 并以全国标准信息公共服务平台收录的612项数据相关标准为数据来源, 构造了包含24 575个冲突术语对的数据集。在此基础上, 从类型分布、时序演化、标准类别、关键术语、核心标准、起草主体差异等多个维度展开系统性分析。分析表明: 层级冲突、引用冲突和范围冲突构成冲突主体; 冲突数量在标准化工作快速推进期出现增幅, 标准的持续推出并未带来术语一致性的提升; 数据技术与数据资源领域标准是当前术语冲突集中区域; “元数据”是存在冲突最明显的术语, 基础类标准与应用类标准的冲突类型差异显著, 交叉领域标准冲突类型分布均衡; 政府与事业单位层级冲突占比较高, 企业引用冲突占比最高。基于上述分析, 从概念体系构建、全生命周期术语协调、多主体协同治理、引用管理规范4个维度提出了针对性工作建议, 为我国数据领域术语工作的完善提供实证支撑与决策参考。

关键词

数据术语; 术语冲突分析; 标准分析; 数据标准; 标准化治理

中图分类号: TP3

文献标志码: A

doi: 10.11959/j.issn.2096-0271.BDR26125

Analysis of the current status of terminology conflicts in data standards

Huang Keman¹, Zhang Dingzhen¹, Zhu Yao¹, Liu Yanfang^{2,3}, Li Yang^{2,3}, Du Xiaoyong^{1,4}

1. Renmin University of China, Beijing 100872, China
2. Travelsky Technology Limited, Beijing 101318, China
3. Key Laboratory of Intelligent Application Technology for Civil Aviation Passenger Service, Beijing 101318, China
4. Key Laboratory of Data Engineering and Knowledge Engineering, Beijing 100086, China

Abstract

Currently, China's standardization efforts in the data domain have entered a new stage of systematic advancement. To systematically identify and manage conflicts in term definitions, this study starts from term definitions and constructs a systematic classification framework for term conflicts comprising ten conflict types, along with a detection process. Using 612 data-related standards collected from the National Standard Information Public Service Platform as the data source, a dataset containing 24,575 conflicting term pairs was established. Based on this dataset, a systematic analysis was conducted from multiple dimensions, including type distribution, temporal evolution, standard categories, key

terms, core standards, and differences among drafting entities. The analysis reveals that hierarchical conflicts, citation conflicts, and scope conflicts constitute the main body of term conflicts; the number of conflicts increased during the period of rapid advancement in standardization, indicating that the expansion of the number of standards does not naturally lead to improved terminological consistency; standards in the fields of data technology and data resources are the most conflict-intensive areas; “metadata” is the term with the most pronounced conflicts, there are significant differences in conflict types between foundational standards and application standards, while the distribution of conflict types in cross-domain standards is more balanced; government agencies and public institutions exhibit a higher proportion of hierarchical conflicts, while enterprises show the highest proportion of citation conflicts. Based on the above analysis, this paper proposes targeted recommendations from four dimensions: conceptual system construction, full-lifecycle terminology coordination, multi-stakeholder collaborative governance, and citation management standardization, providing empirical support and decision-making references for improving terminology in China’s data domain.

Key words

data terminology, terminology conflict analysis, standards analysis, data standard, standardization governance

0 引言

随着数字经济的深入发展，数据已被确立为继土地、劳动力、资本、技术之后的第五大关键生产要素^[1]。在此宏观背景下，数据标准化工作已提升至国家战略层面，成为推动数据要素市场化配置、促进数据高效流通与共享、充分释放数据价值的基础性与先导性工程^[2]。术语构成数据标准化的基石，其统一性与规范性直接关乎核心概念的理解一致性、跨领域协作沟通的有效性，以及数据要素在实际应用中的互操作性。然而，当前数据领域在术语使用方面，仍普遍存在概念边界模糊、定义表述不一、引用来源混杂等问题。不同标准文本、实践主体之间对同一术语的理解时常存在差异，这制约了数据标准体系的内部协调性与外部应用实效^[3]。因此，系统性地识别数据领域术语定义冲突的具体类型、分布特征、演变规律及深层成因，具有重要的理论价值与紧迫的现实意义。

为应对上述挑战，本研究以全国标准信息公共服务平台收录的612项数据相关

标准为实证分析的数据来源，采用大模型结构化信息提取方法，从中自动化抽取了28 133条术语条目；进而，通过同名异义与异名同义的匹配原则，构建了包含48 228个术语对的基础数据集；依据GB/T 10112-2019《术语工作 原则与方法》的规范性指导，构建了一套包含10种冲突类型的术语冲突系统化分类体系与检测方法。在此基础上，本研究对经筛选后确认存在的24 575个冲突术语对，从类型分布、时序演化、标准类别、关键术语、标准引用网络、起草主体差异等多个视角开展多维度、系统性分析，揭示术语定义冲突呈现的宏观态势，剖析冲突产生的原因；最终，为我国数据领域术语工作的完善提供实证依据与决策参考，以助力构建一个统一、规范、协调的数据术语体系。

1 数据领域标准术语工作背景与挑战

1.1 研究背景

近年来，伴随数据要素市场的培育与国家数字化战略的深入实施，我国数据领

域标准化工作已迈入体系化推进的新阶段。

从政策层面看，国家高度重视数据要素市场培育和数据基础制度建设^[4]。2022年12月，中共中央 国务院印发《关于构建数据基础制度更好发挥数据要素作用的意见》（简称“数据二十条”），提出从流通规则、交易市场、服务生态等方面加强数据流通交易顶层设计，建立数据流通准入标准规则，探索开展数据质量标准化体系建设；统筹优化全国数据交易场所规划布局，出台数据交易场所管理办法，构建多层次市场交易体系；培育数据商和第三方专业服务机构两类主体。

从组织层面看，全国信息技术标准化技术委员会大数据标准工作组于2014年12月在北京成立，由梅宏院士任组长，秘书处设在中国电子技术标准化研究院，统筹开展我国大数据领域标准体系建设、国家标准制修订与国际标准化推进工作。2024年10月，全国数据标准化技术委员会正式成立，由国家数据局负责日常管理和业务指导，主要负责数据资源、数据技术、数据流通、智慧城市、数字化转型等基础通用标准制订工作，以及支撑数据流通利用的数据基础设施标准、保障数据流通利用的安全标准等领域的国家标准制修订工作。全国数据标准化技术委员会的成立，为数据领域标准化工作提供了专门的组织支撑和工作平台。

从实践层面看，跨部门、跨行业、跨区域的数据共享与协调面临困难，数据孤岛现象普遍存在，制约了数据产业的进一步发展。数据供给不充分、数据质量参差不齐、数据标准不统一等问题突出，部分可开放利用的数据资源仍处于“休眠”状态^[5]。这些现实困境迫切需要建立健全统一、规范的数据标准体系。

总体而言，数据领域标准化工作进入

体系化推进的新阶段。其中，术语定义作为基础性、先导性的工作，直接关系到数据标准体系的根基建设，对于凝聚行业共识、消除沟通壁垒、促进数据要素流通具有重要意义。然而，由于数据领域概念体系尚在动态演进中，业界对诸多基础术语的认知尚不一致，术语冲突与使用混乱的问题仍然存在，亟须系统性的研究和规范^[6]。

1.2 面临的挑战

尽管我国数据领域术语工作已取得阶段性进展，但在实际推进过程中，仍面临若干结构性挑战。

其一，核心概念体系尚未定型。相较于成熟学科，数据领域的诸多基础概念的内涵边界与逻辑关系尚未形成稳定且公认的体系，整体上仍处于模糊状态。这具体表现为存量标准与新兴增量需求之间的结构性错位。早期数据领域标准多集中于数据处理、系统架构等技术实现层面。随着数据要素化进程加速，数据治理、流通、价值化等新实践催生了大量涉及政策与经济属性的新概念。当人们试图在既有技术框架内修补标准时，常发现原有术语体系难以有效承载这些新的概念维度。新旧术语之间的关系并非简单的增补，而是需要进行体系性的反思与重构。当前，产业界已在积极探索数据资产化等路径，但在标准层面，相应的概念共识远未形成。

其二，术语冲突的实际情况缺乏系统性研判。在术语工作中，人们对术语冲突的认知大多仍停留在零散的个案层面，对现存定义冲突的类型、分布与严重程度，尚缺乏全面、系统的摸底与诊断。当前，数据领域术语使用中究竟存在哪些具体冲突、这些冲突集中分布于哪些环节、冲突程度如何，均未有清晰答案。由于无法精

准定位问题集中的领域，标准研制工作往往依赖专家经验或个案观察，难以确保有限资源投向最亟须规范之处。例如，哪些术语最迫切需要统一，哪些冲突应优先解决。在缺乏系统性数据支撑的情况下，对这些问题的决策多基于主观判断，这制约了标准化工作的针对性与实效性。更需警惕的是，术语定义中潜藏的系统性风险尚未被充分认知。若某些关键概念的歧义长期未被识别与化解，可能在数据流通、资产确认、跨域互操作等关键环节引发误解。

其三，高效的多方协同机制尚不健全。术语的本质在于寻求最大范围内的共识，而数据领域恰恰是技术、法律、经济、管理等多重维度与多元主体交汇的复杂场域，共识凝聚面临显著的结构障碍。不同专业背景的主体对同一术语的解读存在天然侧重：技术方关注实现细节与功能边界，法律方聚焦权利归属与合规要求，经济方则强调价值属性与流通规则，管理方侧重统筹规范落地与整体治理。这导致术语定义工作陷入两难：若定义过于宽泛，则可能失去规范意义；若定义过于精细，又可能降低部分应用场景的适用性。与此同时，该领域尚未建立起成熟、高效的常态化协商解决机制。共识的凝聚往往依赖于特定项目组或少数核心专家的反复协调，这一过程不仅耗费大量时间，其产出的结果也往往因参与方代表性不足而难以获得业界广泛认可。

1.3 术语冲突研究现状

当前，术语冲突问题已引起学术界的广泛关注，围绕其成因、影响及化解路径形成了若干研究脉络。

从基础理论与通用原则层面来看，GB/T 10112-2019《术语工作 原则与方法》等标准确立了概念体系、定义撰写与

指称命名的一般规范，为各专业领域的术语标准化工作提供了方法论指导^[7]。然而，这些通用原则在数据这一新兴领域的适用性和转化路径仍有待深入检验。

在学科领域层面，国内外学者对特定领域的术语冲突展开了一系列研究。在国际上，研究数据联盟（Research Data Alliance, RDA）等组织持续关注数据管理实践中的术语不一致问题，相关研究发现一词多义（polysemy）、模糊性（vagueness）是导致混乱的最关键来源，并尝试借鉴风险评估框架对术语混淆的严重程度进行分级评估^[8]。跨域融合领域的研究同样关注术语差异对协作的阻碍作用，指出即使是“model”等常见术语，在不同学科中也存在不同的语义指称^[9]，并倡导通过构建跨学科术语库（如 Dialogue: Wiki）来促进共识的形成。

我国学者同样注意到数据领域术语工作的紧迫性。全国数据标准化技术委员会已将《数据 基础术语》列为1号国家标准项目，并于2025年10月正式立项，旨在解决数据产业发展中术语定义不统一、概念边界模糊等关键问题。在学术研究中，部分学者对“数据要素”等特定术语的内涵界定进行了理论探讨，但总体上，国内在该领域的系统性、实证性研究仍然有限。

综上所述，现有研究在术语冲突的宏观诊断与系统化解方面尚存在以下研究缺口。

其一，对数据领域术语冲突的系统性诊断十分匮乏。既有研究多集中于特定概念（如数据资产、数据治理等）的个案式辨析或整体性地呼吁加强术语规范建设，对术语冲突的类型构成、分布特征、演变规律和严重程度的实证研究较少，缺乏对大范围标准术语文本的数据驱动分析。此外，既有研究对术语定义冲突的源头缺乏清晰的归因框架与实证证据。

其二，术语冲突的分类体系不够完善。尽管术语学通用原则提供了概念协调的一般性指导，但面向数据领域术语定义实践的可操作冲突分类框架尚未建立。GB/T 10112-2019 确立了术语工作的原则与方法，但其更侧重于规范性的编纂指导，而非对既有术语定义中具体冲突类型的系统归纳与分类。针对数据领域标准文本中大量存在的定义不统一现象，目前尚缺乏一套能够精细识别和归因的冲突类型学体系，这制约了后续系统分析与治理工作的开展。

其三，缺乏对术语冲突治理机制的实证支撑。尽管国家和行业层面已意识到术语标准化的重要性并启动了具体工作，但标准制定者面对“哪些术语最迫切需要统一”“哪些冲突应优先解决”等关键决策问题时，仍缺乏充分的实证数据支撑。术语治理尚处于依靠专家经验和个案判断的阶段，亟须基于大规模数据分析的宏观态势研判来优化资源配置与完善政策干预。

2 术语冲突类别及其定义

2.1 术语定义

为系统探究数据领域标准术语的冲突

状况，首先需对“术语”的内涵予以明确。在术语学中，术语是专业领域中概念的语言指称^[7]；概念是客体在人类意识中的反映，是通过抽象概括客体特性而形成的思维单元。而定义则是对概念的语言描述，其核心功能在于界定某一概念在所属概念体系中的确切位置，并将其与相关概念清晰区分。在术语工作中，根据专业领域、研究视角与具体目标的不同，对同一客体可采取不同的观察与定义方式。如表1所示，常见的定义方法主要包括内涵定义、外延定义和组成部分定义^[10]。其中，内涵定义是构建概念体系时最核心的定义模式，通常采用“上位概念（属概念）+区别特征（种差）”的科学定义结构。该方法首先明确所定义概念所属的更高层级类别，再通过指出其独有的区别特征，使其与同一属概念下的其他并列概念区分开来。

2.2 术语冲突类别

基于术语定义的完整过程及其关键环节，本研究进一步归纳了术语定义过程中可能出现的各类冲突类型。从术语定义的工作流程来看，术语定义并非简单的概念描述，而是一个涉及概念分析、定义撰写与指称命名的系统性过程。在这一过程中，

表1 术语定义方法

定义方法	说明	示例
内涵定义	先指明概念所属的属概念,再通过种差将其与同属的其他概念区分开来,是术语工作中最常用的定义方式	元数据是定义和描述其他数据的数据。其中“数据”为上位概念,“定义和描述其他数据”为区别特征,将元数据与业务数据、主数据等并列概念区分开来 ^[11]
外延定义	通过列举概念所涵盖的全部下位概念或具体实例来完成定义,适用于下位概念数量有限且广为人知的情形	数据资产按应用形态可分为基础数据资产、衍生数据资产、数据产品资产等; 结构化数据包括关系型数据库中的二维表数据、CSV 文件等 ^[12]
组成部分定义	通过描述概念所指客体的结构组成来界定概念,常用于技术产品或系统类术语	数据管理系统由数据采集模块、数据存储模块、数据处理模块、数据分析模块和数据安全管理模块等组成 ^[13]

由于不同主体对概念的认知差异、定义表述的规范程度不一以及术语指称的选用习惯不同,容易产生多种类型的冲突。该分类体系的构建遵循了术语定义的自然工作流程,即从认知层面的“概念分析”到语言表述层面的“定义撰写”,再到体系映射层面的“指称与体系”,并在此逻辑框架下识别每个阶段可能产生的具体冲突。

如图1所示,本研究依据术语定义的工作流程,即“概念分析→定义撰写→指称与体系”三阶段,并参考GB/T 10112-2019对概念、定义和指称的规范,系统推导并构建了上述10类冲突类型。在概念分析阶段,冲突主要表现为归属分歧与特征分歧两类。归属分歧涉及概念在概念体系中的上下位关系定位,具体包括层级冲突和领域冲突。层级冲突指错误地将概念归入不恰当的属概念,导致与其他概念的逻辑关系混乱,例如将“数据仓库”分别归入“数据库系统”与“决策支持系统”两种不同的属概念;领域冲突则指不同学科或专业领域对同一术语的归属存在不同理

解,例如计算机科学将“元数据”定义为“关于数据的数据”,而图书馆学则将其定义为“关于资源的数据”,两者基于各自领域的概念体系,造成跨领域交流时的歧义。特征分歧则体现在对概念区别特征的界定差异上,具体包括程度冲突、时效冲突和数值冲突。程度冲突表现为对概念本质特征在量化程度、等级或强度上的界定不一,如“大数据”有的定义为“规模超过传统数据库软件处理能力的数据集”,有的则具体量化为“数据量达到PB级别以上”;时效冲突是指定义依据的特征或事实随时间的推移已更新,但旧定义仍被沿用,如“数据挖掘”的定义从早期的“从数据库中提取知识”演变为当代“通过算法发现模式的过程”;数值冲突则涉及定义中关键参数、指标或数值标准的直接矛盾,如“数据压缩率”存在压缩后/原始大小与原始/压缩后大小两种相反的定义方式。

在定义撰写阶段和指称与体系阶段,冲突主要表现为形式差异、指称范围差异。定义撰写阶段的形式差异包括逻辑冲突、

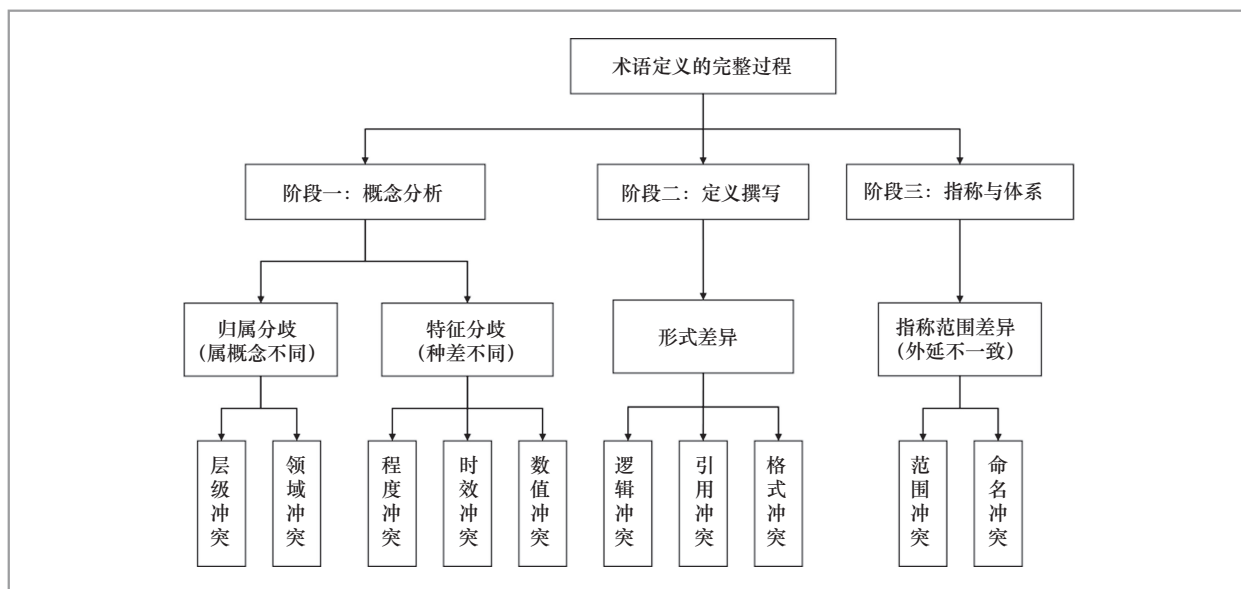


图1 术语冲突类别

引用冲突和格式冲突。逻辑冲突指定义本身存在自相矛盾的说法，如对“数据清洗”是否包含删除重复数据给出相互矛盾的界定；引用冲突指定义所依据的文献、标准或来源相互矛盾，如“数据质量”的定义分别引用 ISO 8000 与 DAMA 两种不同来源；格式冲突指定义撰写不符合规范要求，如使用循环定义（如“数据治理是治理数据的过程”）而非内涵定义。指称范围差异主要表现为范围冲突，即两个定义对术语所指对象的外延划定不同，导致边界宽窄不一，如“数据集”有的仅包含结构化数据，有的则涵盖结构化与非结构化数据两类。指称与体系阶段的核心冲突为命名冲突，即同一概念使用了不同的术语指称，造成概念与指称的映射混乱，例如“数据字段”与“数据列”指向同一概念却使用了不同术语。上述冲突类型共同构成了术语定义过程中需要系统识别与规范解决的

关键问题，为后续构建术语定义质量评估与分析框架提供了完整的理论依据。**表 2**系统定义了各类冲突，并提供了参考示例。

3 数据准备

3.1 数据来源

本研究的数据来源于全国标准信息公共服务平台。该平台是由国家市场监督管理总局、国家标准化管理委员会主办的国家级标准信息权威发布与服务平台，涵盖国家标准、行业标准、地方标准、团体标准及企业标准等多种类型，从而确保数据来源的权威性与全面性^[14]。

本研究以“数据”为关键词，在该平台进行标准名称与标准内容的检索匹配，共收集与数据相关的现行标准 612 项。在

表2 术语冲突类别定义及示例

阶段	冲突类型	定义	示例
概念分析	层级冲突	错误地将其归入不恰当的属概念,导致与其他概念的逻辑关系混乱	数据产品:被归入“数据服务”或“解决方案”
	领域冲突	不同学科或专业领域对同一术语的归属有不同理解	元数据:在计算机科学和图书馆学中解释不同
	程度冲突	对概念的本质特征在量化程度、等级或强度上的界定不一致	大数据:在定义中是否被量化
	时效冲突	定义依据的特征或事实已随时间的推移而更新,但旧定义仍被沿用	数据挖掘:20 世纪 90 年代定义仍在沿用
	数值冲突	定义中涉及的关键参数、指标或数值标准存在直接矛盾	数据压缩率:参数的计算方式不同
定义撰写	逻辑冲突	定义本身存在自相矛盾的说法,或者定义中的属性相互冲突	数据清洗:对是否包括删除重复数据存在不一致
	引用冲突	定义依据的文献、标准或来源相互矛盾	数据质量:应用内容不同
	格式冲突	定义撰写形式上采取不同形式,包括内涵定义(属+种差、循环定义、否定定义或定义项与被定义项同义反复等形式)	数据治理:采取循环定义或内涵定义
指称与体系	范围冲突	两个定义对术语所指对象的外延(即包括哪些具体事物)划定不同,导致边界宽窄不一	数据集:是否包括非结构化数据
	命名冲突	同一概念使用了不同的术语指称(同义术语)	数据字段与数据列定义相同

筛选过程中,采用关键词匹配的方式,确保获取的标准均围绕数据展开,构建了研究所需的基础标准数据集。

国家发展改革委等部门印发了《国家数据标准体系建设指南》,将数据领域标准划分为基础通用、数据基础设施、数据资源、数据技术、数据流通、融合应用、安全保障7个类别,构建了数据标准体系^[3]。本研究参考各类别定义,对这些数据标准进行了类别标注,统计结果如图2所示。

3.2 数据预处理

在完成612项相关标准的采集后,本研究对标准文本进行了数据预处理。标准文档多为非结构化的文本形式,难以直接对其进行定量分析与统计,本研究采用大模型结构化信息提取的方法,对每项标准中的术语相关内容进行自动化抽取。

具体而言,本研究利用大语言模型在语义理解与信息抽取方面的能力,构建统一的术语提取模板,从标准文本中识别并提取术语名称、定义内容、所属标准编号及章节位置等关键字段。例如,笔者从标准文本中提取了如下原始段落:“3.1 元数据 metadata: 定义和描述其他数据的数据[来源: GB/T 18391.1-2009, 3.2]”。大模型根据预设的提取模板,将该段落结构化处理为: {术语名称: 元数据, 定义内容: 定义和描述其他数据的数据, 所属标准编号: GB/T XXXXX-XXXX, 章节号: 3.1, 引用来源: GB/T 18391.1-2009, 3.2}。

经过结构化处理后,共获得28 133条术语条目,构建了标准术语数据集,为后续分析提供了结构良好、语义清晰的底层数据支撑。冲突涉及两个及以上不同术语定义,因此本研究采取同名异义和异名同义的方式,将术语数据集整理成了冲突对

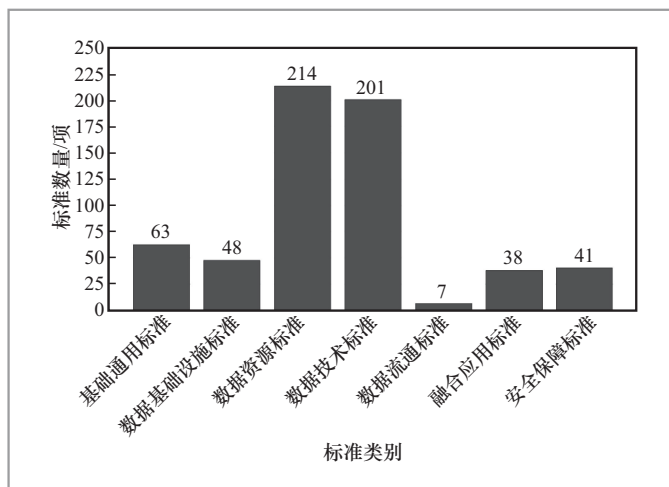


图2 各类标准数量分布

的形式,共得到48 228个术语对。

3.3 冲突类别标注

基于上述冲突类型分类体系与判断逻辑,本研究构建了如图3所示的术语定义冲突检测的标准化流程。第一步开展初始条件筛查,并行完成3类基础判断:一是检查定义表述是否存在互相矛盾,如“A是B,且A不是B”这类自相矛盾描述,判定为逻辑冲突;二是检查定义结构是否失范,如存在“A是A”的循环定义等格式问题,判定为结构冲突;三是比对两个定义内涵,若内涵完全一致但术语名称不同,如“记录”和“行”,直接判定为命名冲突。以上任一条件成立即输出对应冲突类型,全部筛查均无冲突则进入下一环节。第二步,检查引用来源,若两个定义引用了相互冲突的权威标准(如分别引用ISO和DAMA对同一术语的不同定义),则判定为引用冲突。第三步,比较属概念,若属概念存在层次差异(如一个认为“数据仓库”是“数据库”,另一个认为是“决策支持系统”),进一步区分层级冲突与领

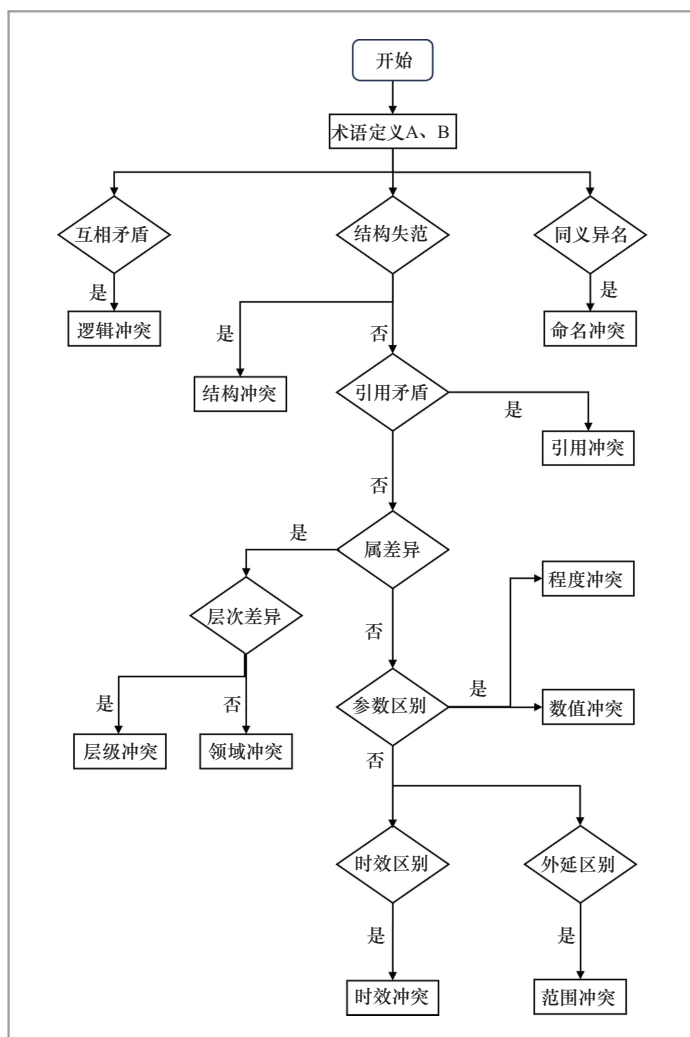


图3 冲突类别标注流程

域冲突。第四步，仅在属概念一致时，再开展种差对比：若种差涉及时间演变，则为时效冲突；涉及数值矛盾，则为数值冲突；涉及程度量级，则为程度冲突；涉及外延宽窄，则为范围冲突。

该流程严格遵循《术语工作 原则与方法》确立的术语工作原则，旨在对同一术语的两个不同定义进行系统性比对，并依冲突类型进行精确归因。整个流程采用顺序判断机制，将检测过程划分为3个阶段：第一阶段为预处理与有效性检查，主要识

别定义是否存在自相矛盾、循环定义、缺属或结构失范等基础性问题，直接归入“逻辑冲突”或“格式冲突”；第二阶段为要素拆解与逐层对比，在抽取属概念与种差要素后，依次判断引用冲突、属概念差异及种差差异；第三阶段为指称映射与最终判定，检查概念与术语的对应关系是否混乱，最终归入命名冲突。该流程以清晰的判断顺序替代了主观经验判断，确保了冲突归因的规范性与一致性。

该流程的核心特征是层次化的判断逻辑与严格的优先级设定。首先，引用冲突被赋予最高判断优先级，因其直接涉及权威依据的差异，对定义的有效性影响最大。其次，属概念的比较先于种差比较，符合术语学的基本逻辑。在种差比较环节，流程进一步区分了时间演化、数值矛盾、量化程度及外延边界等不同维度的差异，分别对应时效冲突、数值冲突、程度冲突与范围冲突，这种精细化的分类体系使原本模糊的“定义不一致”问题得以拆解为可识别、可归因的具体冲突类型。最后，在概念内涵一致的前提下，流程仍保留对指称映射的检查，避免因术语命名混乱而导致的隐性冲突被遗漏。总体而言，该流程为术语定义质量的系统评估提供了可复用的分析框架，既可用于人工比对场景下的规范指引，也可为后续自动化检测工具的开发奠定逻辑基础。

4 冲突分析

基于前文对冲突类型的定义以及冲突类别标注流程，本研究采用大模型标注的方式对术语对进行了冲突类别判断。采用分层抽样方法，从大模型标注结果中抽取100条数据，涵盖所有冲突类型，由领域

专家逐条人工标注。经检验，Cohen's Kappa 系数为 0.7055，说明大模型标注与人工标注结果显著一致。模型在多数类型上达到较好的一致性，可作为辅助标注工具，但对领域冲突等复杂情况仍需人工复核。通过大模型标注的方式，本研究对所有术语对进行了冲突类型判定，在筛除掉无冲突术语对后，剩余 24 575 个冲突术语对。

4.1 宏观态势扫描

4.1.1 类型分布

根据图4所示的各冲突类型数量分布，本研究对数据领域术语定义冲突的类型分布进行分析。从整体分布来看，10类冲突类型的数量存在显著差异，呈现出集中分布的特征。其中，层级冲突、引用冲突和范围冲突3类数量最突出，构成了冲突总量的主体部分；领域冲突、格式冲突和命名冲突数量居中；程度冲突、时效冲突、数值冲突和逻辑冲突的数量较少。

数值冲突和逻辑冲突数量相对较少。

层级冲突数量最高，反映了数据领域术语在概念体系定位上的普遍性问题。数据领域涉及数据库、数据管理、数据分析、数据应用等多个技术层次，不同来源的标准制定者在界定术语的属概念时往往依据各自的技术视角，导致同一术语被归入不同层级的属概念，造成概念体系的逻辑混乱。引用冲突数量次之，表明数据领域术语定义对权威来源的依赖程度较高，但由于引用的标准、文献或规范版本不一，不同定义之间因依据来源的差异而产生矛盾。范围冲突数量也较为突出，说明术语外延边界的划定在数据领域存在较大分歧，尤其在新兴技术领域（如数据湖、数据中台等），不同定义对概念涵盖对象范围的认知尚未统一。

相比之下，程度冲突、时效冲突、数值冲突和逻辑冲突的数量较少。程度冲突较少可能与数据领域术语的量化特征相对明确有关，多数术语的定义倾向于采用定性描述而非量化阈值；时效冲突较少则反

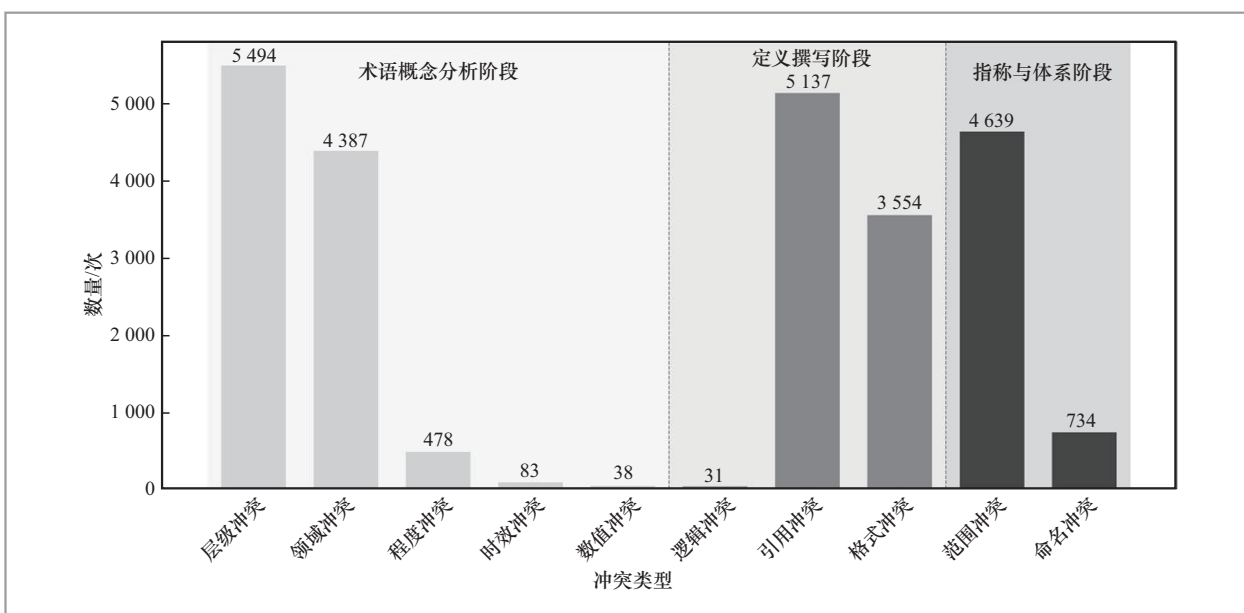


图4 各冲突类型数量分布

映出数据领域标准更新相对及时，新旧定义并存的矛盾尚未大规模积累；数值冲突数量最少，说明涉及关键参数直接矛盾的术语定义在数据领域并不普遍；逻辑冲突数量亦较低，表明大多数定义在逻辑自洽性方面表现良好。

4.1.2 时序统计

根据图5所示的各年份术语定义冲突数量分布，本研究对数据领域术语定义冲突的时间演化态势进行分析。从整体趋势来看，1992年至2025年，冲突数量在不同年份之间存在一定波动，其中2007年、2018年、2021年和2024年等时间点出现了较为明显的增幅，反映出数据领域术语工作在特定历史时期经历了重要的变化与

调整。

整体而言，标准数量的增加未能带来术语冲突数量的减少，反而呈现冲突数量的明显增长趋势。这一现象表明，在标准化工作整体推进的过程中，术语并未实现同步发展。标准数量的扩张，在缺乏系统性整合与协调机制的情况下，难以凝聚共识，在一定程度上加剧了术语层面的认知分歧与使用冲突。

从图5右侧数据可见，早期冲突类型较多且分布均匀，呈现多样化、分散化特征；2010年后，冲突类型逐步收敛，引用冲突、层级冲突、格式冲突、范围冲突、领域冲突成为核心冲突类型，命名冲突、时效冲突等其他类型占比下降。冲突类型从早期的多样化、分散化状态，逐步收敛为高度集中的核心类别，包括引用冲突、

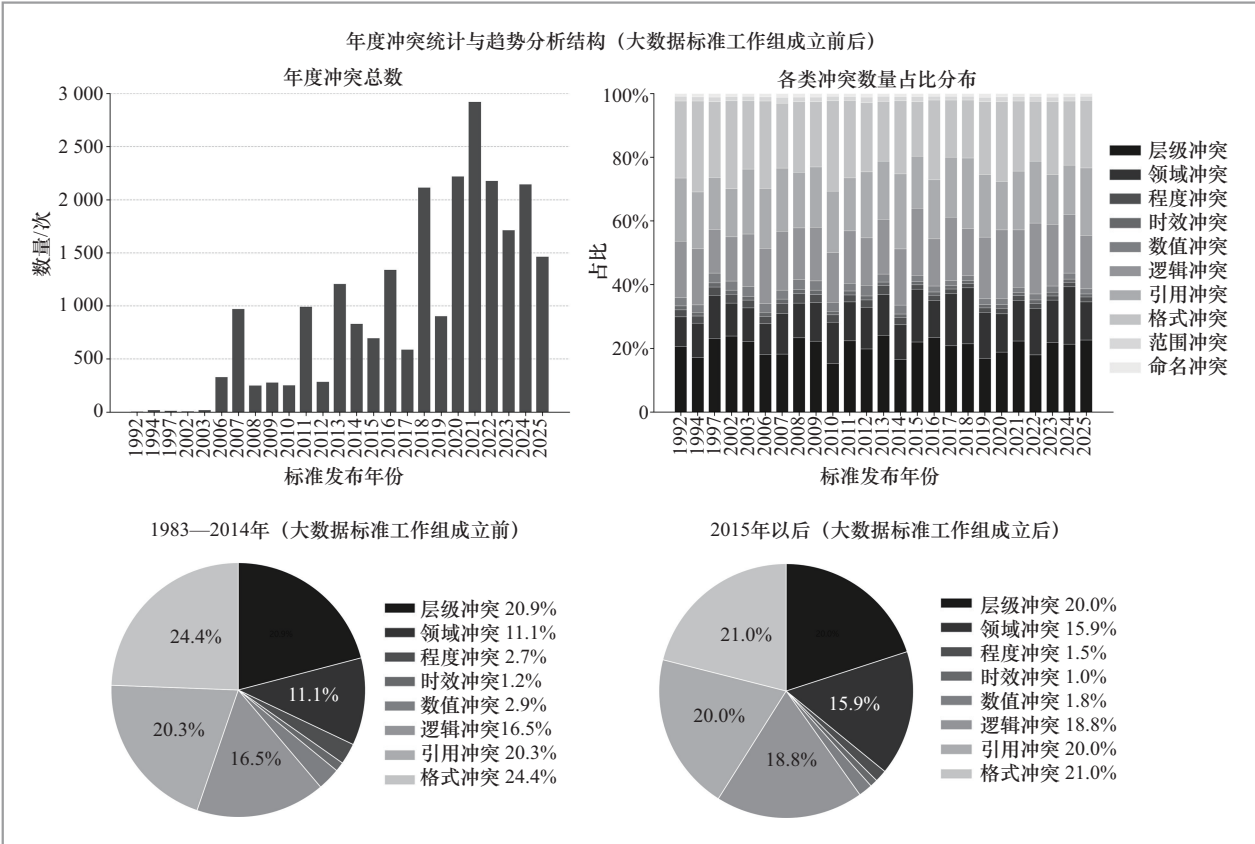


图5 冲突时序图

层级冲突、格式冲突、范围冲突、领域冲突。

从图5下方饼图可见，自2014年全国信息技术标准化技术委员会大数据标准工作组成立之后，层级冲突与范围冲突的占比明显下降，这表明术语适用边界在标准化工作推进过程中得到了一定程度的规范。但同时引用冲突占比大幅上升，成为最主要的冲突类型，这说明随着标准数量的增加和标准体系复杂度的提升，标准之间的相互引用关系成为术语冲突的主要来源。

4.1.3 来源分析

2024年9月，国家发展改革委、国家数据局、中央网信办、工业和信息化部、财政部、国家标准委六部门联合印发了《国家数据标准体系建设指南》，该指南在前期工作的基础上进一步优化完善，构建由基础通用、数据基础设施、数据资源、数据技术、数据流通、融合应用、安全保障7个部分组成的国家数据标准体系^[15]。这一分类框架系统反映了数据要素化背景

下标准化工作的顶层设计思路，为本研究的数据标准分类提供了权威依据。据此，本文将收集的612项数据相关标准归入上述7个部分，分别统计各类别的冲突参与次数，以揭示术语定义冲突在不同类型标准中的分布特征。

如图6所示，数据技术标准冲突记录数最多，这与其标准数量多、技术迭代快密切相关。数据技术标准涵盖数据采集、存储、处理、分析、可视化等全生命周期技术，标准基数大，加之新技术概念不断涌现，不同技术路线下的术语定义差异显著，冲突记录累积较高^[16]。数据资源标准冲突记录数同样较多，该类标准涉及数据分类分级、数据资产、数据质量、元数据等基础性内容，这些概念在不同领域和应用场景中的界定需求各异，标准数量也较多，导致冲突总量较高。基础通用标准、安全保障标准和数据基础设施标准的冲突记录数处于中游水平。基础通用标准作为标准体系的顶层规范，虽然覆盖面广、影响大，但标准数量相对有限；安全保障标

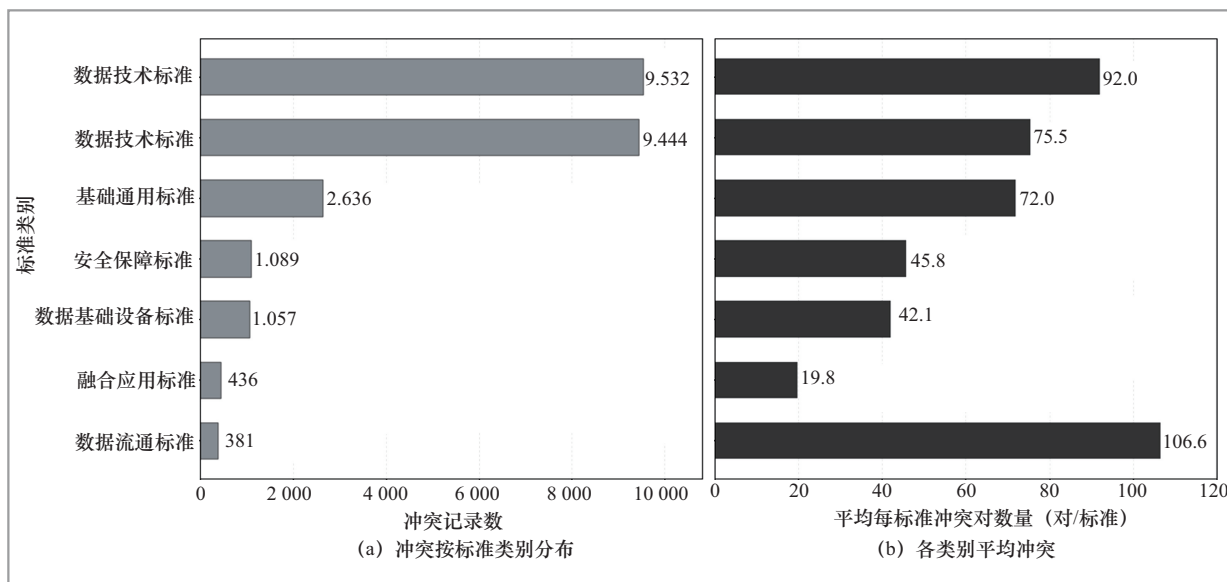


图6 各标准类别冲突统计图

准得益于《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律法规对相关术语的明确界定，冲突相对可控；数据基础设施标准涉及数据中心、网络、存储等硬件设施，术语体系较为成熟稳定，争议空间不大^[17]。融合应用标准和数据流通标准的冲突记录数相对较低。融合应用标准涉及数据在金融、医疗、交通等行业的具体应用，术语具有较强的领域专属性，跨领域交叉较少；数据流通标准虽然涉及共享、交易、开放等复杂环节，但该类标准数量本身较少，导致冲突记录总量不高。然而值得注意的是，如图5右侧所示，数据流通标准领域的平均每标准冲突对数量是最高的，这与当前数据流通领域跟其他领域交叉较多的现状是相关的，再次强调在标准制定过程当中，及时对术语冲突进行处理才能更好地支持数据流通。

总体来看，冲突记录数的分布与各类别的标准数量、技术活跃度以及术语体系成熟度密切相关。数据技术与数据资源标准因标准基数大、概念更新快、跨领域交叉多，成为术语定义冲突的集中区域；而融合应用、数据流通等类别因标准数量有限或领域边界清晰，冲突总量相对较低。

4.2 术语冲突深度解析

4.2.1 关键术语

国家数据局发布的两批数据领域名词解释在数据领域标准术语体系中具有顶层设计与核心枢纽的地位。这两批名词解释由国家数据局牵头、数据领域名词解释起草专家组研究形成，系统性地涵盖了数据从资源化、资产化到资本化全链条的核心概念，包括数据基础概念、数据处理与流通、数据技术、数据产权制度、市场流通

等领域的关键术语^[18]。作为国家数据管理主管部门发布的权威界定，其旨在凝聚广泛共识，为社会各界提供统一、权威的参考标准，本质上有效解决了长期存在的概念多源异构、相互冲突的问题，确保了跨部门、跨行业沟通的基础一致性。

在标准体系建设的具体实践中，这两批名词解释扮演着基础输入与基准参照的关键角色。当前，全国数据标准化技术委员会正在制定的《数据基础术语》等国家标准，其核心内容正是要定义数据领域的基础性术语，已发布的权威名词解释将成为该标准以及后续所有相关领域标准的核心术语来源与定义依据，使后续标准能够在统一的术语基础上构建，从而有效规避层级冲突、领域冲突和引用冲突等问题。同时，结合本研究建立的术语定义冲突检测框架，这两批名词解释为冲突判断提供了无可争议的官方基准：当一个标准中的定义与国家数据局发布的权威解释在属概念或种差上出现显著偏离时，即可依据其进行客观比对与冲突归因，使术语定义冲突分析从相对主观的专家判断转变为有据可依的规范操作，为整个数据标准生态的协调与规范发展奠定了坚实基础。表3列举了存在多次冲突（冲突次数超过2次）的关键术语。

本研究重点关注了这些关键术语的冲突情况，经过具体分析得出了如下结论。

元数据是冲突最集中的关键术语，集中体现为引用冲突与范围冲突。在统计的关键术语中，元数据的冲突数量最多、表现形式最多样。其定义多源引用、多种表述并存状况，导致引用冲突和范围冲突的集中出现。不同标准对元数据的外延界定宽窄不一，有的仅强调描述，有的则扩展至管理属性，使同一术语在不同标准中覆盖的范围存在明显差异。

表3 部分关键术语冲突情况

术语	名词解释	冲突次数/次	差异定义示例
元数据	定义和描述特定数据的数据,它提供了关于数据的结构、特征和关系的信息,有助于组织、查找、理解、管理数据	32	定义和描述其他数据的数据
数据	指任何以电子或其他方式对信息的记录。数据在不同视角下被称为原始数据、衍生数据、数据资源、数据产品和服务、数据资产、数据要素等	4	信息的可再解释的形式化表示,以适用于通信、解释或处理(注:可以通过人工或自动手段处理数据)
原始数据	指初次产生或源头收集的、未经加工处理的数据	3	终端用户存储使用的各种未经处理或简化的数据(注:原始数据有多种存在形式,如文本数据、图像数据、音频数据或者几种数据混合存在)
数据处理	包括数据的收集、存储、使用、加工、传输、提供、公开等	3	包括身份鉴别、访问控制、授权管理安全、数据脱敏、数据加密、数据防泄露、分布式处理安全、数据处理溯源、数据分析安全、安全审计等
数据资产	指特定主体合法拥有或者控制的,能进行货币计量的,且能带来经济利益或社会效益的数据资源	3	组织拥有和控制的、能够产生效益的数据资源

规范引用管理是提升术语一致性的关键路径。在元数据等关键术语的定义中,频繁出现引用不同版本国家标准的现象。虽然核心内涵相近,但具体表述和细节界定存在差异,这导致引用冲突。这一现象提示,术语工作中需加强引用来源的统一管理,优先采用最新版本的国家标准作为引用依据,并在标准制定过程中建立引用来源的协调机制,从源头上减少由版本差异导致的定义分歧。

4.2.2 核心标准

本研究基于每个术语冲突对的来源标准信息,统计出了每个标准发生冲突的数量,对数据领域标准冲突进行了深入考察,针对部分冲突数量较多的核心标准展开分析。表4展示了其中冲突数量最多的10个标准的具体情况。

如表4所示,基础资源类标准与技术应用类标准的冲突类型差异显著。基础资源类标准的冲突集中于层级冲突、范围冲突与引用冲突,本质指向数据领域核心概

念体系尚未定型、概念边界模糊、权威引用缺乏统一规范等根本性问题;技术应用类标准的冲突则以格式冲突为主,反映的是标准制定过程中定义撰写的规范性不足。

格式冲突在技术类标准中高度集中。大量应用交换类标准在定义术语时未遵循内涵定义规范,循环定义、缺属、结构失范等问题普遍存在。这说明,标准审查环节对术语定义格式的审核不足,亟须将格式合规性作为术语定义的基本要求。

交叉领域标准的冲突更加凸显。在涉及生态、航天、行政与商业等多学科交汇的标准中,术语冲突集中,并且冲突类型在层级冲突、领域冲突、范围冲突、引用冲突之间分布相对均衡。这种均衡分布反映出,多学科背景的标准制定者对同一术语的理解天然存在差异,导致概念归属、外延边界、引用依据各环节均可能出现分歧。这表明,交叉领域的术语标准化需建立跨学科协同机制,而非依赖单一学科的逻辑自治。

表4 部分核心标准冲突情况

标准	类型	冲突次数/次	主要冲突类别	占比
生态科学数据元数据	数据资源	716	层级冲突	27.2%
			范围冲突	25.2%
			领域冲突	20.1%
工业机器人云服务平台数据交换	数据技术	664	格式冲突	94.4%
			领域冲突	2.26%
			命名冲突	1.51%
地理空间数据交换格式	数据技术	562	引用冲突	48.7%
			领域冲突	19.9%
			层级冲突	12.5%
机械 科学数据 第2部分:数据元目录	数据资源	495	格式冲突	44.8%
			引用冲突	43.2%
			范围冲突	4.2%
航天术语 空间数据与信息传输	基础通用	434	领域冲突	24.9%
			引用冲突	20.3%
			层级冲突	44.8%
行政、商业和运输业电子数据交换 第2部分:复合数据元目录	数据资源	423	格式冲突	30.0%
			范围冲突	19.8%
			引用冲突	15.6%
信息安全技术 数据库管理系统安全评估准则	安全保障	418	引用冲突	73.7%
			领域冲突	11.7%
			格式冲突	6.7%
信息技术 大数据 术语	基础通用	414	引用冲突	40.1%
			层级冲突	21.3%
			范围冲突	14.9%
术语数据库开发文件编制指南	基础通用	404	层级冲突	39.9%
			引用冲突	14.9%
			范围冲突	13.9%
地下管线数据获取规程	数据技术	397	引用冲突	83.1%
			领域冲突	6.5%
			格式冲突	2.8%

4.2.3 主体分析

本研究对各类别起草单位的术语定义冲突程度进行了分析。标准信息中起草单位以顿号、逗号等分隔为多个单位名称；同一标准内对单位名称去重，再按规则将

各单位映射为主体类别（高校、科研院所、企业、政府/事业单位、标准化委员会、其他等）。首先，计算每一个标准自身的冲突量；其次，利用单个标准的冲突量计算该标准所有起草单位的平均量；再次，对每个起草单位，把该单位参与的所有标准的

平均量求和；最后，为了防止由某类单位参与起草的标准数量多而导致冲突较多的情况出现，本研究对统计数据进行了归一化，用每一个起草单位的总量除以该单位参与起草的标准数量，以此得到该起草单位的术语定义冲突程度。图中数值反映了不同主体类型在标准起草过程中因术语定义差异而产生的冲突强度，数值越高代表冲突越显著。

如图7所示，政府与事业单位的层级冲突占比较高。在各类主体中，政府与事业单位的引用冲突占比最低，而层级冲突占比最高，与其他主体形成显著反差。这表明政府机构在标准制定中更关注概念在体系中的位置定位，强调制度框架的层级性和逻辑自洽性，而非对外部引用来源的依赖。

企业的引用冲突占比最高。这一现象表明企业在术语定义时更倾向于引用既有权威依据，而非自主构建概念体系，但是对引用来源的规范性验证不足。当引用来

源存在版本差异或标准冲突时，标准更容易陷入引用冲突。

引用冲突与格式冲突在各类主体中普遍存在。从数据来看，引用冲突在各主体中均占有较高比重，格式冲突占比同样普遍较高，在所有冲突类型中位居前列。这两类冲突不受主体类型影响，具有跨主体的普遍性特征。这一现象表明，引用管理与格式规范是提升术语一致性的基础性工作，应作为标准化流程中的刚性约束加以落实。

另外，考虑到不同起草单位之间可能存在术语定义冲突，本研究对不同类别起草单位之间的协作情况进行了分析。首先，计算每个标准涉及的冲突总次数；再根据标准信息中的起草单位，把各单位归并为主体类别（高校、科研院所、企业、政府/事业单位、其他等）。若某标准对应 m 个不同类别，则将该标准的冲突总量平均分配到这 m 个类别组上，即每个组合分得总量的 $1/m^2$ 对所有标准求和，得到冲突强度矩阵。为突出实体主体，本研究不考虑标

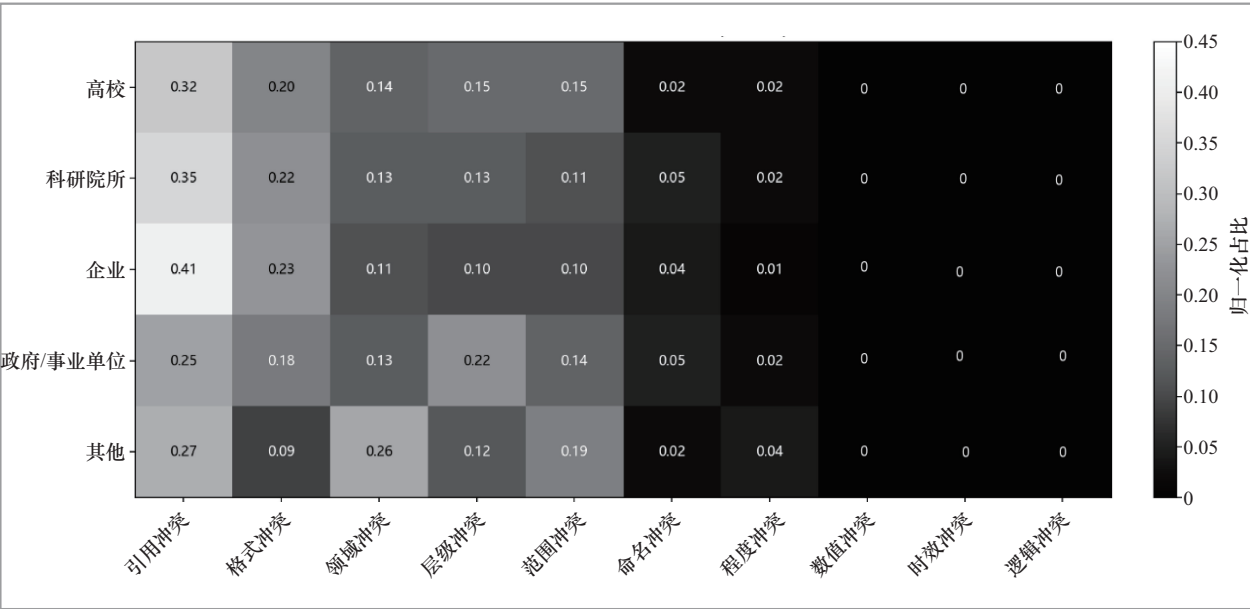


图7 起草主体冲突热力图

准化机构在术语冲突中的定位，在去除这一类别后对矩阵按行归一化，使图中数值表示在每一行对应的主体类别下，冲突在不同列类别上的相对占比。如图8所示，各主体与同类主体冲突高于跨主体冲突。从数据来看，所有主体的自冲突比例均不低于0.35，其中企业、高校、政府/事业单位均超过半数，科研院所和其他单位也远高于其与外部主体的冲突比例。因此，术语冲突并没有因为同种类型的主体而减少，反而可能因为同类主体的局限性，冲突水平更高。

4.3 原因分析

经过对术语冲突现象的系统分析，可将数据领域标准术语产生定义冲突的深层原因归纳为以下4个方面。

(1) 概念体系尚未定型

这是引发术语冲突的根本原因。与数学、物理等传统学科不同，数据领域是一个新兴的交叉领域，其众多核心概念的內

涵边界与逻辑关系仍处于动态演化和探索之中，尚未形成稳定、公认的体系。同一术语在不同技术语境或业务场景下，常常承载差异化的语义。这种固有的模糊性与多义性使标准制定者在引用或界定术语时缺乏统一的语义框架作为基准，从而直接导致了概念归属的层级冲突与外延划定的范围冲突。

(2) 标准化工作机制存在结构性缺陷

一方面，数据标准涵盖国家标准、行业标准、地方标准、团体标准等多个层级，不同制定组织之间缺乏高效、权威的横向协调机制，容易产生定义矛盾。另一方面，标准研制周期与产业和技术创新速度之间存在时滞，当新业态、新模式快速涌现并催生新概念时，标准制定往往难以同步跟进，导致旧标准中的过时定义与产业实践中的新概念并存，进而引发时效冲突。

(3) 多元主体间的认知与诉求存在固有差异

术语定义本质是一种社会共识，而数

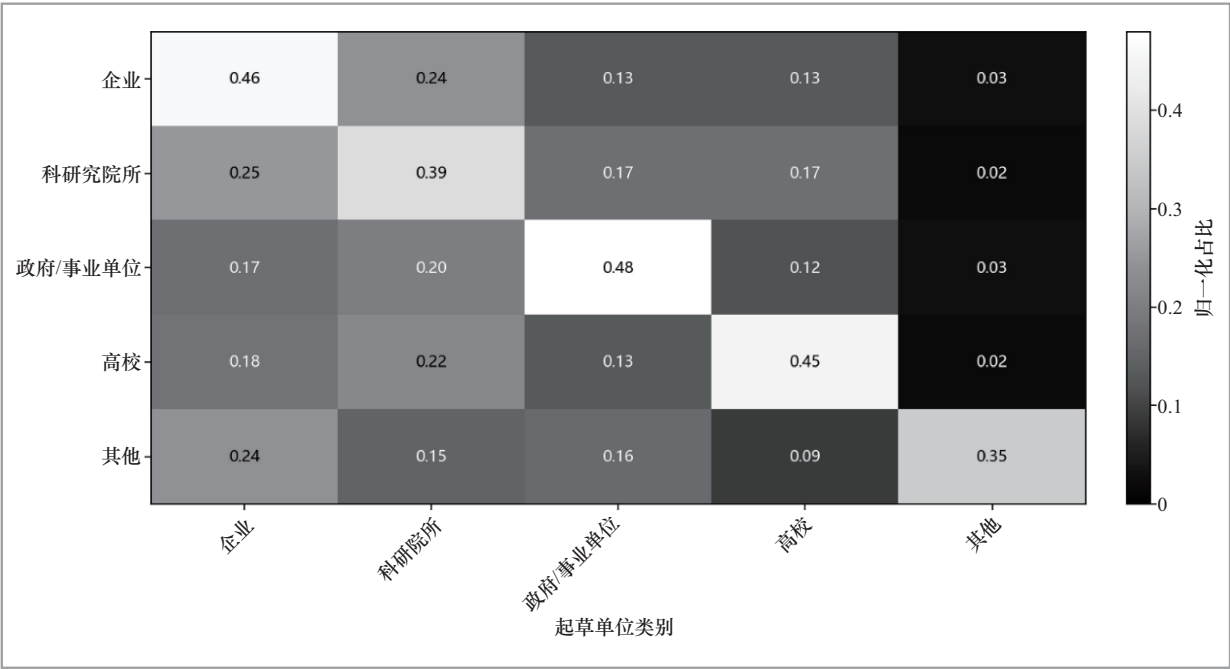


图8 起草主体协作热力图

据领域恰好汇聚了技术、法律、经济、管理等多重维度的专业逻辑与多元利益主体。不同背景的参与者对同一术语的解读存在天然侧重：技术专家关注实现细节，法律专家聚焦权属合规，经济学家强调价值流通。这种深刻的认知差异在标准研制过程中，直接表现为对概念归属的分歧和对特征描述的分歧。此外，学术研究者、政策制定者、行业从业者等不同实践主体，基于其各自的目标与视角，对术语的理解也难以一致，使共识凝聚变得异常困难。

（4）标准化流程中的引用管理机制缺失

从“元数据”等关键术语的冲突情况可知，引用冲突问题尤为突出，这暴露出标准研制流程在引用管理上的不足。首先，缺乏统一的引用规范指引，起草者在撰写定义时往往依据自身熟悉的文献或标准版本，导致对同一术语的引用来源五花八门；其次，当不同来源对同一术语的界定存在差异时，缺乏明确的权威性优先级判断标准，使得起草者无所适从，进一步加剧了引用层面的混乱与冲突。

5 数据领域术语标准化工作建议

基于前文对术语定义冲突的系统分析，本研究从概念体系构建、标准化机制优化、多主体协同治理、引用管理规范4个维度，提出以下工作建议。

（1）构建统一的概念体系框架

针对数据领域概念体系尚未定型、术语多义性突出的问题，建议以《数据基础术语》国家标准的研制为契机，系统梳理数据领域核心概念的内涵边界与逻辑关系。首先，明确基础术语的定义方式，确保术语定义的科学与规范性。

其次，厘清易混淆概念之间的逻辑关系。最后，借鉴国际经验，建立数据领域术语的概念体系图谱，通过可视化方式呈现术语之间的上下位关系、层级结构和关联路径，为标准制定者和使用者提供清晰的参照框架。

（2）建立全生命周期的术语协调体系

针对标准化机制的结构性缺陷，建议从以下3个方面入手。首先，建立标准制定的事前协调机制。在国家标准、行业标准、地方标准、团体标准立项之初，应进行术语冲突预评估，识别与既有标准在术语定义上可能存在的冲突点，从源头减少冲突的产生。其次，建立标准发布的事中审查机制。在标准审查环节，应增设术语一致性审查流程，由术语标准化专家对定义格式、引用来源、概念定位进行规范性审核，确保新发布标准与已有标准在术语体系上保持一致。最后，建立标准实施的事后维护机制。建立术语标准的定期复审与更新制度，及时淘汰过时定义、更新滞后概念、修复冲突问题。

（3）搭建多主体协同平台

针对多元主体认知差异导致的术语冲突，建议构建常态化的多主体协同机制。首先，发挥全国数据标准化技术委员会的统筹作用，邀请来自政府部门、科研院所、高校、企业、行业协会的代表共同参与标准研制。其次，建立术语标准研制的开放参与机制。通过公开征求意见、专题研讨会、线上意见征集等方式，扩大标准研制的参与范围，让更多利益相关方有机会表达对术语定义的看法。同时，针对标准实施过程中发现的术语歧义问题，可通过标准解释、实施指南等形式及时回应，避免因术语理解偏差导致的应用混乱。

（4）规范引用管理机制

针对引用冲突频发的问题，建立系统

化的引用管理规范。首先，明确术语定义引用的优先级顺序：优先引用国家数据局发布的两批名词解释作为基础依据，其次引用已发布的国家标准，再次引用行业标准或国际标准。当不同来源的定义存在冲突时，应遵循“最新版本优先、国家级优先、行业共识优先”的原则。其次，建立引用来源的统一台账。对数据领域常用术语的引用来源进行梳理，明确每个术语推荐采用的标准版本和定义表述，为标准制定者提供权威、统一的引用参照。再次，针对元数据等频繁出现多版本引用的术语，建议由全国数据标准化技术委员会牵头组织专项协调，统一引用来源，消除版本差异导致的冲突。

6 结束语

由于数据要素化进程的深入推进，数据标准化工作正从技术支撑走向制度构建，传统的分散式标准制定模式正在向体系化、协同化方向发展。推动数据领域术语工作的关键因素是保障概念体系的逻辑自洽、定义表述的规范一致、跨域沟通的顺畅高效、引用依据的权威可信。在这一逻辑下，保障数据领域术语标准化建设的路径，应当是构建统一的概念体系框架、建立全生命周期的术语协调机制、搭建多主体协同治理平台、规范引用管理机制。2025年10月，全国数据标准化技术委员会下达的第九批推荐性国家标准计划中，《数据基础术语》正式立项，此举具有标志性意义^[19]。该项标准由全国科学技术名词审定委员会事务中心牵头，联合中国电子技术标准化研究院、国家数据发展研究院、北京大学、中国人民大学、华为技术有限公司等30余家单位共同研制。其研制目标聚

焦于数据术语领域最核心的基础部分，旨在系统梳理并规范数据领域的基本、通用、重要及新兴术语，以应对当前产业实践中存在的术语定义不统一、概念边界模糊、跨领域沟通存在壁垒等关键问题，从而为我国数据要素的市场化配置奠定坚实基础。

本研究立足于我国数据领域术语标准化工作的现实需求，系统构建了包含10类冲突类型的术语定义冲突分类体系与检测流程。基于全国标准信息公共服务平台收录的612项数据相关标准，通过大模型结构化信息提取与人工标注相结合的方式，构建了包含24 575个冲突术语对的数据集，并从类型分布、时序演化、标准类别、关键术语、核心标准、起草主体差异等多个维度展开了系统性分析。本研究揭示了当前我国数据领域术语定义冲突的宏观态势，为识别术语冲突的系统性特征、厘清冲突产生的深层原因提供了实证支撑。希望研究成果能够为进一步规范关键标准的研制工作提供参考，通过在标准制定过程中前置识别与规避潜在冲突，提升术语定义的规范性与一致性，支撑我国数据领域标准化工作的推进。

参考文献：

[1] 吴非，胡慧芷，林慧妍，等. 企业数字化转型与资本市场表现——来自股票流动性的经验证据 [J]. 管理世界，2021，37(7): 130–144, 10.
Wu F, Hu H Z, Lin H Y, et al. Corporate digital transformation and capital market performance: empirical evidence from stock liquidity[J]. Management World, 2021, 37(7): 130–144, 10.

[2] 欧阳日辉，徐远彬. “十五五”时期我国实体经济和数字经济深度融合的阶段性特征与战略举

- 措[J]. 经济纵横, 2025(10): 9-20.
- Ouyang R H, Xu Y B. Phased characteristics and strategic measures of deep integration of China's real economy and digital economy during the 15th Five-Year Plan period[J]. Economic Review, 2025(10): 9-20.
- [3] 杨辉, 王锋, 杜潇霖, 等. 国内外数据要素标准研究综述[J]. 质量与认证, 2024(7): 30-33.
- Yang H, Wang F, Du X L, et al. A review of research on data element standards in China and abroad[J]. China Quality Certification, 2024(7): 30-33.
- [4] 文英姿, 曲杨, 张旭东, 等. 数据交易相关法规比较研究[J]. 大数据, 2022, 8(3): 66-77.
- Wen Y Z, Qu Y, Zhang X D, et al. Comparative study on laws and regulations related to data transaction[J]. Big Data Research, 2022, 8(3): 66-77.
- [5] 欧阳日辉, 李晓壮. 科学数据形成新质生产力的理论逻辑、典型案例与发展策略[J]. 科学管理研究, 2024, 42(6): 61-71.
- Ouyang R H, Li X Z. The theoretical logic, typical cases, and development strategies of scientific data as the new quality productivity[J]. Scientific Management Research, 2024, 42(6): 61-71.
- [6] 代红, 张群, 尹卓. 大数据治理标准体系研究[J]. 大数据, 2019, 5(3): 47-54.
- Dai H, Zhang Q, Yin Z. Study on big data governance standard system[J]. Big Data Research, 2019, 5(3): 47-54.
- [7] 中国标准化研究院, 中国国家标准化管理委员会. 术语工作 原则与方法: GB/T 10112-2019 [S]. 北京: 中国标准出版社, 2019.
- China National Institute of Standardization, Standardization Administration of China. Terminology work—Principles and methods: GB/T 10112-2019[S]. Beijing: Standards Press of China, 2019.
- [8] OSCARS Project. Bridging terminological gaps to strengthen data management quality at RDA P25 [Z]. 2025.
- [9] Gross F, Hundsdörfer M, Jung R, et al. "The good, the bad and the ugly" of terminology in cross domain fusion: introduction of the dialogue: wiki[J]. Informatik Spektrum, 2023, 46(2): 114-117.
- [10] 黎东良, 林媛, 王怡明. 论术语学中的定义与“有缺陷的定义”[J]. 中国科技术语, 2025, 27(3): 43-48.
- Li D L, Lin Y, Wang Y M. On definitions in terminology and “defective definitions” [J]. China Terminology, 2025, 27(3): 43-48.
- [11] 中国电子技术标准化研究所, 中国国家标准化管理委员会. 信息技术 元数据注册系统 (MDR) 第 1 部分: 框架: GB/T 18391.1-2009 [S]. 北京: 中国标准出版社, 2009.
- China Electronics Standardization Institute, Standardization Administration of China. Information technology—Meta-data registries (MDR)—Part 1: Framework: GB/T 18391.1-2009[S]. Beijing: Standards Press of China, 2009.
- [12] 中国电子技术标准化研究院, 中国国家标准化管理委员会. 信息技术 大数据 数据分类指南: GB/T 38667-2020 [S]. 北京: 中国标准出版社, 2020.
- China Electronics Standardization Institute, Standardization Administration of China. Information technology—Big data—Guide for data classification: GB/T 38667-2020[S]. Beijing: Standards Press of China, 2020.
- [13] 中国电子技术标准化研究院, 中国国家标准化管理委员会. 信息技术 大数据 数据管理系统 总体架构: GB/T 37973-2019 [S]. 北京: 中

国标准出版社, 2019.

China Electronics Standardization Institute, Standardization Administration of China. Information technology—Big data—Overall architecture of data management system: GB/T 37973-2019[S]. Beijing: Standards Press of China, 2019.

[14] 马超. 标准数据形态变革视域下我国标准数字化转型发展综述[J]. 信息通信技术与政策, 2024, 50(2): 74-81.

Ma C. Digital transformation and development of standards in China from the perspective of transformation of standard data forms[J]. Information and Communications Technology and Policy, 2024, 50(2): 74-81.

[15] 国家发展和改革委员会, 等. 国家数据标准体系建设指南[Z]. 2024.

National Development and Reform Commission, et al. Guidelines for the construction of national data standard system[Z]. 2024.

[16] 周奕玮. 钢铁企业大数据应用现状与实施策略探索[J]. 现代工业经济和信息化, 2025, 15(6): 47-50.

Zhou Y W. Enterprise big data application status and implementation strategies [J]. Modern Industrial Economy and Informationization, 2025, 15(6): 47-50.

[17] 高志宏. 数字经济时代个人信息的识别标准及规范续造[J]. 社会科学辑刊, 2023(2): 68-77.

Gao Z H. The identification standard and specification of personal information in the digital economy era[J]. Social Science Journal, 2023(2): 68-77.

[18] 国家数据局. 《数据领域常用名词解释》[J]. 工业信息安全, 2025(2): 94-97.

National Data Administration. Common terminology in the data field[J]. Industrial Information Security, 2025(2): 94-97.

[19] 中华人民共和国工业和信息化部. 5项数字化转型国家标准正式立项[J]. 工业控制计算机, 2023, 36(1): 64.

Ministry of Industry and Information Technology of China. Five national standards for digital transformation of - ficially approved for establishment[J]. Industrial Control Computer, 2023, 36(1): 64.

作者简介



黄科满（1988-），男，博士，中国人民大学信息学院副教授，中国人民大学吴玉章青年学者，麻省理工斯隆管理学院网络安全研究中心兼职研究员，国际网络安全专家CISSP认证。主要研究方向为数字生态、数据治理与网络安全治理。



张鼎镇（2002-），男，中国人民大学智慧治理学院硕士生，主要研究方向为数据治理、大语言模型。



朱尧（1999-），男，中国人民大学智慧治理学院硕士生，主要研究方向为数据治理、大语言模型。



刘艳芳（1983-），女，中国民航信息网络股份有限公司高级工程师，主要研究方向为数据分析、航空业数据标准、数据治理、数据安全。



李洋（1977-），男，中国民航信息网络股份有限公司高级工程师，主要研究方向为航空业数字化转型、数据平台架构规划、数据标准、数据安全等。



杜小勇（1963-），男，博士，中国人民大学吴玉章讲席教授、明理书院院长，数据工程与知识工程教育部重点实验室主任，中国计算机学会会士，全国数据标准化技术委员会首批委员，全国信息技术标准化技术委员会委员。主要研究方向为数据库、大数据技术、数据治理。

录用日期: 2026-05-19
通信作者: 李洋, liyang@travelsky.com.cn
基金项目: 国家自然科学基金项目 (No.62441230, No.62172425); 中国工程院咨询项目“实现数据要素高水平共享流通的若干重大问题研究” (No.2024-JZ-03)
Foundation Items: The National Natural Science Foundation of China (No.62441230, No.62172425), Chinese Academy of Engineering Strategic Consulting Project “Research on Several Major Issues for High-level Sharing and Circulation of Data Elements” (No. 2024-JZ-03)