

# 《衍生数据认定方法 第2部分 通信行业》团体标准 编制说明

## 一、任务来源

2023年2月，中共中央、国务院印发《数字中国建设整体布局规划》，明确将数字中国建设作为推进中国式现代化的重要引擎，并指出要夯实数字中国建设基础，必须畅通数据资源大循环。同年12月，财政部发布《关于加强数据资产管理的指导意见》，要求有序推进数据资产化，并鼓励行业根据发展需要，自行或联合制定数据资产相关标准。这些战略部署对数据要素治理的规范化、标准化提出了紧迫要求。

北京市作为全国数字经济发展的“排头兵”，率先出台了一系列纲领性政策法规。例如，《北京市数字经济促进条例》规定，“市经济和信息化部门会同市场监管等有关部门推进数字经济地方标准体系建设，建立健全关键技术、数据治理和安全合规、公共数据管理等领域的地方标准”；又如，《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》明确提出，要“探索建立结构性分置的数据产权制度”，推动“界定数据来源、持有、加工、流通、使用过程中各参与方的合法权利”。

要实现数据产权的“结构性分置”，首先必须能科学、客观地区分不同形态的数据。特别是经过“加工”环节产生的“衍生数据”，是“数据加工使用权”和“数据产品经营权”等新型权利的核心载体。若缺乏一套公认的认可标准，产权分置将沦为空谈。

基于以上需求，北京市政数局在国家数据要素综合试验区（北京）2025年试验任务书中，设立了《衍生数据认定标准试验任务》，由北京联通、智慧足迹作为任务承担方，同时与北京市大数据中心、百度、电子四院、海天瑞声、认知洞察、煤科总院、北京健康有益科技有限公司、滴滴出行等衍生数据研发专业团体组织共同形成了专项工作组，并形成了一套1+N的《衍生数据认定方法标准》整体标准制定与实践体系。其中，北京联通、智慧足迹承担总则与《通信行业衍生数据认定标准》的制定。

## 二、制定规范的必要性和意义

### （一）解决数据流通交易的市场需要

在数字经济高速发展的现在，数据已经成为企业竞争的核心资源，而对于原始数据与衍生数据的权利归属问题常引发争议，特别是对于衍生数据的准确认定，已经成为数据管理、资产管理和合规活动中的重要环节。

在目前很多实际一线的数据流通交易场景当中，因为对衍生数据的认定不清，已经成为很多商业诉讼的导火索。

因此，对衍生数据的准确认定，已经成为明确数据权益、评估数据价值、实施分类分级保护、优化数据资产管理、保障数据安全以及促进数据合规高效流通交易的必要前提。

而衍生数据认定方法标准，能为数据产权的“三权分置”提供技术标尺，能为数据资产“入表”提供可信依据，为数据交易流通提供“标准化商品说明书”，是北京市数据要素市场化配置改革实践中的市场需求。

### （二）适应通信行业实际应用场景需求

通信行业作为国家重要的基础设施，其业务运行和管理天然伴随着海量、高并发、多源异构的数据产生。这些数据包括用户通信记录(通话、短信)、上网行为日志(流量、时长、访问内容)、设备信令、网络性能指标、位置信息、客户资料等。通信行业对数据的使用需求多种多样，例如:网络优化与规划、用户行为分析、增值业务开发、风险控制、满足监管要求等。

然而，通信行业的数据处理过程也面临诸多挑战:数据来源复杂、处理链条长且隐蔽、涉及个人信息保护、数据量巨大且实时性高等。

在这样的背景下，准确识别和认定通信行业中产生的大量衍生数据，对于数据资产管理、合规经营、数据安全和价值挖掘至关重要。

### （三）有利于明晰通信行业权责边界，筑牢安全底线

通过为通信衍生数据提供统一、规范的“身份证”，本标准作为通信运营商、互联网合作伙伴及第三方数据服务商等在数据处理活动中的各方主体，确立了清晰的权利、义务与责任边界。这有助于从源头上规范用户信令、网络日志、业务记录等多源数据的处理行为，有效防范在用户画像、位置洞察等典型场景下的数据滥用与隐私泄露风险，为行业筑牢数据安全防线。

#### （四）有利于稳定市场预期，激活通信数据创新动能

明确的规则是通信数据要素市场活力的催化剂。通过本标准，企业能够清晰地预见其在网络数据、用户行为数据等基础资源上的加工投入，可以形成受法律保护、可在市场交易的高价值合法资产。这将极大激发通信企业及生态伙伴投资于数据挖掘技术、开发多元化数据服务产品的积极性，为赋能千行百业的数字化转型注入强大动能。

#### （五）为数据产权界定、资产入表、交易流通等核心环节提供操作性支撑

一是为数据产权的“三权分置”提供技术标尺：本标准针对通信数据多源异构、实时流动的特性，提出的“数据处理过程分析法”，为判断针对原始通信数据的“加工”行为是否达到“实质性改变”、从而创设新的数据产品经营权，提供了客观、可量化的技术标尺。

二是为数据资产“入表”提供可信依据：本标准通过规范化的认定流程与统一的结果表述要求，确保了在网络智能运维、客户关系管理等业务中产生的衍生数据资产，其生成过程可追溯、可验证、可审计，为会计师事务所与资产评估机构采信此类数据资产提供了可靠依据。

三是为数据交易流通提供“标准化商品说明书”：本标准实际上为“衍生数据”这一在通信交易市场中的核心产品，提供了一份详尽的“商品说明书”，明确了其数据来源、加工方法、质量指标与适用场景，有助于降低交易成本，减少交易纠纷，建立市场信任，促进通信数据要素安全、合规、高效地流通。

### 三、主要编制过程

1. 第一阶段，成立标准研制起草工作组（以下简称工作组），编制项目立项材料和标准草案。2025年11月，智慧足迹牵头成立标准编制工作组，开始标准起草工作。工作组通过查阅文献，形成理论框架和工作方法，并对已有相关内容基础进行了梳理，多轮研讨沟通，形成立项申请材料和标准草案。工作组内部多次召开立项准备沟通会，为项目立项制定完整方案。

2. 第二阶段，项目立项评审。工作组于2025年12月在中关村知识产权大厦A座，北京软件和信息服务业协会会议室召开了项目立项专家评审会。评审专家包括：首经贸数据要素市场化专题团队带头人、财税学院副院长陈蕾、清华大学软件学院研发主任敖然、京东集团公共业务副总裁孔鑫、北京大成律师事务所高

级合伙人陈福、百度资深标准化专家杨庄媛。标准立项通过专家评审后，信息在全国团体标准信息平台和协会官方网站和北京软协官网发布。

3. 第三阶段，编写规范征求意见稿。2026年1月-7月，围绕标准草案及各方意见，工作组开展了进一步调查研究，通过多轮内部研讨和标准内容文件修订，形成征求意见稿。

#### 四、制定规范的原则和依据，与现行法律、法规、标准的关系

本标准依据《中华人民共和国标准化法》及其《实施细则》，并参照《标准化工作导则第1部分：标准的结构和编写》GB/T 1.1-2020 编制，严格遵守以下五个原则。

**科学性原则：**立足通信行业“实时精准、网络强关联、数据多源异构”的核心特性，结合信息科学、通信工程及网络工程学科原理，确保衍生数据的定义、分类与判定指标等内容符合业务实际与技术规律，如采用信息熵量化分析等科学方法设定判定阈值。

**实用性原则：**确保标准能够满足实际工作需要，易于理解和操作，具有实用性。同时，标准应易于理解和实施，方便各相关方使用和推广。

**动态性原则：**充分考虑通信网络技术（如5G-Advanced/6G、网络数字孪生、通信大模型）、数据法规政策及数据服务模式的演进变化，预留指标扩展与认定方法更新的接口与机制，支持判定阈值根据网络负载、业务优先级等动态因素灵活调整，确保标准持续适配通信行业技术迭代与业务创新的长期发展需求。

**统一性原则：**注重与《衍生数据认定方法第1部分：总则》以及行业规范的衔接，统一术语定义、认定逻辑与编码规则，避免冲突矛盾，提升跨区域、跨企业互操作性。

**开放性原则：**坚持公开透明的编制流程，广泛征求通信运营商、设备供应商、科研院所、监管部门及法律机构等各方意见建议，确保标准制定过程公平公正，全面覆盖各类主体的合规与应用需求。

## 五、主要条款的说明，主要技术指标、参数、试验验证的论述

本标准结合通信行业特性，围绕衍生数据认定核心需求，核心内容如下：

### （一）主要条款说明

本标准结合通信行业“数据融合、算法驱动、场景多样”的特性，围绕衍生数据认定的客观性与一致性核心需求。核心内容如下：明确“用户信令数据”、“网络流量数据（DPI）”、“客户关系管理数据（CRM）”、“测试对象”等7项关键术语，界定衍生数据为“经关键处理操作形成的具备新信息价值的信息”，清晰划分与原始数据的边界。通信行业衍生数据认定需遵循四大原则，其中客观性原则要求基于真实数据处理逻辑与量化证据；可追溯性原则确保认定结论能追溯至输入数据、处理脚本、算法模型及测试过程；全面性原则需覆盖通信典型场景（如位置洞察、用户画像、流失预警等）全链条；审慎性原则规定当定性结论与定量结论不一致或涉及高敏感数据时，必须启动人工专家复核。处理过程优先原则明确以数据处理过程分析为基础判定依据，量化计算为验证手段。认定条件包括测试环境配置（硬件、软件、数据集）、测试程序（处理分析、量化计算、结果判定）与人员能力要求。认定要点包括处理过程分析法认定要点（识别关键转换操作）、量化计算验证要点（ECR、D\_KL 计算）与人工复核启动条件。判定准则：处理过程含关键操作（如聚合、模型预测、多源 JOIN）则初步判定为衍生数据；量化指标（ECR 或 D\_KL）超过阈值则定量判定为“显著差异”；两者一致则形成最终结论，不一致则启动复核。记录与报告：测试记录需结构化归档，测试报告应包含基本信息、测试结论、环境、过程、结果及附录，报告需测试工程师与审核人签字并加盖机构公章，结论变更需重新出具报告并说明原因。

### （二）主要技术指标与参数

本标准结合信息论原理与通信行业数据处理特点，设定核心技术指标如下：采用信息熵变化率（ECR）与相对熵（D\_KL）量化衍生数据与原始数据的差异，确保衍生数据具备信息内涵的显著变化。典型场景阈值如下：位置洞察类 ECR > 40%；用户画像类 ECR > 25% 或 D\_KL > 0.3；网络智能运维类 ECR > 20% 或 D\_KL > 0.2；客户流失预警类 D\_KL > 0.15。数据质量方面，输入数据需满足真实性、代表性要求，如位置洞察需至少 10 万用户 24 小时信令记录，用户画像需

至少 1 万用户多维度信息。测试环境配置需满足：CPU  $\geq$  8 核、内存  $\geq$  32GB、存储  $\geq$  1TB，操作系统为 Linux 或 Windows Server，编程环境为 Python 3.8+并配备 pandas、numpy、scipy、scikit-learn 等核心库。处理过程需可解析、可复现，算法模型需提供说明文档，全程记录完整可追溯。安全合规上，测试对象不得包含未脱敏个人信息，数据处理需符合 GB/T 35273 等国家标准，测试环境应具备数据访问控制与审计能力。

### （三）试验验证论述

为验证标准可行性，编制工作组联合通信运营商、第三方检测机构及数据服务企业，在多业务场景下开展试验验证：验证对象涵盖位置洞察报告、用户画像标签库、客户流失概率评分、网络运维指标等典型衍生数据，场景包括第三方检测、运营商自评、数据流通前认定等，覆盖不同区域与业务规模。通过案例验证（如静安寺人流热力图、高价值客户流失预警）、重复验证（多家机构独立测试同一数据集）、阈值灵敏度验证（调整 ECR/D\_KL 阈值观察结论稳定性）、实操验证（评估测试流程复杂度与耗时）等方法，得出以下结果：认定结果与业务专家判断高度吻合，能有效区分“真衍生”与简单转换数据；不同机构对同一数据集认定结论一致性显著提升，解决了以往依赖主观判断导致的认定分歧；标准在运营商、合作伙伴及第三方机构中均能顺利实施，测试流程清晰、代码工具可复用，经培训后可快速掌握，显著降低认定成本与操作门槛。综上，本标准符合通信行业数据合规流通与资产化管理需求，可有效解决衍生数据认定过程中主观性强、标准不一等问题，为数据价值释放与跨主体流通提供技术依据。

## 六、重大意见分歧的处理依据和结果的说明

本规范研究制定过程中未出现重大分歧。

## 七、采用国际标准或国外先进标准程度的说明，以及国内、外同类产品或标准的对比情况

### （一）参考和引用标准的标准号和标准名称

GB/T 35295—2017 信息技术大数据术语

GB/T 36344—2018 信息技术数据质量评价指标

GB/T 38667—2020 信息技术大数据数据分类指南

## （二）与国内外相关标准的内容异同

本标准在数据要素标准体系中具有独特的定位，它不重复现有标准，而是填补了关键空白，为其他标准的落地应用提供了必要支撑。

对比维度	本标准《衍生数据认定方法》	现有数据治理/资产/质量标准
核心目标	身份认定与界定：提供科学的原则、流程和方法，用于识别和确认衍生数据。	过程优化与状态评估：建立管理框架、评估能力、评价质量，以提升数据管理水平。
关注焦点	数据的“生成与演变”过程：聚焦从原始数据到新数据的转换过程、加工方式和血缘关系。	数据的“管理与应用”状态：聚焦数据在组织内的存储、流动、使用、安全等状态。
关系定位	前置基础与底层支撑：衍生数据的认定是开展数据资产管理、质量评价等活动的前提。	后置应用与管理框架：在数据形态被明确后，应用这些标准进行有效的管理和使用。

## 八、实施标准的措施建议

按照标准要求，推动企业构建、优化、评估及验证本标准。

## 九、其他说明

无。

《衍生数据认定方法 第2部分 通信行业》起草工作组

2026年7月8日