

河北省质量信息协会团体标准

《数据资产整合与挖掘规范》

(征求意见稿) 编制说明

标准起草工作组

2025年10月

一、任务来源

依据《河北省质量信息协会团体标准管理办法》，团体标准《数据资产整合与挖掘规范》由河北省质量信息协会于2025年9月16日批准立项，项目编号为：T2025396。

本标准由联城科技（河北）股份有限公司提出，由河北省质量信息协会归口。本标准起草单位为：联城科技（河北）股份有限公司、唐山国控科创集团有限公司、唐山联信数据有限公司。

二、重要意义

随着数字经济的快速发展，公共服务数据（如政务服务、医疗健康、教育民生、城市规划等领域数据）已成为政务服务高效便捷办理、医疗健康精准优质供给、教育民生资源均衡普惠、城市规划科学智慧推进的核心支撑要素。近年来，我国先后出台《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》《“十四五”数字经济发展规划》等政策文件，明确要求规范数据挖掘利用、保障数据安全隐私、促进数据要素合规流通，为公共服务数据挖掘工作提供了政策依据。

从实践来看，公共服务数据具有多源异构、敏感程度高、应用场景复杂等特点。例如，在政务服务领域，需整合人口、法人、政务办理等数据，以实现“一网通办”，提升政务服务效率与便捷性；在医疗健康领域，要融合患者诊疗记录、电子病历、医疗设备运行等数据，助力精准诊断、个性化治疗方案制定以及医疗资源的优化调配；在教育民生领域，需整合学生学籍、教学质量评估、教育资源分布等数据，为因材施教、教育资源均衡配置和教育质量提升提供支撑；在城市规划领域，通过整合地理信息、城市基础设施分布、人口流动监测等数据，能为城市空间布局优化、交通规划、公共服务

设施选址等提供科学依据，推动城市的智慧化、可持续发展。然而，当前公共服务数据挖掘过程中，存在缺乏统一技术规范的问题：数据采集缺乏标准化流程导致质量参差不齐，数据挖掘过程中隐私保护措施不到位存在泄露风险，数据流通与使用缺乏可信管控机制，多源数据融合缺乏统一技术要求，这些问题严重制约了公共服务数据价值的有效释放，也增加了数据合规使用的风险。

本标准基于综合治理等领域的实际数据应用场景，规范数据整合的技术要求、挖掘算法选型原则及安全管控措施，可解决跨系统数据接入难、挖掘结果可用性低等问题，为数据资产识别与评估提供高质量数据源，助力数据要素市场化配置。

三、编制原则

《数据资产整合与挖掘规范》团体标准的编制遵循规范性要求、一致性和可操作性的原则。首先，标准按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》、《河北省质量信息协会团体标准管理办法》等编制起草；此外，工作组在制定标准过程中遵循“面向市场、服务产业、自主制定、适时推出、及时修订、不断完善”原则，不断满足下游企业实际生产中对技术的需求，推动数据资产整合与挖掘技术多领域发展。

四、主要工作过程

2025年7月，联城科技（河北）股份有限公司牵头，组织开展《数据资产整合与挖掘规范》编制工作。2025年8月，起草组制定了《数据资产整合与挖掘规范》立项文件，进行了征求意见稿草案的编制，明确了编制工作机制、目标、进度等主要要求。主要编制过程如下：

(1) 2025年7月-8月：联城科技（河北）股份有限公司联合其他参编单位召开标准编制预备会，会议组织各单位开展资料收集和编制准备等相关工作。

(2) 2025年8月中旬：召开第一次标准起草讨论会议，初步确定起草小组的成员，成立了标准起草工作组，明确了相关单位和负责人员的职责和任务分工。

(3) 2025年8月中旬-8月下旬：起草工作组积极开展调查研究，检索国家及其他省市相关标准，调研数据资产整合与挖掘技术现状，分析了相关科研中积累的技术文件，并进行总结分析，为标准草案的编写打下基础。

(4) 2025年9月上旬：分析研究调研材料，由标准起草工作组的技术人员编写标准草案，通过研讨会、电话会议等多种方式，对标准的主要内容进行了讨论，确定了本标准的名称为《数据资产整合与挖掘规范》。并听取了相关专家和领导的意见和建议，确定了标准的大纲的各条款和指标的调研方案，在各参编单位的积极配合下，调研数据陆续反馈回主编单位。

(5) 2025年9月中旬：本标准起草牵头单位联城科技（河北）股份有限公司向河北省质量信息协会归口提出立项申请，经归口审核，同意立项。

(6) 2025年9月16日：《数据资产整合与挖掘规范》团体标准正式立项。

(7) 2025年9月-10月：起草工作组通过讨论，对标准草案进行商讨。确定本标准的主要内容包括数据资产整合与挖掘的基本原则、参与主体要求、数据架构、技术架构、数据流转模式、数据源接入要求、数据资源整合流程、挖掘算法选型、安全管控要求，初步形成标准草案和编制说明。起草组将标准文件发给相关标准化专家进行初审，根据专家的初审意见和建议进行修改完善，形成征求意见稿。

五、主要内容及依据

《数据资产整合与挖掘规范》团体标准的制订主要基于企业数据空间产品的数据资产整合与挖掘技术材料，并结合实际应用情况，作为本标准起草制定依据。

1. 范围

当前公共服务领域数据整合与挖掘缺乏统一规范，政府部门、企事业单位、第三方机构在数据对接、算法应用、安全管控中存在协同障碍，需通过标准界定适用范围与主体，解决“谁来做”“做什么”的问题。

2. 规范性引用文件

本标准没有规范性引用文件。

3. 术语和定义

下列术语和定义适用于本标准。

3.1

数据资产 data asset

合法拥有或者控制的，能进行计量的，为组织带来经济和社会价值的数
据资源。

3.2

数据资产整合 data asset integration

将分散在不同来源、不同格式、不同结构的数据资产，通过采集、清洗、转换、关联、虚拟化等一系列技术手段和管理流程，形成逻辑上统一、物理上可灵活调度的数据集合，实现数据资产的统一访问、共享和管理的过程。

3.3

数据资产挖掘 data asset mining

从整合后的数据资产中，运用统计分析、机器学习、深度学习、隐私计算等技术方法，发现潜在规律、关联关系、趋势特征等有价值信息，并将其转化为业务决策支持、产品创新、服务优化等成果的过程。

3.4

洋葱加密 onion encryption

对数据资产的每个字段根据使用场景、使用方身份、使用价格等因素，采用不同的加密算法进行多层加密处理，形成类似“洋葱”的加密结构，实现数据资产的细粒度访问控制和使用控制，防止数据未经授权拷贝、传输和泄露的多层加密技术。

4. 基本原则

合规性：依据《中华人民共和国个人信息保护法》第十三条“处理个人信息应当取得个人同意”、《数据安全法》第二十八条“开展数据处理活动以及研究开发数据新技术，应当有利于促进经济社会发展，增进人民福祉，符合社会公德和伦理”，明确敏感数据处理需获同意、跨部门流通需安全评估；

安全性：参考《信息安全技术 数据安全分级指南》（GB/T 35273—2020）中“分层防护”要求，结合隐私计算、洋葱加密等技术发展，提出“原始明文数据不出域”“密文流转可追溯”；

标准化：明确数据接口、格式、算法评估等统一要求，解决“数据不通”问题；

实用性：基于公共服务领域业务需求（如政务审批效率、医疗诊断时效），避免过度技术化导致资源浪费；

可追溯：依据《数据安全法》第三十八条“数据处理者应当建立数据安全保护制度，采取技术措施和其他必要措施，保障数据安全”，结合区块链“不可篡改”特性，提出“全流程日志存证”。

5. 参与主体要求

明确数据提供方、消费方、运营方、监管方、开发者 5 类参与主体的职责与技术要求，构建多主体协同的责任体系。

数据提供方：依据《数据安全法》中数据提供者的权属责任，要求其具备合法数据所有权/使用权，同时结合数据质量管控实践，提出分类分级、质量保障、使用监控等要求；

数据消费方：参考数据使用授权的常规流程，要求其通过身份认证获取权限，遵守使用协议，同时结合数据销毁的合规要求，明确使用后数据安全销毁义务；

数据运营方：基于平台运营的核心职责，提出空间云端搭建、认证授权、监控审计等要求，参考云计算平台的高可用与灾备标准，确保平台稳定性；

数据监管方：依据监管机构的法定职责，要求其具备全生命周期监管能力，参考数据安全监管的常规流程，明确监管标准制定、违法违规处理等要求；

数据开发者：结合安全开发实践，提出沙箱环境开发、数据申请授权、日志同步等要求，防范开发过程中的数据泄露风险。

6. 数据架构、技术架构

数据架构：围绕“本地数据-共享数据-空间服务-消费使用”闭环，明确数据在多主体间的流转路径，以客户端与空间服务端为核心载体，适配集中化管理与安全需求；

技术架构：采用“客户端-云端”协同架构，兼容中心化与去中心化部署模式，细化数据加密、隐私计算、协议控制、区块链存证、数据网格、客户端 6 大核心技术模块的要求。

7. 数据流转模式与数据源接入

数据流转模式：提出密文流转、按需调度、全链路存证3种模式，确保数据流转安全与高效；数据流转模式参考数据安全流通的最佳实践，密文流转基于加密传输技术，按需调度基于资源优化理念，全链路存证基于区块链不可篡改特性；

数据源接入：明确接入范围（政务、医疗、教育等领域数据）、接入方式（MQTT流式、文件导入、API对接）、接入流程（申请-审核-配置-验收）与接入后管理（监控、更新、下线）。数据源接入范围依据公共服务领域的核心数据类型确定，接入方式参考物联网（MQTT）、API接口（REST）的通用标准，接入流程参考数据接入的常规审批机制，接入后管理参考数据生命周期管理实践，确保数据源全周期可控。

8. 数据资产整合流程

将整合流程分为数据准备、数据采集、数据预处理、整合交付4个阶段，明确各阶段的核心任务与要求，如需求分析、方案设计、数据清洗、质量评估、数据注册等。

参考数据ETL（抽取-转换-加载）的经典流程，结合数据资产化的特殊需求（如合规审核、授权配置），补充需求分析、方案设计、整合交付等环节，形成覆盖“需求-采集-处理-交付”的全流程规范，确保整合后数据资产的质量与合规性。

9. 挖掘算法选型

明确算法部署模式（本地、云端协同、沙箱）与算法成果管理（注册、发布、监控更新）要求，适配不同场景与数据敏感级别的需求。

算法部署模式结合算法复杂度与数据敏感性：轻量级算法本地部署参考边缘计算理念，复杂算法云端协同参考分布式计算标准，高敏感数据沙箱部署参考安全开发环境实践；

算法成果管理参考模型生命周期管理（MLOps）标准，明确注册、发布、监控更新流程，确保模型可用性与时效性。

10. 安全管控要求

从身份认证与权限管控、数据安全管控、审计与合规管控3个维度，提出全链路安全要求，如三重身份认证、最小权限原则、数据存储/传输/使用安全、全链路审计日志等。

身份认证与权限管控参考零信任“持续验证”理念，结合多因素认证（MFA）标准，提出三重认证机制；权限管控基于“最小权限原则”，参考RBAC（基于角色的访问控制）模型，实现字段级、操作级精细管控；

数据安全管控参考数据安全分级保护标准，针对存储、传输、使用不同环节的风险，提出加密存储、SSL/TLS传输、使用行为监控等要求；

审计与合规管控参考数据审计的通用规范，明确日志内容、存证要求与合规检测规则，确保满足监管审计需求。

六、与有关现行法律、政策和标准的关系

本标准符合《中华人民共和国标准化法》等法律法规文件的规定，并在制定过程中参考了相关领域的国家标准、行业标准、团体标准和其他省市地方标准，在对基本原则、参与主体要求、数据架构、技术架构、数据流转模式、数据源接入要求、数据资源整合流程、挖掘算法选型、安全管控要求等内容的规范方面与现行标准保持兼容和一致，便于参考实施。

七、重大意见分歧的处理结果和依据

无。

八、提出标准实施的建议

建立规范的标准化工作机制，制定系统的团体标准管理和知识产权处置等制度，严格履行标准制定的有关程序和要求，加强团体标准全生命周期管理。建立完整、高效的内部标准化工作部门，配备专职的标准化工作人员。

建议加强团体标准的推广实施，充分利用会议、论坛、新媒体等多种形式，开展标准宣传、解读、培训等工作，让更多的同行了解团体标准，不断提高行业内对团体标准的认知，促进团体标准推广和实施。

九、其他应予说明的事项

无。

标准起草工作组

2025年10月