ICS 35.030
L80

团　　体　　标　　准

T/ISC 0008—2021

可信数据服务
多方数据价值挖掘体系框架

Trusted data service-Multi-party value mining system framework

2021－02－04 发布　　2021－05－01 实施

中　国　互　联　网　协　会　　发 布

目　　次

前　言

本文件按照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》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由中国互联网协会标准工作委员会（T/ISC）提出并归口。

本文件起草单位：中国信息通信研究院、蚂蚁科技集团股份有限公司、北京市竞天公诚律师事务所、京东云计算有限公司、中国联合网络通信有限公司、同盾科技有限公司、杭州数梦工场科技有限公司、北京三快在线科技有限公司、京东数字科技控股股份有限公司、咪咕文化科技有限公司、山东伏羲智库互联网研究院等。

本文件主要起草人：闫树、袁博、姜春宇、魏凯、吕艾临、李雪妮、冯坚坚、刘朋、周泉、袁立志、孙中伟、张新、黄一珉、刘洪波、贾晓菁、李金夏、陈涛、曲远汶、胡国华、李楷、李丹、刘笑岑、叶串、高雨冰、李然辉、黄琼峰、桂祖宏、于涛。

引　言

2020年4月，中共中央国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，要求加快培育数据要素市场，“加强数据资源整合和安全保护。探索建立统一规范的数据管理制度，提高数据质量和规范性，丰富数据产品。”作为丰富数据产品的必由之路，多方数据价值挖掘将有效促进数据作为生产要素的价值发挥。

目前，对于多方网络运营者的数据流通、共享及如何在保护个人信息的同时更好发挥数据价值、实现数据增值等方面还没有相关立法和国家标准，然而，产业发展却存在着迫切的数据利用、共享的需求。以数据为纽带的产学研融合，数据驱动型创新体系的发展都离不开不同网络运营者在保护个人信息的前提下进行数据融合、分析，发挥数据价值，从而促进产业创新和经济发展。因此，制定网络运营者多方数据价值挖掘的可信体系标准，为不同网络运营者的数据流通、数据价值挖掘提供解决方案，形成一定的行业共识，是对中央政策在行业的具体落实，是对创造数据价值的有益探索，对于未来相关立法和政策制定也有重要的参考意义。

可信数据服务 多方数据价值挖掘体系框架

1 范围

本文件确立了网络运营者针对多方数据价值挖掘的可信体系建设框架，给出了让多方数据价值挖掘达到可信程度的基本要求和保障要求。

本文件适用于网络运营者，为其开展多方数据价值挖掘工作提供参考和指引。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 29828-2013 《信息安全技术 可信计算规范 可信连接架构》

GB/T 30847.1-2014 《系统与软件工程 可信计算平台可信性度量 第1部分概述与词汇》

GB/T 35273-2020 《信息安全技术 个人信息安全规范》

GB/T 37964-2019 《信息安全技术 个人信息去标识化指南》

3 术语和定义

下列术语和定义适用于本文件。

3.1

网络运营者 network operators

网络的所有者、管理者和网络服务提供者。

3.2

可信 trust

信任者对于被信任者关于能否在没有直接管控的情况下履行承诺的一种认知。

3.3

可信性 trustworthiness

被信任者的一种内在的、动态的属性，由被信任者所展示的履行承诺的能力和基于监督证据持续改进的能力所反映。

3.4

可信计算平台 trust computing platform

具有可信保证机制的计算平台。

3.5

可信第三方 trusted third party

一个安全的权威方，它为其他安全相关实体所信任。

注:本文件中的可信第三方是指多方数据价值挖掘区域的运营者，不能由为多方数据价值挖掘提供数据的网络运营者担任，不能为本文件以外的目的使用挖掘区域中的数据和数据价值。

3.6

原始数据 raw data

由不同网络运营者合法控制的数据，通常基于其业务、职能或提供的服务所收集、加工产生。

3.7

数据价值 data value

本文件中的数据价值特指无法反推出原始数据或个人信息的挖掘结果及其所具有的分析、决策和预测意义，同时对于具体的数据价值应用场景具备有效性和可用性，可以提高数据在价值发挥上的广度和深度。具体类型包括但不局限于：聚合统计结果；探查、分析、预测、决策结论；可运算的加密数据；算法模型；策略化建议；无法识别个人的特征标签；经过混淆的不精准数据等。

3.8

数据价值挖掘 data value mining

在对原始数据进行包括去标识化等处理的基础上，对数据进行加工、分析、建模、聚合等处理，挖掘数据背后隐藏的价值，以用于分析、决策、预测等应用场景。

3.9

个人信息 personal information

以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息，包括但不限于自然人的姓名、出生日期、身份证件号码、个人生物识别信息、住址、电话号码等。

3.10

个人信息主体 personal information subject

个人信息所标识或者关联的自然人。

3.11

去标识化 de-identification

通过对个人信息的技术处理，使其在不借助额外信息的情况下，无法识别或者关联个人信息主体的过程。

注:去标识化建立在个体基础之上，保留了个体颗粒度，采用假名、加密、哈希函数等技术手段替代对个人信息的标识。

4 可信体系框架

本文件给出了针对多方数据价值挖掘，建设可信体系的基本框架，其中包括：针对可挖掘数据的预处理、可信挖掘区域建设、数据进入挖掘区域、数据挖掘实施、数据价值输出、数据价值使用、可信体系保障等。

4.1 可信基本原则

可信基本原则包括：数据价值打通而原始数据不打通、保护个人信息主体权益不受侵害、保护对数据拥有合法权利的商业主体正当利益不受侵害、确保数据挖掘区域运营者独立性和可信性、不得滥用数据价值、全链路可审计追溯。

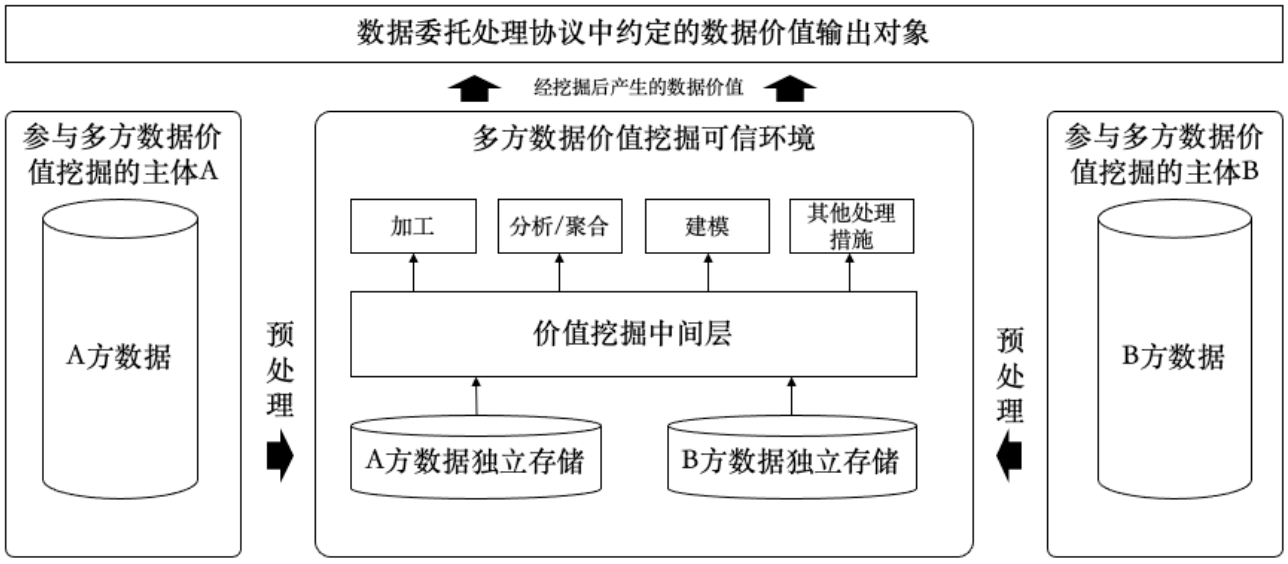
4.2 多方数据价值挖掘的典型需求和可应用场景

典型需求包括：聚合统计、决策分析、数据探查、数据资产建设、模型训练及模型预测、提供策略化建议以及其他无法精准反推出原始数据的数据价值输出场景。上述需求可能出现在政府数据开放、学术科研数据分析、商业数据应用等常见的数据应用场景中。

4.3 多方数据价值挖掘可信体系的基本架构特点

多方数据价值挖掘可信体系的基本架构特点为：数据宽进严出，充分挖掘，最小使用；价值挖掘中间层资产可沉淀复用；数据链路可追溯，行为可审计，风险可管理；多方数据价值挖掘区域为可信计算平台；如果原始数据中包含个人信息，应当在进入价值挖掘区域之前做去标识化处理，或采用隐私计算技术（可信执行环境、多方安全计算、同态加密、功能加密等）对多方数据进行加密处理。

基本架构如下图所示。



注：
预处理：指针对进入数据价值挖掘区域的数据进行清洗、加工、去标识化等措施。

图 1　多方数据价值挖掘体系的基本架构

4.4 可挖掘数据范围

原则上网络运营者收集、产生的数据均可以根据本文件搭建的可信体系进行数据价值的挖掘，但以下数据除外：

——根据国家法律法规或者监管部门规定，禁止进行多方数据价值挖掘的数据；

——没有经过去标识化处理的个人信息；

——来源违法或者违反合同义务约定的数据；

——个人生物识别信息；

——进行数据价值挖掘后可能对国家安全、社会公共利益以及他人的合法权益造成危害的数据。

4.5 确保多方数据价值挖掘可信的关键技术

4.5.1 概述

为确保多方数据价值挖掘的可信性，可采取以下技术能力：个人信息去标识化、全链路列级血缘、未去标识化个人信息和个人生物识别信息识别/使用拦截、数据真实性验证、主体打标、数据分类分级、数据安全网关、全景权限管控、数据流转管控、数据访问风控、动态数据脱敏、智能行为审计、隐私计算。

4.5.2 个人信息去标识化

去标识化是对直接标识符和准标识符进行删除或变换，而且能控制避免被重标识的风险，最后确保去标识化后的数据集可用。

处理标识步骤分为预处理、选择模型技术、实施去标识化三个阶段。

去标识化的技术主要有：加密、加盐哈希、删除、K匿名、L多样性等。

4.5.3 数据安全网关

数据安全网关技术能力包括：

a) 对于网关账户，建立账户安全体系，控制账号接口访问权限；
b) 对于数据的访问，加密记录数据请求日志；
c) 在网关稳定方面，集群需要可弹性扩容并支持超大并发；
d) 在安全方面，支持HTTPS协议，并提供防DDoS/CC攻击等功能。

4.5.4 数据真实性验证

数据真实性验证技术能力包括：

a) 使用区块链、数字签名、电子存证、可信机构的数据认证等，保证数据可验证；
b) 通过使用“零知识证明”的密码学手段，在各方不泄露信息的情况下，对数据真实性进行验证。

4.5.5 数据安全流转管控

详细记录数据流转环节，分析数据流转痕迹，主要使用的办法：

a) 对数据流转事件采用人工审核、机器审核等多种审核机制；
b) 记录数据流转的详细过程，保证数据流转行为可追溯；
c) 可以采用文档加密、数据脱敏等手段，保证数据的安全流转。

4.5.6 数据访问风控

数据访问风控技术能力包括：

a) 控制用户对数据资源的访问；
b) 接入网站应用级入侵防御系统（WAF，即 Web Application Firewall），保障数据安全；
c) 根据数据属性、用户及行为，建立模型计算数据访问风险级别；
d) 进行人机操作识别，拦截非法的数据访问请求。

4.5.7 动态数据脱敏

对时效性要求很高的请求进行数据脱敏和日志脱敏，主要技术有：

a) 数据抽样；
b) 字符子链屏蔽等模糊化处理；

c）屏蔽、局部抑制、记录抑制等抑制技术；
d）噪声添加、置换、微聚集等随机化技术。

4.5.8 隐私计算技术

常见的隐私计算技术有：
a）可信执行环境（TEE）；
b）多方安全计算（MPC）；
c）联邦学习；
d）差分隐私；
e）其他能够在多方数据价值挖掘的具体应用场景中提高隐私安全性、可信性的技术措施。

5 可信体系建设的基本要求

5.1 多方数据价值挖掘区域运营者的可信性要求

多方数据价值挖掘区域运营者应当满足可信性的要求。可信性要求具体包括：
a） 多方数据价值挖掘区域应由可信第三方运营、控制和管理。该可信第三方能够被为多方数据价值挖掘提供数据的网络运营者所信任，从技术和管理上可以防止任何为多方数据价值挖掘提供数据的网络运营者不经可信第三方同意访问导出挖掘区域中的数据。但符合本文件规定的访问或数据价值输出除外。
b） 多方数据价值挖掘区域运营者应与参与多方数据价值挖掘的网络运营者分别签署数据委托处理协议。为了确保挖掘区域运营者的业务中立性和可信任程度，除了基于数据委托处理协议约定，为多个网络运营者提供多方数据价值挖掘之外，挖掘区域运营者不从事其他任何业务，没有其他任何数据来源，既不会做出任何再识别去标识化数据的行为，也不会将挖掘区域内的数据及数据价值用于数据委托处理协议以外的目的。多方数据价值挖掘区域运营者应当根据本文件的要求定期接受可信性审计。
c） 多方数据价值挖掘区域运营者应当具备充分的数据安全能力，严格按照网络安全等级保护制度等履行安全保护义务。

5.2 多方数据价值挖掘的申请流程

任何参与多方数据价值挖掘的网络运营者，应针对其每一个数据价值挖掘需求向挖掘区域运营者发起申请流程。申请中应当包含以下要素：各方参与挖掘的原始数据说明、去标识化情况、具体需求场景、挖掘持续周期、对本文件的遵守情况说明及承诺。挖掘区域运营者应当按照本文件要求核实申请后，方可同意进行多方数据价值挖掘。

5.3 可挖掘数据的可信要求

可挖掘数据在进入挖掘区域之前，参与多方数据价值挖掘的网络运营者应对原始数据进行清洗、加工、去标识化等预处理，尤其对其中的个人信息应当进行去标识化处理或采用隐私计算技术（可信执行环境、多方安全计算、同态加密、功能加密等）对可挖掘数据进行加密，以确保挖掘区域运营者无法重新识别或者关联个人信息主体，或者确保挖掘区域运营者无法将可挖掘数据用于价值挖掘以外的目的。针对进入挖掘区域的数据，挖掘区域应当具备相应的技术和管理能力，确保将不符合本文件4.4款规定的数据识别并拦截在挖掘区域之外。

注：上述去标识化处理不影响挖掘区域运营者利用去标识化后新生成的唯一标识字段进行数据表的融合扩列，以确保多方数据价值可被进一步挖掘。

5.4 挖掘区域中的可信要求

挖掘区域中的可信要求包括数据流转管控、可信应用访问、数据访问风控、用户行为审计。

a） 数据流转管控：挖掘区域应基于保护模式，保证数据只进不出，严格管控。

b） 可信应用访问：基于应用签名+用户签名双签名方式，做到应用未经挖掘区域认证不可访问数据。

c） 数据访问风控：基于数据安全风控能力，进行实时风控识别，并可对风险进行实时的动态脱敏、阻断等处置。

d） 用户行为审计：在挖掘区域中对全链路上的应用行为和数据访问行为进行日志采集，一旦发现数据安全风险，系统可自动发起审计，进行风险追溯和排查。

5.5 挖掘区域输出数据价值的可信要求

挖掘区域仅可输出经挖掘后的数据价值，不得直接输出任何未经聚合统计、模糊化、策略化或者混淆化处理的数据。挖掘区域输出的数据价值对于接收方而言不得构成个人信息，不得危害国家安全、公共利益，不得侵犯他人的合法权益。

5.6 数据价值使用的可信要求

经挖掘后产生的数据价值，仅可用于基于本文件5.2款申请流程中所声明的需求使用场景。参与多方数据价值挖掘并使用数据价值的网络运营者，应当接受针对数据价值使用的可信核查或审计。

6 可信体系保障的要求

6.1 组织人员保障

参与多方数据价值挖掘的网络运营者和挖掘区域运营者均应当配备专门人员负责多方数据价值挖掘的可信体系保障工作。

a） 应明确其法定代表人或主要负责人对多方数据价值挖掘体系的可信性负全面领导责任，包括为可信体系提供人力、财力、物力保障等；

b） 应任命可信体系负责人和可信体系工作机构，可信体系负责人应由具有相关管理工作经历和专业知识的人员担任，有关多方数据价值挖掘的重要决策直接向组织主要负责人报告工作；

c） 可信体系负责人和可信体系工作机构的职责应包括但不限于：

1) 全面统筹实施多方数据价值挖掘的可信体系工作，对多方数据价值挖掘的可信性负直接责任；

2) 组织制定多方数据价值挖掘的可信体系工作计划并督促落实；

3) 制定、签发、实施、定期更新多方数据价值挖掘可信体系相关政策和相关规程；

4) 建立、维护和更新组织已开展的多方数据价值挖掘活动记录和数据价值使用情况；

5) 开展多方数据价值挖掘的可信性评估，提出可信体系建设的对策建议，督促整改可信隐患；

6) 组织开展多方数据价值挖掘的可信体系培训；

7) 负责组织实施内部的可信性审计，定期邀请外部机构进行可信性审计；

8) 与监督、管理部门保持沟通，通报或报告多方数据价值挖掘的相关情况。

d） 应为可信体系负责人和可信体系工作机构提供必要的资源，保障其独立履行职责。

6.2 管理与培训要求

参与多方数据价值挖掘的网络运营者和挖掘区域运营者均应当做好相关管理与培训工作，具体要求为：

a） 应与从事多方数据价值挖掘岗位上的相关人员签署保密协议，对大量接触数据的人员进行背景审查，以了解其犯罪记录、诚信状况等；

b） 应明确内部涉及多方数据价值挖掘不同岗位的可信体系职责，建立惩戒处分机制；

c） 应要求多方数据价值挖掘岗位上的相关人员在调离岗位或终止劳动合同时，继续履行保密义务；

d） 应定期(至少每年一次)对多方数据价值挖掘岗位上的相关人员开展专业化培训和考核。

6.3 多方数据价值挖掘活动记录

参与多方数据价值挖掘的网络运营者和挖掘区域运营者应建立、维护和更新多方数据价值挖掘活动记录，记录的内容可包括:

a） 根据本文件 5.2 款提出的多方数据价值挖掘申请的具体内容；

b） 输出的数据价值使用情况；

c） 与多方数据价值挖掘活动各环节相关的信息系统、组织或人员。

6.4 可信体系审计

对参与多方数据价值挖掘的网络运营者和挖掘区域运营者的要求包括：

a） 应对多方数据价值挖掘的可信体系有效性，依据本文件的要求进行定期（至少每年一次）内部和外部审计；

b） 应建立自动化审计系统，监测记录多方数据价值挖掘活动；

c） 审计过程形成的记录应能对威胁可信性事件的处置、应急响应和事后调查提供支撑；

d） 应防止非授权访问、篡改或删除审计记录；

e） 应及时处理审计过程中发现的违反本文件规定的情况；

f） 审计记录和留存时间应符合法律法规的要求。