ICS 35.020
CCS L 70

T/GZAS

贵州省标准化协会团体标准

T/GZAS 018—2022

数据要素安全可信流通技术标准

Technical standard for secure and trusted flow of data elements

2022－12－29 发布　　2022－12－29 实施

贵州省标准化协会　　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专利的责任。

本文件由贵州省数据流通交易服务中心提出。

本文件由贵州省标准化协会标准化技术委员会归口。

本文件起草单位：贵州省数据流通交易服务中心、中国信息通信研究院、云上贵州大数据产业发展有限公司、多彩贵州印象网络有限公司、中电科大数据研究院有限公司、年华数据科技有限公司、贵阳大数据交易所有限责任公司、贵州商学院、中汇数字创新（贵州）科技服务有限公司。

本文件主要起草人：潘伟杰、陈辉、王佳卫、黄明峰、王仕品、刘军、程序、张婧慧、赵定喜、吕东、徐翼凌、王似巍、周雪卿、李俊男、姚千喜、韩坤洁、陈蔚、张燕、吴明娅、刘润、吴俊、叶玉婷、黄煜、姚昕金、周万青、刘泥君、龙婕、余先昊、杨杰。

数据要素安全可信流通技术标准

1 范围

本文件规定了数据要素安全可信流通过程中数据要素的采集、处理、质量、分类分级、描述、目录、准入、流通方式、流通安全保障和流通安全监管的要求。

本文件适用于数据要素流通工作。

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 11643-1999 公民身份号码

GB/T 25069-2022 信息安全技术 术语

GB/T 34960.5-2018 信息技术服务 治理 第5部分：数据治理规范

GB/T 35273-2020 信息安全技术 个人信息安全规范

GB/T 36343-2018 信息技术 数据交易服务平台 交易数据描述

GB/T 36344-2018 信息技术 数据质量评价指标

GB/T 36625.3-2021 智慧城市 数据融合 第3部分：数据采集规范

GB/T 37932-2019 信息安全技术 数据交易服务安全要求

GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型

GB/T 39477-2020 信息安全技术 政务信息共享 数据安全技术要求

GB/T 41479-2022 信息安全技术 网络数据处理安全要求

JR/T 0197-2020 金融数据安全 数据安全分级指南

DB52/T 1540.6-2021 政务数据 第6部分：安全技术规范

DB52/T 1541.6-2021 政务数据平台 第6部分：面向全网搜索应用的数据处理规范

3 术语和定义

GB/T 25069界定的以及下列术语和定义适用于本文件。

3.1

数据要素 Data elements

由大数据形成的数据要素，既来自个人衣食住行、医疗、社交等行为活动，又来自平台公司、政府、商业机构提供服务后的统计、收集等。

3.2

数据流通 Data flow

对数据进行开发、交易的过程。

3.3

政务数据 Government data

各级行政部门及其技术支撑单位在履行职责过程中依法采集、生成、存储、管理的各类数据资源。

3.4

企业数据 Enterprise data

所有与企业经营相关的信息、资料，包括公司概况、产品信息、经营数据、研究成果等，其中不乏涉及商业机密。

3.5

个人数据 Personal data

以电子或者其他方式记录的能够单独或者与其他数据结合识别特定自然人身份或者反映特定自然人活动情况的各种数据。

3.6

敏感数据 Sensitive data

由权威机构确定的受保护的信息数据。

3.7

数据采集 Data acquisition

指从传感器和其它待测设备等模拟和数字被测单元中自动采集非电量或者电量信号，送到上位机中进行分析、处理。

3.8

数据处理 Data processing

指对数据的采集、存储、检索、加工、变换和传输。

3.9

数据质量 Data quality

数据的一组固有属性满足数据消费者要求的程度。

3.10

数据分类 Data classification

按照公共数据具有的某种共同属性或特征（包括数据对象、重要程度、共享属性、开放属性、应用场景等），采用一定的原则和方法进行区分和归类，以便于管理和使用公共数据。

3.11

数据分级 Date grading

按照公共数据遭到破坏（包括攻击、泄露、篡改、非法使用等）后对国家安全、社会秩序、公共利益以及个人、法人和其他组织的合法权益（受侵害客体）的危害程度对公共数据进行定级，为数据全生命周期管理的安全策略制定提供支撑。

3.12

隐私计算 Privacy computing

在保护数据本身不对外泄露的前提下实现数据分析计算的技术集合，达到对数据“可用、不可见”的目的。

3.13

多方安全计算 Secure Multi-party Computation

指在无可信第三方的情况下，多个参与方共同计算一个目标函数，并且保证每一方仅获取自己的计算结果，无法通过计算过程中的交互数据推测出其他任意一方输入数据的计算方式。

3.14

联邦学习 Federated Learning

是一种分布式机器学习框架，可以做到在保障数据隐私安全及合法合规的基础上，实现数据共享，共同建模。

3.15

可信执行环境 Trusted Execution Environment

通过软硬件方法在中央处理器中构建一个安全区域，保证其内部加载的程序和数据在机密性和完整性上得到保护。

3.16

数据安全 data security

是指通过采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力。

4 数据要数流通过程

数据要素安全可信流通的过程包括以下几个步骤：

a) 数据提供方合理合法的采集数据；

b) 数据提供方对采集的数据进行必要的处理；

c) 数据提供方增加对数据的描述，形成数据目录；

d) 数据提供方将数据目录同步到数据要素安全可信流通平台；

e) 数据要素安全可信流通平台选取多方数据创建数据产品，经过数据审批和模型审定后形成数据产品；

f) 数据要素安全可信流通平台将数据产品推送到隐私计算平台构建任务；

g) 数据要素安全可信流通平台将数据产品发布到各交易门户；

h) 数据需求方通过交易门户购买数据产品并执行计算；

i) 隐私计算平台通过隐私计算技术从数据提供方调用数据进行计算；

j) 模型审定平台对调用数据进行计算的过程进行审定；

k) 数据需求方获取到数据产品的计算结果；

l) 数据流通运营方对数据流通过程进行管理；

m) 数据流通监管方对数据流通过程进行审计。

5 数据要素流通方式

5.1 多方安全计算

多方安全计算（Secure Multi-party Computation，MPC）是指在无可信第三方的情况下，多个参与方共同计算一个目标函数，并且保证每一方仅获取自己的计算结果，无法通过计算过程中的交互数据推测出其他任意一方输入数据的计算方式。多方安全计算在数据要素安全可信流通中的作用如下：

a) 提供秘密共享、不经意传输、混淆电路、差分隐私、同态加密、零知识证明等基于密码学算法的多方安全计算协议；

b) 具有很高的计算安全性，要求数据资源和中间计算结果均不可泄露，实现数据可用不可见；

c) 可与可信执行环境等硬件隐私计算技术结合进一步强化安全性。

5.2 联邦学习

联邦学习（Federated Learning，FL）本质是一种分布式机器学习框架，可以做到在保障数据隐私安全及合法合规的基础上，实现数据共享，共同建模。联邦学习在数据要素安全可信流通中的作用如下：

a) 提供横向联邦学习、纵向联邦学习、迁移联邦学习、模型评估、模型预测等服务；

b) 多个数据源共同参与模型训练时，不需要进行原始数据流转，仅通过交互模型中间参数进行模型联合训练，原始数据可以不出本地；

c) 实现多个机构间构建统一的数据安全、高效、合规的多源数据应用生态系统，实现跨机构的数据共享融合，通过系统扩大样本量、增加数据维度为大数据应用提供高精度模型构建的有力支撑，进而提供更丰富、高质量的大数据服务，为社会发展创造更多价值。

5.3 TEE 可信计算环境

可信执行环境（Trusted Execution Environment，TEE），通过软硬件方法在中央处理器中构建一个安全区域，保证其内部加载的程序和数据在机密性和完整性上得到保护。可信执行环境在数据要素安全可信流通中的作用如下：

a) 将系统的硬件和软件资源划分为两个执行环境：可信执行环境和普通执行环境。两个环境是安全隔离的，有独立的内部数据通路和计算所需存储空间。普通执行环境的应用程序无法访问 TEE，即使在 TEE 内部，多个应用的运行也是相互独立的，不能无授权而互访；
b) 可信执行环境没有对隐私区域内的算法逻辑语言有可计算型方面的限制，支持更多的算子及复杂计算，可实现联合统计、联合查询、联合建模及预测等多种计算，业务表达性强；
c) 利用可信执行环境计算度量功能，可实现身份、数据、算法全流程的计算一致性证明，解释性和逻辑可信度高；
d) 支持多层次、高复杂度的算法逻辑实现，运算效率高；
e) 硬件的可信度是中心化的，芯片设备厂商声誉及产品安全的可信度决定了技术路径的可信度；
f) 为进一步提高安全性，可信执行环境常结合多方安全计算等密码学算法来实现加密。

6 数据要素采集

6.1 采集过程

数据要素采集的过程包括以下几个步骤：

a) 数据源选择：根据需要采集数据的数据源类型（如：文件、数据库、传感器等），确定数据源连接通讯的方式，明确采集标准范围及属性；
b) 数据采集方式选择：数据采集分为人工采集和系统采集两种，通过分析相关数据源类型，根据可操作性、成本导向等原则选定数据采集方式；
c) 数据汇聚：对采集的原始数据进行清洗、转换、分析等处理，确保数据的完整性、准确性和时效性；
d) 数据存储：处理后的数据存储应满足海量、安全、高性能、高可靠、易管理。

6.2 采集方式

数据要素采集的方式包括但不限于以下几种：

a) 针对结构单一、数据量相对较小的结构化数据，可通过数据库表、文件、网络服务（Web Service）、REST、HTTP/HTTPS、消息订阅/发布等技术进行数据采集；
b) 针对传感器、智能手机、PDA 设备、网络等渠道产生的类型丰富、数据量较大的数据，可通过分布式系统接口、分布式流数据收集、网络爬虫等技术进行数据采集；
c) 针对由麦克风、摄像头等设备产生的海量音视频数据，可通过语音图像识别、编解码等技术转化后进行数据采集；
d) 针对问卷调查、实地调研、资料分析等产生的数据，可通过在线填报、离线导入等人工转化方式进行数据采集。

6.3 采集内容

数据要素采集的内容包括但不限于以下几个方面：

a) 基础数据：如人口、法人单位、自然资源、地理空间、宏观经济、电子证照等数据；
b) 专题数据：如房屋、城市部件、网格等公共共享数据；
c) 业务专属数据：如涉及公安、公共卫生和医疗、教育、民政、交通、水利、人力资源和社会保障、市场监管等众多领域的业务数据；
d) 其他数据，如互联网、工业、商业等数据。

6.4 采集要求

数据采集应满足以下要求：

a) 支持全量、历史数据采集：应提供数据传输服务、高并发的离线数据上传下载服务，支持 TB、PB 级别的数据导入（全量数据或历史数据的批量导入）及导出；
b) 支持实时或定时增量数据采集：宜提供实时同步、定时采集、数据订阅、日志采集等服务；
c) 支持条件过滤：按照指定条件进行指定过滤采集，例如字段内容；
d) 支持采集作业管理和调度：采集作业支持条件触发、并发调度、周期循环调度等模式；支持对作业启动、停止、暂停、恢复等操作；
e) 支持数据标签：依据数据清洗要求为数据标记数据标签；
f) 支持数据建模：提供基于不同业务需求进行数据建模功能；
g) 数据采集应具备复杂网络环境下、不同异构数据源之间高速、稳定、弹性伸缩的数据移动及同步能力。

7 数据要素处理

7.1 处理方法

数据要素处理的方法包括但不限于以下几种：

a) 数据格式处理：对错误的数据格式进行治理、修改；
b) 关键信息缺失补全：对需要搜索的关键信息出现缺失的部分进行补全；
c) 明显逻辑错误修正：核查数据元间业务逻辑关系，对明显错误逻辑进行修正；

示例：对数据上下级、归属关系进行修正。

d) 数据类型错误修正：根据搜索目标对数据类型进行判断，对不满足数据类型的错误进行修正。

7.2 处理要求

数据要素处理的要求有以下几个方面：

a) 对时间、区划地点、对象等通用型数据的处理，应满足以下要求：
   1) 行政区划代码应符合 GB/T 2260-2007《中华人民共和国行政区划代码》规定；
   2) 性别代码应符合 GB/T 2261.1-2003《个人基本信息分类与代码 第 1 部分：人的性别代码》规定；

3） 日期和时间应符合 GB/T 7408-2005《数据元和交换格式 信息交换 日期和时间表示法》规定；

4） 公民身份号码应符合 GB 11643-1999《公民身份号码》规定；

5） 法人和其他组织统一社会信用代码应符合 GB 32100-2015《法人和其他组织统一社会信用代码编码规则》规定。

b） 对非通用型的描述类数据处理，应满足以下要求：

1） 可计算型数据元的计算类型，应满足当前系统可连接数据库类型的可计算数据类型的要求，计算单位根据业务进行描述；

2） 可分组维度的数据字段应按其特征进行分组；

3） 可比较数据元应按照不同维度、不同规则进行比较；

4） 主体数据元应对主体的核心内容进行描述。

c） 对指标型数据的处理，应满足以下要求：

1） 指标型数据分类：按照其反映的内容或其数值表现形式，分为总量指标、相对指标和平均指标；

2） 处理原则：对数据中涉及数值与统计数据的指标数据，应按同一指标内部相对差距不变、不同指标间的相对差距不确定、标准化后极大值相等的原则；

3） 处理方法：采用数据同趋化、无量纲化等方法进行处理；

4） 处理内容包括：确定指标的含义和范围；指标指向的对象或现象应具有同类性；有统一的计量单位；两个对比指标要有可比性。

8 数据要素质量

8.1 质量评价指标

数据要素质量的评价指标包括以下几个方面：

a） 规范性：数据符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度；

b） 完整性：按照数据规则要求，数据元素被赋予数值的程度；

c） 准确性：数据准确表示其所描述的真实实体（实际对象）真实值的程度；

d） 一致性：数据与其他特定上下文中使用的数据无矛盾的程度；

e） 时效性：数据在时间变化中的正确程度；

f） 可访问性：数据能被访问的程度。

8.2 质量要求

数据要素安全可信流通应确保数据要素满足以下质量要求：

a） 数据提供方应向数据要素安全可信流通平台提供流通数据获取渠道合法，权利清晰无争议的承诺或证明材料；

b) 数据提供方应向数据要素安全可信流通平台提供拥有流通数据完整相关权益的明确声明；

c) 数据提供方应向数据要素安全可信流通平台提供数据真实性的明确声明；

d) 数据提供方应对流通数据进行分类，并对交易数据进行安全风险评估，出具安全风险评估报告；

e) 数据提供方应明确流通数据的限定用途、使用范围、交易方式和使用期限；

f) 数据提供方应对流通数据进行准确描述，明确数据类别等内容，描述内容满足准确性、真实性要求；

g) 数据要素安全可信流通平台应对流通数据描述和样本的准确性、真实性进行审核；

h) 数据要素安全可信流通平台应对流通数据的安全风险评估报告进行审核，确保数据可交易；

i) 数据要素安全可信流通平台应对流通数据分类结果进行审核。

9 数据要素分类分级

9.1 分类

根据数据要素安全可信流通平台数据流通过程数据来源的特点，将流通数据分为以下几种：

a) 企业数据

企业数据泛指所有与企业经营相关的信息、资料，包括公司概况、产品信息、经营数据、研究成果等，其中不乏涉及商业机密。

商业性企业数据是商业公司负责收集、加工整理并发布的，是作为有价商品进行开发的，所属权归商业公司所有，并由商业公司负责发布销售，其目的是为其他有需求的商业公司提供潜在客户的获取渠道，帮助其他企业开发有效客户，为中小企业提供数据服务。

b) 政务数据

政务数据是指各级行政部门及其技术支撑单位在履行职责过程中依法采集、生成、存储、管理的各类数据资源。

结合政府的职能，不难看出政务数据的独特价值性。随着政务数据的不断开放及规范，政务大数据会在包括城市规划、交通管理、环境保护等多领域被应用。政务数据能够有效的集成各类经济、生态领域的信息资源，帮助政府与民众的沟通建立在科学的数据分析之上，优化公共服务流程、简化公共服务步骤、提升公共服务质量。为政府制定各种决策，提供技术基础和支撑。

c) 个人数据

个人敏感数据是一旦泄露、非法提供或滥用可能危害人身和财产安全，极易导致个人名誉、身心健康受到损害或歧视性待遇等的个人信息。个人敏感信息包括身份证件号码、个人生物识别信息、银行账户、通信记录和内容、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息、14岁以下（含）儿童的个人信息等。

9.2 分级

根据数据要素遭篡改、破坏、泄露或非法利用后，可能带来的潜在影响的范围和程度进行安全分级，其中：

a） 影响范围包括：国家安全，全社会、多个行业、行业内多个组织，单个组织或个人；

b） 影响程度包括：极其严重、严重、中等、轻微、无。

根据公共数据破坏后对国家安全、社会秩序、公共利益以及对公民、法人和其他组织的合法权益（受侵害客体）的危害程度来确定数据的安全级别，共分为5级，由高至低分别为：

a） 高敏感数据（L5 级）；

b） 敏感数据（L4 级）；

c） 较敏感数据（L3 级）；

d） 低敏感数据（L2 级）；

e） 非敏感数据（L1 级）。

详细数据要素级别及分级参考判断标准见表1。

表 1　数据要素分级表

数据级别	敏感程度	敏感判断
L5级	高敏感数据	✦数据要素遭篡改、破坏、泄露或非法利用后，对国家安全造成影响，或对社会秩序、公共利益造成严重影响。 ✦依据国家法律法规和强制性标准或法规规定的特别重要数据，主要用于特定职能部门、特殊岗位的重要业务，只针对特定人员公开，且仅为必须知悉的对象访问或使用的数据。
L4级	敏感数据	✦对社会秩序、公共利益、多个行业、行业内多个组织造成严重影响；对单个组织的正常运作造成极其严重影响；对人身和财产安全、个人名誉造成严重损害。 ✦国家法律法规和强制性标准定义的重要数据，一般只针对特定人员公开，且仅为必须知悉的对象访问或使用。
L3级	较敏感数据	✦对社会秩序、公共利益、多个行业、行业内多个组织造成中等程度的影响；对单个组织的正常运作造成严重影响；对个人名誉造成中等程度的损害。 ✦涉及政府的内部信息，用于一般业务使用，针对受限对象共享或开放的政务数据；仅对受限内部对象共享或开放的企业数据；法律法规明确保护的个人隐私数据。
L2级	低敏感数据	✦对社会秩序、公共利益、多个行业、行业内多个组织造成轻微影响；对单个组织的正常运作造成中等程度或轻微影响；对个人的合法权益造成轻微损害。 ✦经规定程序审核后，可以向社会公开的政务数据；用于一般业务使用，对受限对象共享或开放的企业数据和个人数据。
L1级	非敏感数据	✦对国家安全、社会秩序、公共利益、行业发展、信息主体均无影响。 ✦已经被政府、企业或个人明示公开或主动披露的数据，可从公开渠道获取。

10　数据要素描述

10.1　基本描述

数据要素的基本描述包括但不限于以下几个方面：

a） 数据编号：流通数据在数据要素安全可信流通平台的唯一编号；

b） 数据名称：流通数据在数据要素安全可信流通平台发布时采用的名称，突出数据集内容、特点、产生时间等；

c） 数据来源：数据来源的机构或部门；

d) 系统名称：数据来源的系统名称；

e) 数据库实例：数据来源的数据库实例；

f) 字段分类：数据字段的分类，包括系统字段、普通业务字段、核心业务字段等；

g) 数据字段：数据中包含的字段中、英文内容；

h) 关键词：描述流通数据内容的关键词语，多个关键词以空格隔开；

i) 所属行业：流通数据所属的国民经济行业的行业名称；

j) 数据种类：流通数据的种类；

k) 数据规模：流通数据占据的存储空间大小、记录条数、吞吐量等；

l) 数据存储格式：流通数据的存储格式（如：TXT、EXCEL 等）；

m) 采集时间：流通数据的采集时间，可以是单一时间也可以是时间区间；

n) 数据发布时间：流通数据在数据要素安全可信流通平台的发布时间，格式 YYYY-MM-DDHH:MM:SS；

o) 数据质量：数据的质量水平（可通过提供数据需求方所要求的数据质量证明文件来说明）；

p) 更新频度：流通数据更新的时间间隔，如每隔 24h 更新一次数据；

q) 更新方式：流通数据的更新方式等（修改更新、增量更新）。

10.2 业务描述

数据要素的业务描述包括但不限于以下几个方面：

a) 数据价格：流通数据的流通价格。可以按流通数据记录的条数、使用次数、数据文件包大小、数据提供方式等制定价格；

b) 数据计费方式：流通数据的计费方式。可按数据使用量、使用时长等计费；

c) 流通方式：需求方获取流通数据价值采用的方式；

d) 提供方权属范围：数据提供方对流通数据的权属范围（数据所有权、收益权、使用权等）；

e) 需求方权属范围：数据需求方对流通数据的权属范围（数据使用权、收益权、所有权等）。

10.3 扩展描述

数据要素的扩展描述包括但不限于以下几个方面：

a) 数据要素安全属性：数据要素的分级情况，包括 L5 级、L4 级、L3 级、L2 级、L1 级；

b) 数据要素授权属性：包括批量授权、全量授权、单次授权等；

c) 数据要素对接方式：包括库表、接口、文件、线下拷贝等；

d) 数据要素加工方式：包括加密、脱敏、抑制、假名化、泛化、随机化、混淆、接口编排等；

e) 数据要素流通方式：加密传输、隐私计算、明文计算等；

f) 数据要素使用属性：包括输入参数查询、核验、直接下载、接口调用等；

g) 数据要素服务属性：批量服务、单次服务、校验服务、建模服务等。

11 数据要素目录

数据要素目录是指数据提供方按照一定的分类方法，对数据资源进行排序、编码、描述，便于检索、定位与获取信息资源，形成的描述数据资源基本信息、业务信息和扩展信息的数据。数据要素目录应满足以下要求：

a) 数据提供方应将数据要素目录同步到数据要素安全可信流通平台进行展示；
b) 应按照数据类别或主题形成数据要素目录；
c) 应定义数据要素目录对应数据资源的内容、安全分级与传输方式；
d) 应对数据要素目录发布进行审核，检查资源目录的规范性、准确性；
e) 在数据要素目录发布过程中，应对数据提供方进行身份鉴别；
f) 应对数据要素目录发布过程进行详细记录，包括发布日期和时间、发布人、审批人、发布资源详细内容等；
g) 应对数据要素目录类型变更、目录迁移等操作进行授权审计；
h) 应保证数据要素目录在传输过程中信息的保密性和完整性。

12 数据要素准入

12.1 可流通数据

数据要素安全可信流通平台可流通的数据包括：

a) 政府、企业或个人明示公开或主动披露的非敏感数据；
b) 企业愿意进行流通的，且不会对社会稳定和国家安全造成危害的企业非敏感数据；
c) 非敏感的个人数据；
d) 经过脱敏处理，且评估可以进行流通的政务数据；
e) 经过脱敏处理，且评估可以进行流通的企业敏感数据；
f) 经过脱敏处理，且评估可以进行流通的个人敏感数据。

12.2 不可流通数据

数据要素安全可信流通平台不可流通的数据包括：

a) 数据遭篡改、破坏、泄露或非法利用后，会对国家安全和社会稳定造成危害的高敏感数据；
b) 不可向公众开放的政务数据；
c) 敏感个人数据；
d) 涉及他人知识产权和商业秘密等权利的数据；
e) 从非法或违规渠道获取的数据；
f) 与原供方所签订的合约要求禁止转售或公开的数据；
g) 其他法律法规明确禁止流通的数据。

13 安全保障

13.1 基础设施安全

数据要素流通的基础设施安全要求有以下几个方面：

a) 支撑数据要素安全可信流通的机房、基础网络、云平台系统等基础设施的通用安全应符合等级保护三级安全要求；

b) 基础网络除通用要求外，电子政务外网用户还应符合国家电子政务外网安全标准。

13.2 数据采集安全

数据采集过程中应全方位防御，避免病毒、攻击、非授权的访问与内部泄密，同时应保障访问记录的审查和监督。应包括但不限于以下几个方面：

a) 对不同数据进行分类并标识，采用安全技术进行安全维护；

b) 监控数据使用情况，防止数据在采集过程中被非法访问、破坏、篡改、丢失、阻止；

c) 设立访问和使用权限控制机制；

d) 制定应急响应预案及相应处理措施，并定期进行应急演练，及时发现安全问题并处理；

e) 定期对数据采集的安全性进行风险评估，并据此制定相应的风险处理计划，及时排查安全漏洞，加固安全技术；

f) 采用安全技术维护数据安全，包括但不限于对称与非对称密码技术及其硬化技术、VPN 技术、身份认证与鉴别技术、CPK 技术、CCKS 技术、PKI 技术、完整性验证技术、数字签名技术、秘密共享技术等；

g) 制定数据采集操作规程，规范数据采集的数据格式、数据质量、流程和方法等；

h) 制定数据采集原则，明确采集数据的目的和用途，确保数据采集的合法性和正当性；

i) 建立安全管理规范，避免人为因素导致数据泄露、损坏等安全事故。

13.3 数据存储安全

数据要素安全可信流通对数据存储的安全要求包括：

a) 应对数据存储环境进行分域分级设计；

b) 应根据数据重要性、量级、使用频率等因素将数据分域分级存储；

c) 应对敏感数据分布式存储；

d) 宜对敏感数据设置在线双活或多活存储机制；

e) 应按照 GB/T 35273-2020 的要求存储个人信息，防止个人信息通过关联分析等技术手段被恢复；

f) 应在存储个人生物识别特征信息时，按照 GB/T 35273-2020 的要求采用技术措施确保信息安全后再进行存储，例如仅存储个人生物识别特征信息的摘要；

g) 应建立数据冗余一致性校验策略；

h) 应对访问用户进行身份鉴别和权限控制，并对用户权限变更进行审核并记录；

i) 应为存储系统安全管理员提供用户标识与鉴别策略、数据访问控制策略，包括访问控制时效的管理和验证，以及接入数据存储的合法性和安全性认证；

j) 应严格限制批量修改、拷贝、下载等操作的权限；

k） 应提供控制机制限制获得访问权的用户将数据传递给非授权的用户；

l） 应对访问通道进行授权许可和访问方式限制；

m） 应建立敏感数据防护区域或敏感数据集群管控访问方式；

n） 应具备数据泄露的发现、阻断等安全机制；

o） 应进行数据血缘关系梳理，建立数字表字段级的上下游关系，建立不同数据源数据合并的分析、核对机制。

13.4 数据传输安全

数据要素安全可信流通对数据传输的安全要求包括：

a） 应采用符合 GM/T 0054 等国家相关标准规定的密码技术，保证通信过程中数据的保密性和完整性；

b） 应具备监控数据传输过程的能力，发现问题时及时告警并进行阻断；

c） 应在数据交换不完整时清除传输缓存数据；

d） 应在交换完成后清除传输历史缓存数据；

e） 应定期检查或评估数据传输的安全性和可靠性。

13.5 数据流通安全

数据要素安全可信流通过程的安全包括：身份鉴别、访问控制、授权管理、数据脱敏、数据防泄漏等。

a） 数据流通过程中对身份鉴别的安全要求包括：

1） 应对访问数据处理系统、服务器操作系统、数据库系统、备份系统的管理员进行身份鉴别；

2） 应建立用户口令长度、口令生存周期、口令复杂度等口令管理策略，保证基于口令的身份鉴别安全性；

3） 应对敏感数据或重要模块的操作复合采用两种或两种以上的鉴别技术进行身份认证。

b） 数据流通过程中对访问控制的安全要求包括：

1） 应针对服务器系统、数据库系统等重要系统设置用户访问控制策略，为不同用户授予其完成各自承担任务所需的最小权限，限制超级管理员等默认角色；

2） 应及时清除系统中无用账号、默认账号，杜绝多人共用同一个系统账号的情况；

3） 用户和管理员账号应采用实名认证，实现追责溯源：

4） 应阻断对数据、应用、系统等的任何非授权访问，提出告警并记录审计日志；

5） 应限制对重要服务器的远程管理，若需要远程管理时应采用 SSH 等安全方式实现；

6） 应只开启业务所需的最少系统服务及端口，并定期核查。

c） 数据流通过程中对授权管理的安全要求包括：

1） 应明确授权目的和范围，保留授权记录，并遵照授权执行；

2） 应采用技术措施防止数据受到未授权的使用；

3） 对敏感数据的使用应经过二次授权，并进行授权审计。

d） 数据流通过程中对数据脱敏的安全要求包括：

1） 应根据不同的业务、应用、部门等采用不同的数据脱敏方式对数据处理过程中产生的敏感数据进行数据脱敏；

2） 应实现动态适配不同数据类型的数据脱敏机制；

3） 应建立对敏感数据脱敏有效性的评价机制，实现效果量化管理。

e） 数据流通过程中应建立共享数据业务的数据透明加密处理能力。

f） 数据流通过程中对数据防泄漏的安全要求包括：

1） 应按数据分级分类预先对每类数据设置访问策略、传播策略和传播范围等；

2） 应采取技术措施防止所有数据在未授权条件下的下载、复制、截屏等方式的数据输出，同时应采取措施防止敏感数据泄露；

3） 应禁止数据处理过程中调试信息的输出；

4） 应防止数据处理过程中日志记录数据的泄露。

13.6 个人数据安全

数据要素安全可信流通平台应确保数据流通在个人数据安全保护方面满足以下要求：

a） 满足 GB/T 35273-2020 中关于个人数据的委托处理、共享、转让、公开披露安全要求；

b） 要求数据提供方对流通数据进行个人数据安全风险评估，提供个人数据安全风险评估报告；

c） 数据对个人数据安全风险评估报告进行审核，确保数据可流通。

14 安全监管

14.1 区块链存证

数据要素流通过程产生的数据需要在区块链平台进行上链存证，防止恶意篡改和行为抵赖。

a） 对用户的注册信息、登录信息、机构信息、审批信息等进行上链存证；

b） 对数据要素目录的增加、修改、删除等信息进行上链存证；

c） 对需求任务信息进行上链存证；

d） 对算法的创建、修改、使用、删除等进行上链存证；

e） 对数据产品创建信息、审批信息、模型审定信息、授权信息、发布信息、上架信息、购买信息、使用信息、结果获取信息等进行上链存证；

f） 对数据要素的调用信息进行上链存证。

14.2 模型审定

数据流通过程中把所使用到的数据、算法提到模型审定平台，由模型审定平台对流通数据和算法进行检查、审定，确保数据流通安全、合规。对数据要素安全可信流通过程的模型审定内容包括：

14.2.1 模型上线审定

模型上线审定流程包括以下几个步骤：

a) 完备性审定

1) 核心信息完整检查：模型标题是否为空；模型描述是否为空；运算开始时间、运算结束时间是否为空；

2) 核心文档完整检查：功能测试报告是否已提供；性能测试报告是否已提供；上线程序是否已提供。

b) 合法性审定

数据需求检查：数据需求申请记录是否存在；数据需求申请快照文档是否存在。

c) 合规性审定

模型设计规范检查：出参字段是否归属敏感信息；运算时间是否在容许范围；原始数据是否涉及敏感信息；该授权运营方下的所有申请的原始数据组合后是否为敏感信息。

14.2.2 模型运行审定

模型运行审定包括以下几个步骤：

a) 标准审定

1) 运算输出参数个数与模型设计是否一致；

2) 运算输出参数类型与模型设计是否一致；

3) 运算输出参数值长度与模型设计是否一致；

4) 运算输出参数值是否在模型设计枚举值范围内；

5) 运算时间是否在设计的运算时间范围内；

6) 模型运行状态与模型实际运行情况是否一致。

b) 质量审定

1) 及时性检查：运算耗时是否超过设置阈值；

2) 波动性检查：运算输出参数值环比波动是否超过设置阈值；运算输出数据量环比波动是否超过设置阈值；

3) 准确性检查：运算输出字段姓名长度大于 5 位，数据可能不正确；运算输出字段姓名出现数字、字母，数据可能不正确；运算输出字段姓名首字母不在百家姓清单，数据可能不正确；运算输出字段地市不在全国地市清单中，数据可能不正确。

c) 安全审定

1) 运算时间是否在设定禁止时间内；

2) 运算时间是否在设计时间内；

3) 运算输出数据是否涉及身份证；

4) 运算输出数据是否可能涉及户口簿；

5） 运算输出数据是否可能涉及护照；

6） 运算输出数据是否涉及军官证；

7） 运算输出数据是否涉及港澳居民通行证；

8） 运算输出数据是否涉及驾驶证；

9） 运算输出数据是否涉及手机号码；

10）运算输出数据是否涉及电子邮箱；

11）运算输出数据是否涉及地址信息；

12）运算输出数据是否可能涉及家庭关系信息，如出现“父亲”、“女儿”等关键字；

13）运算输出数据是否可能涉及银行卡信息；

14）运算输出数据是否可能涉及收入信息，如出现“收入”、“工资”、“奖金”等字符；

15）运算输出数据中 ANSIC 码数量超过设定数量阈值（用于发现可能有加密情况）；

16）运算输出数据是否可能组合出身份证（输出数据同时发现有户籍地址、性别、出生日期）。

14.3 全流程安全监管

对数据要素流通全部流程进行安全监管，随时掌握数据要素流通的整体安全状况，做到数据安全及时感知、安全事件的及时追踪。全流程安全监管的要求有以下几个方面：

a） 数据要素流通过程要满足国家及省、市相关法律法规、行政命令、政策和标准要求；

b） 数据要素流通参与各方以确定的在线方式，向安全监管方提交支撑数据要素流通监管活动的相关数据；

c） 安全监管方应通过平台或系统，实现相关监管内容的自动分析处理与判断；

d） 数据要素流通参与各方提交的日志，日志内容应包括但不限于事件主体、事件客体、起止时间，操作内容和操作结果等元素；

e） 开展等级保护测评、等级保护备案、个人信息保护规范等安全检查，确保数据要素流通各参与方落实安全合规建设；

f） 分析评估各平台系统特权账号操作日志记录，重点监管审计数据库特权账号对数据要素资源的增、删、改、查；

g） 结合国家法规标准与数据分类分级要求，判定数据要素分类分级设置合法性与合理性；

h） 对数据要素处理、传输、分析等数据操作行为进行日志分析与流量分析审核，确保敏感数据的使用符合申请需求；

i） 确保数据要素落实相应等级的传输加密保护，必要时可对安全通道方案、身份鉴别和认证机制、数据加密算法等进行评估审核；

j） 确保数据要素落实相应等级的存储加密手段，必要时可对数据库配置、数据加密算法等进行评估审核；

k） 对敏感数据的访问行为进行分析审计，确保敏感数据调用合法合规。

参 考 文 献

[1] 中华人民共和国网络安全法

[2] 中华人民共和国数据安全法

[3] 中华人民共和国个人信息保护法

[4] DB2201/T 17-2022　政务数据安全分类分级指南

[5] DB3301/T 0322.1-2020　数据资源管理 第1部分：政务数据安全监管

[6] TC260-PG-20212A《网络安全标准实践指南—网络数据分类分级指引》