ICS 35.020
CCS L 70

团 体 标 准

T/ISC 0047—2024

数据流通备份与审查技术通则

General principles for backup and review techniques of data circulation

2024-06-12 发布 2024-07-12 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归口。

本文件起草单位：华东师范大学、北京大学长沙计算与数字经济研究院、清华大学丘成桐数学科学中心、运易通科技有限公司、上海汇付支付有限公司、上海工业自动化仪表研究院有限公司、数力聚（北京）科技有限公司、中原工学院、国科华创认证有限责任公司、智研高科（北京）信息技术发展有限公司、天翼电子商务有限公司。

本文件主要起草人：杨艳琴、金澈清、文伟平、尹昊、丁津泰、刘海峰、刘利、梁星元、李艳丽、周红福、赵浩延、李乐平、杨晨光、付昆、邵奇峰、杨要科、杨小琴、张礼、喻博。

引　言

数据是与土地、劳动力、资本、技术并列的生产要素之一。数据流通是指以数据作为流通对象，按照一定规则从数据提供方传递到数据需求方的过程，即数据资源先后被不同主体获取、掌握或利用的过程。数据通过流通和共享产生了更大的经济价值，为保障数据的质量和安全，数据的流通共享需要被有效监管。数据流通备份审查机制，针对数据流通过程中的重要数据备份进行审查，确保数据的整体质量、合法合规和保护数据的安全，支持数据的恢复，避免数据的泄露。本文件通过对数据流通备份和审查通用技术要求的定义，对规范化的数据流通和交易具有指导意义。

数据流通备份与审查技术通则

1 范围

本文件规定了数据流通备份审查通用技术要求，包括数据流通备份通用要求和数据流通审查技术要求。

本文件适用于政府部门、交易平台、第三方机构的监管、评估、审查。

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 22239-2019 信息安全技术 网络安全等级保护基本要求

GB/T 43697-2024 数据安全技术 数据分类分级规则

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据流通 data circulation

以数据作为流通对象，按照一定规则从数据提供方传递到数据需求方的过程。

3.2

数据流通备份 data circulation backup

将数据流通过程涉及的核心数据与关键操作进行备份，以对数据流通全流程的持续审查提供技术支撑。

注：数据流通中的数据既指传统数据和大数据，也指对其加工处理后的衍生数据。与用于数据恢复的传统数据备份不同，本文件中的数据备份着重用于数据流通过程中的数据审查。

3.3

数据审查 data review

数据监管方对流通数据的数据质量和违法违规问题进行检测评估，以实现事前评估或事后追责。

3.4

数据完整性 data integrity

数据的内容在整个流通过程中始终是正确的、未被篡改的。

3.5

数据一致性 data consistency

数据库或分布式系统多个节点存储的数据副本之间保持了相同的状态和值，以确保数据的准确和可靠，避免数据的丢失或错误。

4 概述

本文件规定了数据流通备份与审查活动的行为，主要涉及数据交易、数据公开、数据共享等场景。数据交易是数据提供方和数据需求方之间以数据商品为交易对象，以货币或其等价物来交换数据商品的行为，典型场景如医疗数据交易。数据公开是数据提供方因社会职责要求而非营利性地公开数据以供数据需求方使用的行为，典型场景如政府数据公开。数据共享是数据提供方以共享己方生产数据为代价而换取其他数据提供方生产数据的行为，典型场景如行业数据共享。作为数据提供方的企业、机构或组织，有责任对流通数据进行备份以接受监管方的评估和审查。

本文件包含数据流通备份通用要求和数据流通审查技术要求两部分。其中，数据流通备份通用要求规定了与数据流通备份相关的对象和媒体、频率和时间、完整性和一致性、数量和可恢复性等内容。数据流通审查技术要求指明了与数据流通备份相关的数据来源、数据质量、数据描述、数据完整性等审查技术。规范化备份的目的是为了实现对备份数据的规范化审查。备份的目的是保障流通数据可靠存证，审查的目的是保障备份存证真实有效。因此，数据流通审查技术要求是对数据流通备份通用要求的补充拓展。

5 数据流通备份通用要求

5.1 备份对象和媒体

5.1.1 备份对象

备份对象是需要被备份的数据集合，应满足以下要求：

a） 备份对象要根据流通审查的数据重要性进行区分，有行业监管标准的按行业监管标准进行区分，没有行业监管标准的按数据对企业的经济影响和社会影响程度来确定；

b） 备份对象要以量化形式确定数据的分级标准，按GB/T 43697-2024的要求执行；

c） 备份对象需要包含数据关联性；

d） 备份对象包括但不限于结构化数据，半结构化数据，非结构化数据及混合模式数据。

5.1.2 备份媒体

备份媒体也称为备份介质，是存放备份数据的物理载体，应满足以下要求：

a) 备份媒体包括但不限于磁带设备、通用NAS存储、物理或虚拟带库、光存储、分布式对象存储等；

b) 备份媒体可以根据审查备份数据特性进行选择；

c) 备份媒体的选择因素包括但不限于备份效率、空间及成本；

d) 备份媒体存取效率影响数据的恢复目标，备份媒体的空间要求及扩展要求与审查备份对象的量级相关，备份媒体是否可以使用公有云资源由审查备份对象的重要性和安全性决定；

e) 备份媒体存放环境及物理基础设施的安全保护等级按GB/T 22239-2019的要求执行。

5.2 备份时间和频率

5.2.1 备份时间

备份时间是备份的时间点，备份时间的设定需保证监管可对指定版本的数据进行审查，应满足以下要求：

a） 备份时间需要结合审查备份对象的量级、备份对象的具体类型、极端条件下对数据恢复时间及数据丢失量的容忍程度、数据备份平台及备份媒体的性能特性；
b） 备份时间需要结合备份策略，备份策略根据备份对象等级分为全量备份、增量备份、差量备份或混合备份；
c） 备份时间通过备份作业设置，备份的作业时间需在规定的时间窗口完成，根据备份作业实际情况，动态调整备份时间窗口及离线归档；
d） 根据审查要求确定备份归档及保存时间。

5.2.2 备份频率

备份频率是备份的时间间隔，备份频率的设定需保证监管可对指定版本的数据进行审查，应满足以下要求：

a） 备份频率需要结合备份对象的量级、变化速度及待备份系统的繁忙程度进行设置；
b） 备份频率通过备份作业设置，备份的作业需在设置的频率下正常触发，如果有中断，需要有补发机制保证备份作业正常执行；
c） 备份频率可以根据审查对象变化动态调整。

5.3 备份完整性和一致性

5.3.1 备份完整性

备份完整性对流通数据生成备份副本的过程进行了规范和保障，备份完整性确保备份未被篡改，满足以下要求：

a） 备份的流通数据应包含必要的元数据描述，包括但不限于备份名称、备份总数、备份创建者、备份创建日期、备份数据容量等数据字段；
b） 备份的流通数据若应用了数据压缩技术，应确保被压缩数据能通过相应压缩软件的完整性检测，未出现损坏或缺失报错；
c） 备份的流通数据应能通过基本数据格式验证并成功被相应的业务程序解析，包括但不限于txt、json、xls、xml、csv等常用的数据存储格式；
d） 备份的流通数据应包含至少1个完整性校验码，宜采用国家密码局推荐的国密哈希算法SM3，可与原始数据的校验码对比以检查备份内容是否被篡改；
e） 备份的流通数据应包含备份者的数字签名，宜采用国家密码局推荐的国密签名算法SM2，用以保障对创建备份副本的不可否认和不可抵赖。

5.3.2 备份一致性

备份一致性对流通数据生成备份副本的过程进行了规范和保障，以确保备份数据与原始数据的数据一致性，应满足以下要求：

a） 确保任一备份副本内容均与数据提供方提供的原始数据内容保持一致，可采用完整性校验码检验数据一致性；
b） 确保多个备份副本之间除数据描述外的备份内容保持一致，可通过多种不同的哈希算法计算哈希值进行比对，可采用国家密码局推荐的国密哈希算法以及国际通用的标准哈希算法；
c） 确保多个备份副本采用的数据存储方式保持一致，例如内容的数据格式、压缩的算法类型等；
d） 确保备份元数据描述中的备份总数与实际备份副本数量保持一致；

e） 确保能够成功验证备份副本数字签名的公钥与备份方提供身份认证的公钥保持一致。

5.4 备份数量和可恢复性

5.4.1 备份数量

备份数量对利用备份副本还原原始数据的过程进行了规范和保障，备份数量确保有足够冗余，应满足以下要求：

a） 确保流通数据至少拥有2份以上的备份副本，用于做相互验证以增加备份鲁棒性，确保备份数据能快速恢复与应对区域级的故障；

b） 确保流通数据与备份副本存储在不同机器，以避免单点故障导致原始数据和备份副本同时丢失，在合法合规前提下，也可考虑采用云存储托管备份副本；

c） 确保多个备份副本中至少有1份存储在不同城区的物理机房，以防止自然灾害等不可抗力因素导致的备份副本完全丢失，在合法合规前提下，可考虑在云存储中选取不同的地理位置。

5.4.2 可恢复性

可恢复性对利用备份副本还原原始数据的过程进行了规范和保障，可恢复性确保能够正确还原原始数据，应满足以下要求：

a） 当原始数据损坏或丢失时，可通过多备份副本校验完整性并修复还原原始数据；

b） 当检查到部分备份副本损坏导致不一致时，可通过其他备份副本校验完整性并修复被损坏的备份副本；

c） 当检查到部分备份副本丢失时，可通过其他备份副本校验完整性并还原被丢失的备份副本；

d） 当备份副本内容被数据压缩时，可解压缩还原备份副本内容；

e） 当流通数据内容被数据压缩时，可解压缩还原流通数据内容。

6 数据流通审查技术要求

6.1 数据来源审查

6.1.1 数据来源验证

数据来源验证用于验证流通数据来源的合法性，审查数据采集过程是否合规。

a） 文档审查。文档审查是识别数据来源的重要方法，应满足以下要求：

1） 获取相关文档：收集与数据来源有关的所有文件，包括数据采集方法、数据处理流程、数据存档等；

2） 分析文档：分析文档以了解数据来源的性质、用途、数据生成过程和质量控制措施；

3） 检查数据源标识：确保文档中包含有关数据来源的明确标识，包括来源的名称、负责机构和联系信息；

4） 评估数据处理流程：审查数据处理步骤，确认数据在采集和处理过程中是否受到适当的管理和监督。

b） 采样审查。采样审查适用于大规模数据源，应满足以下要求：

1） 制定采样计划：明确定义采样的范围、目标和方法，确保采样具有代表性；

2） 数据采集：从数据来源中随机或系统性地选择样本以获取数据的代表性快照；

3） 数据验证：验证采样数据的准确性和一致性，以确定数据来源可靠；

4） 确认来源标识：确保每个样本都有明确的来源标识以便追溯数据。

c） 数据供应商审查。如果数据来源是数据供应商，应满足以下要求：

1） 供应商背景调查：研究供应商的信誉、历史和专业背景，以确定其可信度；
2） 合同审查：审查合同以确保合同中明确定义了数据的质量、交付和支持要求；
3） 绩效监控：建立供应商绩效监控机制以定期评估数据质量和供应商的服务；
4） 风险管理：评估供应商引入的风险，包括数据不一致性、数据泄露和数据安全风险，采取适当的风险管理措施。

6.1.2 数据可信度评估

数据可信度评估用以审查流通数据来源的可信度，以对数据的质量和准确性进行评估。

a） 数据比对与验证。数据比对与验证将备份数据与原始数据进行比对以检验数据质量，应满足以下要求：
1） 参考数据：获得来自可信数据源的参考数据，检验参考数据的准确性和可信性；
2） 数据比对：将要审查的数据与参考数据进行比对，检查数据之间的差异性和不一致性；
3） 验证准确性：确定数据的准确性，识别任何不一致性和错误。

b） 专家评估。专家评估是一种系统性的评估方法，依赖于领域专家或数据管理专家的经验和判断，应满足以下要求：
1） 专家资格：招募具有相关领域知识和经验的专家，确保专家组成员具备可靠的专业背景和行业经验；
2） 评估标准制定：制定清晰的评估标准和指导原则，以帮助专家评估数据可信度。评估标准应涵盖数据来源的可靠性、数据收集和处理的合规性、数据完整性等方面；
3） 数据审查：专家组对数据进行全面审查，考虑数据的来源、采集方法、处理过程和相关文档。审查过程应注重发现数据质量问题、潜在的误差或不一致性；
4） 专家评分：专家根据评估标准和经验，对数据的可信度进行评分。评分应采用一致的标准和严谨的方法进行，确保评估结果的客观性和准确性。

6.2 数据质量审查

数据质量审查是指在数据流通过程中实施数据有效性检查，确保数据发挥应有的流通价值，应满足以下要求：

a） 数据容量检查：获取流通数据的容量尺寸，数据容量过大则不利于数据传输，可考虑拆分成多个数据包，确保数据流转顺利；
b） 数据格式检查：获取流通数据的文件格式，检查数据内容与格式是否匹配，结合数据格式判断数据类型是否合适，是否支持通用的文件格式或能以兼容模式读取；
c） 数据约束检查：读取流通数据内容，检查数据值是否满足对应约束，如整型是否超出界限、字符型是否包含特殊字符、布尔型是否满足规范等；
d） 数据字段检查：读取流通数据内容，检查是否存在空白字段，是否存在多余字段，以及字段数量与描述是否不匹配，确保数据定义有效；
e） 数据条目检查：利用统计和分析手段，检查数据是否包含重复条目，是否有不满足约束检查和字段检查的条目，通过人工比对发现无用条目；
f） 数据时间检查：从元数据中读取流通数据的创建时间、不同版本的修改时间、数据有效的截止时间等，检查当前时间是否在有效范围内。

6.3 数据描述审查

数据描述审查用以确保数据集的元数据与描述信息足够详细，使需求方能够理解和有效使用数据，应满足以下要求：

a） 元数据检查：检查数据集的元数据，包括数据类型、字段名称、单位等信息，验证元数据是否准确，是否与数据内容相匹配；

b） 描述清晰度检查：评估数据描述的清晰度和易读性，确保数据集的描述信息包括适当的背景知识和上下文信息；

c） 数据质量检查：确保数据描述中包括数据质量信息，如数据完整性、准确性和更新频率等，验证数据质量信息是否与实际数据质量一致；

d） 分类分级信息检查。对来自不同业务场景的数据（如：研发、生产、运维、管理等）建立验证机制，确定其分类信息；并通过数据的领域、群体、区域、精度、规模、深度、覆盖度、重要性等指标，评估其分级信息；以能够限定相关人员的访问权限，防止非授权访问。

6.4 数据完整性审查

数据完整性审查是指在数据流通过程中实施可用性检查，确保流通数据整体质量，应满足以下要求：

a） 数据描述检查：读取流通数据提供的元数据内容，检查相应字段是否与数据匹配，如数据条目数量、数据存储位置、数据文件类型等；

b） 数据压缩检查：若应用了数据压缩技术，需检查数据是否通过相应压缩软件的完整性测试；

c） 数据校验检查：读取流通数据提供的校验码，根据描述信息采用相应的函数计算校验码并检查是否与数据备份时提交的校验码匹配；

d） 数据签名检查：读取流通数据提供的签名信息，检查公钥提供者身份是否与描述信息匹配，执行相应的验证算法判定签名是否有效；

e） 数据备份检查：读取流通数据提供的备份信息，检查备份的副本数量和存储位置是否与描述信息匹配，根据备份的完整性和一致性要求检查备份是否可用。

6.5 数据安全审查

数据安全审查用以确保流通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，应满足以下要求：

a） 数据识别：识别流通数据中涉及的敏感数据，包括个人信息、重要数据和其他数据，形成数据保护目录，并及时更新；

b） 分类分级：按照相关国家标准、合同规定和业务运营需要，对所识别的数据按照行业领域以及本身业务属性完成分类，并根据其在具体业务场景中的重要程度完成分级；

c） 风险防控：对流通数据应履行数据安全保护义务，加强风险监测，应采取加密、脱敏、访问控制、审计等技术或者其他必要措施，保护数据免受泄露、窃取、篡改、损毁、不正当使用；

d） 审计追溯：对流通数据的全生命周期进行记录，确保流通过程可审计、可追溯；

e） 隐私技术审查：结合密码学及隐私计算前沿技术（如：安全多方计算、同态加密、可搜索加密、可信执行环境TEE、后量子密码等），对涉及国家利益、公共安全、商业秘密、个人隐私等重要数据实现隐私保护，但需要对相应技术及其保护结果的安全性、有效性和可靠性进行审查和验证。

参 考 文 献

[1] GB/T 20984 信息安全技术 信息安全风险评估方法
[2] GB/T 29765—2021 信息安全技术 数据备份与恢复产品技术要求与测试评价方法
[3] GB/T 35273—2020 信息安全技术 个人信息安全规范
[4] GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型
[5] GB/T 39335—2020 信息安全技术 个人信息安全影响评估指南
[6] GB/T 39461—2020 国际物流信息系统数据接口
[7] GB/T 40217—2021 财经信息技术 养老保险基金审计数据接口
[8] GB/T 41479—2022 信息安全技术 网络数据处理安全要求
[9] GB 50174—2017 数据中心设计规范
[10] JR/T 0196—2020 多方安全计算金融应用技术规范
[11] JR/T 0223—2021 金融数据安全 数据生命周期安全规范
[12] YD/T 3801—2020 电信网和互联网数据安全风险评估实施方法
[13] YDT 4690—2024 隐私计算 多方安全计算产品安全要求和测试方法
[14] DB33/T 1329—2023 数据资产确认工作指南
[15] DB3311/T 127—2020 公共数据共享安全管理规范
[16] DB4403/T 271—2022 公共数据安全要求
[17] T/PCAC 0009—2021 多方安全计算金融应用评估规范
[18] IEEE 2842—2021 Recommended Practice for Secure Multi-Party Computation
[19] IEEE P3117 Standard for Interworking Framework for Privacy-Preserving Computation
[20] ISO/IEC 4922—3 Information security – Secure multiparty computation – Part 3: Mechanisms based on garbled circuit
[21] 《中华人民共和国数据安全法》 中华人民共和国第十三届全国人民代表大会常务委员会第二十九次会议通过 2021 年 6 月 10 日
[22] 《中华人民共和国个人信息保护法》 中华人民共和国第十三届全国人民代表大会常务委员会第三十次会议通过 2021 年 8 月 20 日
[23] 《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》 2022 年 12 月 2 日