ICS

T/GXDSL

团　　体　　标　　准

T/GXDSL　—2026

数据隐私脱敏加工技术规范多源异构数据融合场景

Technical Specification for Data Privacy Desensitization Processing in Multi-source Heterogeneous Data Fusion Scenarios

（工作组讨论稿）

（本草案完成时间：2026-01-29）

2026 - - 发布　　2026 - - 实施

广西电子商务企业联合会　发布

目　　次

前　言

本文件依据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。

本文件由广西产学研科学研究院提出。

本文件由广西电子商务企业联合会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

数据隐私脱敏加工技术规范多源异构数据融合场景

1 引言

为贯彻落实国家关于构建数据基础制度、更好发挥数据要素作用的战略部署，有效应对多源异构数据融合场景下日益复杂的隐私安全挑战，保障数据要素依法有序、安全可信流通，依据《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》《关于构建数据基础制度更好发挥数据要素作用的意见》等国家法律法规与政策文件，制定本规范。旨在确立一套科学、严谨、可操作的技术与管理体系，为多源异构数据融合全过程中的隐私脱敏活动提供统一标准。通过规范敏感数据识别、分级防护、脱敏处理与效果评估，力求在筑牢安全底线的前提下，充分释放数据融合价值，服务于数字经济高质量发展、国家治理能力现代化与国家安全能力提升。

2 范围

本规范规定了在多源异构数据融合场景下，开展数据隐私脱敏加工的总体原则、技术架构、技术方案、实施流程、管理要求与评估标准。适用于国家机关、企业、事业单位、科研机构、社会组织等各类数据处理者，在开展跨系统、跨领域、跨层级的多源异构数据采集、汇聚、整合、加工、分析、共享与开放等融合活动时，进行的隐私脱敏相关设计、开发、实施、运维、评估与审计工作。

3 规范性引用文件

下列文件对于本规范的应用必不可少。凡是注日期的引用文件，仅注日期的版本适用于本规范。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本规范。

《中华人民共和国网络安全法》（2017 年 6 月 1 日起施行）

《中华人民共和国数据安全法》（2021 年 9 月 1 日起施行）

《中华人民共和国个人信息保护法》（2021 年 11 月 1 日起施行）

GB/T 35273-2020 信息安全技术个人信息安全规范

GB/T 37988-2019 信息安全技术数据安全能力成熟度模型

GB/T 39335-2020 信息安全技术个人信息安全影响评估指南

GB/T 40685-2021 信息技术数据脱敏技术指南

GB/T 42460-2023 信息安全技术个人信息去标识化效果分级评估规范

GB/T 42590-2023 信息安全技术数据分类分级规则

GB/T 25069-2022 信息安全技术术语

GB/T 43697-2024 数据安全技术数据分类分级规则

ISO/IEC 29100:2011 信息技术安全技术隐私框架

ISO/IEC 27001:2022 信息安全管理体系要求

4 术语和定义

GB/T 25069-2022 界定的以及下列术语和定义适用于本规范。

4.1 多源异构数据

来源于不同管理域、业务系统或采集渠道，在数据结构（如关系型、非关系型）、数据格式（如文本、图像、流式）、数据模式（Schema）及语义上存在显著差异的数据集合。其核心特征表现为来源多样性、结构复杂性和语义差异性。

4.2 数据隐私脱敏

依据法律法规要求和业务场景需要，通过技术手段对数据中的敏感信息进行变形、替换、遮蔽或抽象处理，以降低或消除其直接或间接可识别特定自然人、组织或国家秘密、核心利益关联性的过程，同时力求保持数据在特定分析目标下的统计特征与业务效用。

4.3 静态脱敏

对处于静止存储状态（如数据库、数据文件、数据仓库）的数据集进行一次性、批量的脱敏处理。脱敏后的数据通常用于开发测试、数据分析、科学研究等非生产环境，原始数据与脱敏数据物理隔离。

4.4 动态脱敏

在数据查询、调用、展示等实时访问过程中，根据访问主体身份、角色、上下文环境及访问策略，对返回结果中的敏感字段进行即时脱敏处理。生产环境中的敏感数据访问控制应优先采用此方式。

4.5 脱敏规则

定义特定类型敏感数据应如何被识别与处理的指令集合。包括但不限于：敏感模式识别规则（如正则表达式）、脱敏算法（如哈希、加密、泛化）、脱敏强度参数（如泛化区间、遮蔽位数）及适用场景约束条件。

4.6 脱敏效果评估

基于量化指标与定性分析，系统化评价脱敏处理后数据集的隐私保护水平、残留风险、数据可用性损失及对业务流程影响的综合性活动。评估须遵循 GB/T 42460-2023 等相关标准。

4.7 数据融合

将来自多个独立来源的数据进行关联、比对、拼接、聚合与深度分析，以形成新的、更全面、更精准的数据视图或知识，从而支持决策、创新或服务的过程。本规范特指涉及敏感信息的多源异构数据融合。

5 总体要求

5.1 基本原则

5.1.1 国家利益至上与合法合规原则：所有脱敏活动必须将维护国家安全和公共利益置于首位，严格遵守国家数据安全与个人信息保护法律法规，履行数据分类分级保护义务，不得利用脱敏技术规避监管或从事非法活动。

5.1.2 安全与利用平衡原则：坚持统筹发展与安全，脱敏策略的设计应在确保敏感信息得到有效保护、风险可控的前提下，最大限度促进数据的高质量利用与价值释放。

5.1.3 目的明确与最小必要原则：脱敏的范围、方式与强度必须与数据融合的具体目的严格对应，避免过度脱敏损害数据价值，或脱敏不足导致隐私泄露。处理活动应限于实现处理目的的最小范围。

5.1.4 分级分类与精准防护原则：严格依据 GB/T 42590-2023 等国家标准对融合数据进行分类分级，针对不同级别数据及其应用场景，采取差异化、精细化的脱敏技术与管控措施。

5.1.5 全程可控与可审计原则：建立覆盖数据融合与脱敏全生命周期的监控、审计与追溯机制。所有操作须留存完整、防篡改的日志，确保任何数据流动与处理行为可追溯、可定责。

5.1.6 责任协同与联防联控原则：明确数据提供方、融合处理方、使用方等各方安全责任，建立跨组织、跨系统的协同治理与应急联动机制，共同防范系统性数据安全风险。

5.2 组织与管理要求

5.2.1 责任体系：数据处理者应明确数据安全负责人及隐私脱敏管理专职岗位，建立健全从决策层

到执行层的责任体系。关键岗位设置应遵循职责分离与最小权限原则。

5.2.2 制度体系：制定并持续完善涵盖数据资产梳理、脱敏策略管理、技术实施、应急响应、审计评估等环节的规章制度与操作规程，确保流程化、规范化运作。

5.2.3 能力建设：定期开展面向全员的数据安全意识教育及面向技术人员的前沿脱敏技术培训。培训记录与考核结果应归档，保存期限不少于 3 年。

5.2.4 协同机制：建立业务部门、数据管理部门、安全合规部门及技术支撑部门间的常态化协同工作机制，确保脱敏需求准确传递、技术方案合规可行、业务目标有效达成。

6 技术架构

6.1 总体架构设计理念

应采用“数据驱动、安全内生、分层管控、灵活扩展”的架构理念。系统应围绕数据流转路径，将隐私保护能力深度嵌入数据处理各环节，实现安全与业务的深度融合。

6.2 核心层级与功能

架构自下而上应包括：

6.2.1 数据源层：对接各类结构化、半结构化、非结构化数据源，支持数据库、API、文件、流数据等多种接入方式。

6.2.2 接入与感知层：实现多源数据的统一接入、格式解析、元数据抽取及初步的数据质量检查。内置数据血缘捕捉能力，记录数据来源与转换关系。

6.2.3 核心脱敏层：智能发现引擎：综合运用规则匹配、模式识别、自然语言处理（NLP）、计算机视觉（CV）等技术，实现多模态敏感信息的自动、精准识别，识别准确率应不低于 97%，误报率应低于 2%。策略执行引擎：支持静态批量脱敏与动态实时脱敏双模式。静态脱敏单节点处理吞吐量应不低于 15 万条/秒；动态脱敏平均延迟应低于 50 毫秒。引擎须支持横向扩展与高可用部署。规则知识库：内置符合国标及行业最佳实践的脱敏规则模板不少于 80 种，支持基于图形化界面的自定义规则编排与复杂逻辑配置。规则版本需严格管理。融合处理层：在脱敏后的安全数据空间内，进行数据的关联、挖掘、建模与分析。此层应具备对关联分析可能导致的隐私再识别风险进行预警的能力。安全服务层：提供统一的脱敏数据服务接口，集成基于属性/角色的访问控制（ABAC/RBAC）、数据水印、安全多方计算（MPC）或联邦学习接口等高级安全服务能力。统一管控层：提供策略管理、任务调度、全景监控、综合审计、风险可视化与度量评估的一体化管控平台。支持与组织现有的安全信息与事件管理（SIEM）

系统对接。

6.3 非功能性要求

6.3.1 性能：系统应满足大规模数据融合场景的性能需求，具体指标见各层级要求。

6.3.2 可靠性：系统可用性应不低于 99.9%，具备故障自动转移与数据一致性保障机制。

6.3.3 安全性：系统自身应采用国密算法进行数据传输与存储加密，实现细粒度的身份认证与权限管理，并通过定期的安全渗透测试。

7 脱敏技术方案

7.1 数据分类分级与脱敏策略映射

严格依据 GB/T 42590-2023 等标准进行分类分级。脱敏策略必须与数据级别、使用场景强关联：

7.1.1 核心数据/一级（高敏感）数据：包含可直接识别特定自然人身份或关系国家安全、重大公共利益的信息（如身份证号、生物识别信息、国家秘密载体信息、核心算法参数等）。必须采用不可逆的强脱敏技术（如符合国密标准的加密哈希、安全删除、强泛化至无法还原的区间），且原则上禁止在未获最高级别授权下直接参与融合分析，应以高度抽象的统计结果或标签形式参与。

7.1.2 重要数据/二级（中敏感）数据：包含间接识别个人身份或涉及企业重要经营秘密的信息（如精确位置、组合职业信息、特定健康指标、未公开的经营数据等）。优先采用不可逆脱敏，在严格评估风险且确有必要时，可对部分字段在强访问控制下使用可逆脱敏（如格式保留加密）。

7.1.3 一般数据/三级（低敏感）数据：主要为经处理后无法识别到特定主体的统计性、群体性信息（如地域统计、年龄分组、匿名化标签等）。可采用轻量级脱敏或直接使用，但仍需进行使用监控。

7.2 分数据类型脱敏技术

7.2.1 结构化数据：强匿名化：应用 k-匿名（k≥5）、l-多样性、t-贴近性等模型，确保个体在发布数据中无法被区分。差分隐私：在统计查询或数据发布中注入经过数学证明的随机噪声，提供可量化的隐私保障（ε通常建议≤1.0）。同态加密/安全多方计算：用于支持在加密态或分散数据上进行融合计算，实现“数据可用不可见”。格式保留加密（FPE）：在需要保持数据格式与业务逻辑（如数据库索引、外键关联）时使用，密钥管理须符合最高安全标准。

7.2.2 半结构化与非结构化数据：文本与日志：采用高精度 NER 模型（召回率≥95%）识别实体，结合上下文语义进行遮蔽、泛化或同义词替换。图像与视频：使用像素级的目标检测与分割技术，对人脸、车牌、证件、敏感背景等进行不可逆的模糊或像素化处理，处理后的敏感区域 PSNR 值应低于 20dB。

音频与语音：应用声纹抹除与语音内容识别技术，对涉及个人身份的语音段进行变声、静音或文本替换处理。

7.3 多源关联脱敏特殊要求

7.3.1 统一标识符管理：为同一实体在不同源中的标识（如用户 ID、设备 ID）生成全域唯一的、不可逆的伪标识符，映射关系应由受控的独立服务管理，严禁在各数据源中直接暴露。

7.3.2 关联风险控制：在融合前与融合后，均需执行关联攻击模拟，评估“数据马赛克”效应。对于通过多源弱关联可还原强标识的高风险组合，必须提升脱敏强度或切断关联链路。

7.3.3 联邦学习/分析：鼓励采用联邦学习等技术范式，使原始数据不出域，仅交换加密的中间参数或梯度，从源头规避融合中的隐私泄露。

8 实施流程

8.1 全生命周期管理流程

必须遵循规划、建设、运行、优化的闭环流程：数据资产测绘→分类分级定级→隐私影响评估→脱敏策略制定→规则部署测试→脱敏作业执行→效果合规验证→持续监控优化→归档与销毁。

8.2 关键阶段控制点

8.2.1 隐私影响评估阶段：对融合场景进行强制性、标准化的评估，明确风险点与保护等级，形成评估报告作为方案设计依据。

8.2.2 策略制定与测试阶段：策略需经法务、安全、业务三方评审。脱敏规则须在独立的仿真测试环境中进行充分验证，测试覆盖率应达 100%，验证数据效用损失在可接受范围。

8.2.3 效果合规验证阶段：须引入第三方专业测评机构或内部独立审计团队，依据 GB/T 42460-2023 等标准，对脱敏后数据集进行攻击测试与效用评估，出具正式报告。未通过验证的数据不得流出。

8.2.4 持续监控阶段：对脱敏数据的使用情况进行持续审计，监测异常访问模式。每半年至少进行一次脱敏策略的有效性复盘与重评估。

9 管理要求

9.1 策略与合规管理

建立企业级统一的脱敏策略库，所有策略的制定、变更、废止必须经过严格的审批流程，并记录于

审计日志。每年至少开展一次全面的合规性审查，确保所有脱敏实践与最新的法律法规及国家标准保持一致。涉及向境外提供数据或跨境融合的场景，脱敏方案必须额外通过国家网信部门组织的安全评估。

9.2 人员与权限管理

严格执行权限分离，系统管理员、策略配置员、审计员的角色必须分离。所有特权操作需执行“双人授权、操作复核”机制。对能够接触原始敏感数据或脱敏映射关系的人员，实施背景审查并签订严格的保密协议。

9.3 应急与事件管理

制定专项应急预案，明确发生脱敏失效、敏感数据疑似泄露等事件的应急响应流程、上报机制（包括向国家网信、公安等监管部门的报告时限与内容）及公关应对措施。恢复时间目标（RTO）应不超过2小时，恢复点目标（RPO）应趋近于零。每季度至少开展一次应急演练，演练后须出具整改报告。

9.4 审计与问责

建立独立的数据安全审计职能，定期（至少每季度）对脱敏全流程进行审计，审计记录至少保存5年。对审计中发现的违规行为，实行“零容忍”，依法依规追究相关单位和个人责任，并纳入考核体系

10 附则

本标准由广西电子商务企业联合会负责解释。本标准自发布之日起试行，试行期为一年。试行期满后，根据实施反馈情况进行修订和完善。各相关单位可依据本标准制定具体的实施细则。若本标准与国家新颁布的法律法规或强制性标准有不一致之处，应以国家法律法规和强制性标准为准。本标准所引用的规范性引用文件如有更新，其最新版本适用于本标准。广西电子商务企业联合会将根据技术发展和应用需求，适时组织对本标准的复审与修订工作，以保障其持续的先进性和适用性。本标准的有效实施，有赖于各级医疗机构、主管部门、技术服务商和各相关方的共同努力，通过规范智慧医院数据互联互通共享技术，推动医疗健康数据资源有效整合与安全共享，提升医疗服务质量和效率，促进智慧医院建设规范化发展，为推进健康中国建设提供技术支撑。