ICS 35.020
CCS L 67

DB51

四 川 省 地 方 标 准

DB51/T 3058—2023

政务数据 数据脱敏规范

Government affairs data
specification for data desensitization

2023－04－28 发布　　2023－06－01 实施

四川省市场监督管理局　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》给出的规则起草。

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。

本文件由四川省大数据中心提出、归口并解释。

本文件起草单位：四川省大数据中心、北京启明星辰信息安全技术有限公司、信息产业电子第十一设计研究院科技工程股份有限公司、中国电子系统技术有限公司、浪潮云信息技术股份公司。

本文件主要起草人：余东亮、赵启斌、黄健、吴晓蓉、刘冰、卢彬、刘雯、杨颋、任轲正、齐翌、孙光春、张铭宇、杨燕、吴凤、尹嘉奇、周奕含、刘盛懋、蒋晓、刘宏、曾刚、朱孟凯、王燕、伏晓龙。

本文件为首次发布。

政务数据 数据脱敏规范

1 范围

本文件规定了四川省范围内政务数据共享开放过程中的数据脱敏原则、数据脱敏应用场景、数据脱敏流程。

本文件适用于指导四川省范围内政务数据的数据脱敏工作，以及各级政务部门对政务数据脱敏工作机制的建立和实施。

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 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

3 术语和定义

下列术语和定义适用于本文件。

3.1

敏感数据 sensitive data

由权威机构确定的受保护的信息数据。

注：敏感信息数据的泄露、修改、破坏或丢失会对人或事产生可预知的损害。

[来源：GB/T 39477-2020]

3.2

数据脱敏 data desensitization

通过一系列数据处理方法对原始数据进行处理以屏蔽敏感数据的一种数据保护方法。

[来源：GB/T 37988-2019]

4 脱敏原则

4.1 有效性

经过数据脱敏处理后，原始信息中包含的敏感信息已被移除，无法通过处理后的数据得到敏感信息；或者需通过巨大经济代价、时间代价才能得到敏感信息，其成本已远远超过数据本身的价值。此外，在处理敏感信息时，应注意根据原始数据的特点和应用场景，选择合适的脱敏方法。

4.2 真实性

脱敏后的数据应尽可能真实体现原始数据的特征，且应尽可能多的保留原始数据中的有意义信息，以减小对使用该数据的系统的影响。为达到真实性要求，在开展数据脱敏工作时，宜注意以下事项:

a) 保持原数据的格式；
b) 保持原数据的类型；
c) 保持原数据之间的依存关系；
d) 保持语义完整性；
e) 保持引用完整性；
f) 保持数据的统计、聚合等特征；
g) 保持频率分布；
h) 保持唯一性。

4.3 高效性

数据脱敏过程中，可通过程序自动化实现，并可重复执行。在不影响有效性的前提下，需注意平衡脱敏的力度与所花费的代价，将数据脱敏工作控制在一定时间和经济成本内。

4.4 稳定性

数据脱敏时需保证对相同的原始数据，在各个输入条件一致的前提下，无论脱敏多少次，其最终结果是相同的。

4.5 可配置

数据脱敏需要确保脱敏工作的可配置性，按照输入条件不同，能够生成不同的脱敏结果，从而可按数据不同的使用场景等因素，为不同的最终用户提供不同的脱敏数据。

4.6 可审计

在数据脱敏的各个阶段应加入安全审核机制，严格、详细的记录数据脱敏过程中的相关信息，形成完整的数据脱敏记录，用于后续问题排查和追踪分析。

5 脱敏场景

5.1 共享场景脱敏

共享场景脱敏，是指在公共服务管理机构之间数据共享时，将敏感数据脱敏后再进行共享使用。在数据共享场景下，需根据数据的敏感等级来选择对应的条件进行共享。

5.2 开放场景脱敏

开放场景脱敏，是指政务部门以非排他形式有条件地面向公民、法人和其他组织开放部分政务数据时，将敏感数据脱敏后再进行开放使用。在数据开放场景下，需根据数据的敏感等级来选择对应的条件进行开放。

5.3 分析场景脱敏

分析场景脱敏，是指数据分析工程师进行数据挖掘分析时，将敏感数据脱敏后再进行数据分析使用。在数据分析场景下，需要重视数据之间的关联性、分析结果，宜采用泛化、均化等技术，脱敏后的数据

应保留原有的数据关系与格式，确保数据脱敏后不会影响分析结果，例如多个表格内相同人员的姓名，需要确保脱敏后结果一致。

5.4 开发测试场景脱敏

开发测试场景脱敏，是指系统测试、联调时，将敏感数据脱敏后再进行开发测试使用。在开发测试场景下，需要重视数据的可用性，因此主要采用替换、变形等技术，敏感数据脱敏可以采用相同含义的数据替换原有的敏感数据，例如身份证信息脱敏后仍然为有效的身份证信息。

6 脱敏流程

6.1 确认脱敏范围

脱敏工作实施主体可基于数据安全管理和保障需要，遵照相关法律法规、政策文件、标准规范，结合实际工作管理需要确定数据脱敏范围，明确哪些数据分类属于敏感数据。有些信息本身可能并不直接是敏感数据，但是可通过与其他一些信息结合后推断出敏感信息，此时也应将此类信息纳入数据脱敏的范围。

6.2 标识敏感数据

通过业务梳理发现敏感数据之后，需要对敏感数据进行标识，包括敏感数据位置、敏感数据的格式等信息，如字段名称、类型、长度等属性，以便后续对敏感数据的访问、传输和处理进行跟踪和监督。

6.3 确定脱敏方法

在对敏感数据脱敏前，需要根据脱敏场景和业务需求选择数据脱敏方法。具体数据脱敏方法参见附录A。

6.4 实施数据脱敏

数据脱敏步骤应包括：使用申请、脱敏审批、选择脱敏方法、脱敏操作、脱敏内容审核、脱敏过程审计。

a) 使用申请：数据申请者需提交《数据申请表》（参见附录 B），明确数据使用目的、范围、字段内容以及使用时间等信息；
b) 脱敏审批：建立敏感数据脱敏审批管理工作机制，并进行审核，再决定是否批准使用申请；
c) 选择脱敏方法：评估敏感数据，并基于主要使用场景选择对应的脱敏方法；
d) 脱敏操作：通过自动化脱敏工具，下发技术性的脱敏策略、并对脱敏过程中产生的操作行为进行记录；
e) 脱敏内容审核：评估脱敏后的数据内容是否仍包含敏感数据，确认不含敏感数据后，数据申请者方可获取数据；
f) 脱敏过程审计：确保所有的操作过程被审计，并做好记录留存，定期开展过程审计。

附 录 A
（资料性）
数据脱敏方法一览表

数据脱敏方法一览表见表A.1。

表A.1 数据脱敏方法一览表

序号	脱敏方法	方法描述	适用数据类型	示例	使用场景
1	遮蔽	对数据项的部分或全部数据用符号替代	日期、时间、数字	将身份证号码“510125196709210019”进行遮蔽得到“510**********0019”	共享场景脱敏、开放场景脱敏
2	扰乱	利用加密、重排等方式对原始数据进行修改，保留数据原始特征，并能经过业务校验	通用	如将“123”变为“abc”	开发测试场景脱敏
3	泛化	保留原始数据局部特征的前提下使用其他方式替代原始数据的方式	通用	年龄泛化，从“20岁”泛化成“0-30岁”	分析场景脱敏、开发测试场景脱敏
4	均化	针对数值型敏感数据，在保证脱敏后数据集总值或平均值与原数据集相同的情况下，改变原始数值	数据集	将“23，44，65”均化为“18，86，28”	分析场景脱敏、开发测试场景
5	时间随机	用随机生成时间去替换原始时间数据的一种脱敏算法	时间	将“2021年6月18日”随机替换为“2020年3月12日”	开发测试场景脱敏
6	数值随机	用随机生成某个范围的一个值去替换原始值的一种脱敏算法	数字	将“32245”随机替换为“42314”	开发测试场景脱敏、共享场景脱敏、开放场景脱敏
7	文字随机	用随机生成某个文字去替换原始值的一种脱敏算法	文字	将姓名：“伏慕”随机替换为“雷琴”	分析场景脱敏、共享场景脱敏、开放场景脱敏

附 录 B
（资料性）
数据申请表

数据申请表见表B.1。

表B.1 数据申请表

数据申请表	
申请人	XXX
申请部门	XX部
申请目的	XXX
申请时间	XXXX.XX.XX
审批人	XXX
审批部门	XX部
数据范围	数据库：X；表：XX；字段：字段X，脱敏方法：遮蔽
	数据库：X；表：XX；字段：字段X，脱敏方法：扰乱
	数据库：X；表：XX；字段：字段X，脱敏方法：泛化
	数据库：X；表：XX；字段：字段X，脱敏方法：均化
	数据库：X；表：XX；字段：字段X，脱敏方法：时间随机
	数据库：X；表：XX；字段：字段X，脱敏方法：数值随机
	数据库：X；表：XX；字段：字段X，脱敏方法：文字重写
使用场所	数据库：A；IP：XX.XX.XX.XX 数据库：B；IP：XX.XX.XX.XX
使用时间	XXXX.XX.XX-XXXX.XX.XX