

# DB4403

深 圳 市 地 方 标 准

DB4403/T XXX—XXXX

## 公共数据质量管理规范

Specification for public data quality management

(送审稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施



目 次

前言 ..... II

1 范围 ..... 3

2 规范性引用文件 ..... 3

3 术语和定义 ..... 3

4 缩略语 ..... 4

5 基本要求 ..... 4

6 相关方职责 ..... 4

7 质量检查 ..... 5

8 质量评价 ..... 5

附录 A （资料性） 市公共数据质检规则 ..... 8

附录 B （资料性） 省“一网共享”的质检规则 ..... 10

参考文献 ..... 11

## 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。

本文件由深圳市政务服务和数据管理局提出并归口。

本文件起草单位：深圳市大数据资源管理中心、深圳市政务服务和数据管理局、深圳市信息安全管理中心、中电科新型智慧城市研究院有限公司、深圳市标准技术研究院、国家信息中心、深圳市润鹏华通创新科技有限公司。

本文件主要起草人：张军、孙飞、赵娜、傅裕、赵延、唐增来、周文治、单志广、张旭杰、罗宇民、邹滔、王鹏、莫家勤。

# 公共数据质量管理规范

## 1 范围

本文件规定了公共数据质量管理的基本要求和相关方职责，以及开展公共数据质量检查和质量评价的管理要求。

本文件适用于公共数据质量管理、开发、开放、共享和评价。

## 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。

DB4403/T 271—2022 公共数据安全要求

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### 数据 data

信息的可再解释的形式化标识，以适用于通信、解释或处理。

注：可以通过人工或自动手段处理数据。

[来源：GB/T 36344—2018，2.1]

### 3.2

#### 公共数据 public data

公共管理和服务机构在依法履行公共管理职责或者提供公共服务过程中产生、处理的数据。

注：本文件提及的数据均为公共数据。

### 3.3

#### 数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源：GB/T 36344—2018，2.3]

### 3.4

#### 数源单位 data source department

按照相关法律法规确定的某一类公共数据的法定采集部门。

### 3.5

#### 元数据 metadata

关于数据或数据元素的数据（可能包括其数据描述），以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

[来源：GB/T 36344—2018，2.2]

## 4 缩略语

下列缩略语适用于本文件。

PDCA: 计划-执行-检查-处理 (Plan-Do-Check-Act)

IP: 互联网协议 (Internet Protocol)

JSON: JavaScript对象表示法 (JavaScript Object Notation)

SQL: 结构化查询语言 (Structured Query Language)

XPATH: 可扩展标记语言路径语言 (XML Path Language)

## 5 基本要求

- 5.1 公共数据应遵循“一数一源一标准”，避免同一数据有两个及以上来源，且数据执行标准不一致。
- 5.2 各相关方应遵循“PDCA”管理理念，不断提升公共数据质量水平。
- 5.3 应遵循“谁所有谁负责、谁持有谁负责、谁管理谁负责、谁使用谁负责、谁采集谁负责”的原则落实相关方数据安全主体责任。

## 6 相关方职责

### 6.1 数源单位

- 6.1.1 应保障提供公共数据的效率、透明度和公平性，明确来源于本单位的公共数据清单，并在本单位的职权范围内采集、核准、提供和管理公共数据。
- 6.1.2 应按照公共数据资源目录编制要求对本单位的公共数据进行编目和管理。
- 6.1.3 应按照公共数据质量管理要求，建立和完善本单位的数据质量管理体系，参照 GB/T 36073—2018 中第 11 章的方式加强数据质量管理。
- 6.1.4 应按照 DB4403/T 271—2022 的要求建立公共数据安全保障能力，保障本单位的公共数据安全。
- 6.1.5 应建立公共数据更新机制，确保纳入公共数据资源目录中的公共数据与本单位的业务数据的一致性。
- 6.1.6 应对公共数据使用单位或社会公众反馈不符合业务需求的公共数据进行研判和核查，改进数据质量。
- 6.1.7 应配合公共数据主管单位开展数据质量检查、质量评价和改进工作。
- 6.1.8 公共数据在共享与公开过程中，应加强公共数据中涉及商业秘密和个人信息等敏感信息的保护。
- 6.1.9 若本单位采集的数据由上级主管部门存储和管理的，应会同上级主管部门做好本单位数据的质量和安全管理等工作。

### 6.2 公共数据使用单位

- 6.2.1 可向数源单位提出公共数据质量要求。
- 6.2.2 在公共数据使用过程中，应及时向数源单位和公共数据主管单位反馈发现的公共数据质量问题。
- 6.2.3 在授权范围内使用并保管好获取的公共数据，采取必要防护措施，保障公共数据安全，不得损害国家利益、公共利益和第三方合法权益。

### 6.3 公共数据主管单位

- 6.3.1 应建立公共数据管理平台，为数源单位和公共数据使用单位提供数据质量检查、共享与开放支撑服务。

- 6.3.2 应建立公共数据资源目录体系，制定公共数据资源目录编制规范。
- 6.3.3 应建立公共数据安全保障制度，制定公共数据安全等级保护措施，保障公共数据管理平台内公共数据的安全。
- 6.3.4 应组织对数源单位的公共数据质量评价工作。
- 6.3.5 宜建立公共数据质量管理协调机制，积极协调公共数据使用单位、不同数源单位之间的公共数据质量管理协作关系，包括明确质量需求和质量问题改进等。

7 质量检查

7.1 检查范围

公共数据质量检查范围为纳入公共数据资源目录的公共数据。

7.2 检查要求

- 7.2.1 公共数据主管单位应会同数源单位明确公共数据质量检查规则，宜为数源单位提供通用的公共数据质量检查工具，协助数源单位将纳入公共数据资源目录的公共数据进行编目并开展质量检查。
- 7.2.2 数源单位可在公共数据管理平台进行数据资源登记，并依托公共数据管理平台对本单位需要纳入公共数据资源目录的公共数据进行质量检查，检查要求如下：
  - a) 应在公共数据共享至市公共数据管理平台前，按照公共数据质量检查要求开展数据质量检查工作；
  - b) 应对检查中发现的问题数据进行及时整改，整改完成后的问题数据应重新进行质量检查；
  - c) 检查完成后，形成公共数据质量检查结果报告；
  - d) 公共数据管理平台质量检查结论为合格的，方可通过市公共数据管理平台进行共享。

注：市公共数据质检规则参见附录A。
- 7.2.3 公共数据主管单位应按照省“一网共享”公共数据编目和质检要求，对需要纳入省公共数据管理平台的公共数据开展质检。质检不合格的，应及时通知数源单位对问题数据进行改进。

注：省“一网共享”质检规则参见附录B。

8 质量评价

8.1 评价范围

公共数据质量评价范围为数源单位非涉密的公共数据。

8.2 评价内容

- 8.2.1 公共数据从数据管理制度、元数据质量、业务数据质量和数据安全维度进行质量评价，见图 1。

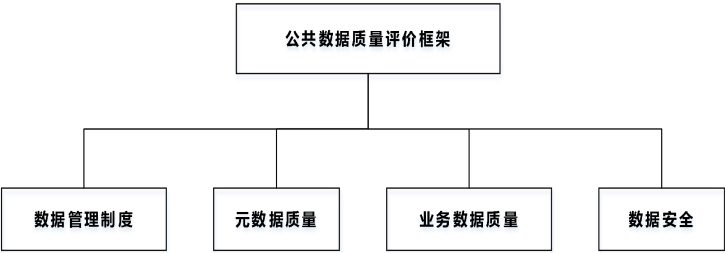


图 1 公共数据质量评价框架

8.2.2 数据管理制度评价内容包括数据共享和开放制度和数据安全管理制度等，具体要求如下：

- a) 数据共享和开放制度：至少包括数据共享原则、元数据标准、数据标准、共享调用标准、数据共享流程等；
- b) 数据安全管理制度：至少包括数据安全管理制度及人员要求、数据分级、数据保密性规范、数据完整性规范、数据备份与恢复等；
- c) 其他：数据管理组织架构、数据质量管理、元数据管理、主数据管理或业务数据管理等。

8.2.3 元数据质量评价内容包括业务元数据、技术与操作元数据、流程元数据和管理元数据，其中应根据实际情况确定元数据质量评价范围，业务元数据质量评价是必选项，技术与操作元数据、流程元数据、管理元数据是可选项。具体要求如下：

- a) 业务元数据：至少包括主题和概念领域、实体及属性的业务名称和业务定义、属性的数据类型和其他特性；
- b) 技术与操作元数据：至少包括数据的技术细节、存储数据的系统以及在系统内和系统之间数据流转过程的信息，如数据存储特性、数据迁移信息等；
- c) 流程元数据：至少包括数据流程、业务规则、流程工具等；
- d) 管理元数据：至少包括数据管理岗位、监管制度流程和责任分配等。

8.2.4 业务数据质量评价内容包括数据的规范性、完整性、准确性、一致性、时效性和可访问性，具体要求如下：

- a) 数据规范性：评价业务数据符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度。包括数据标准度量、模型度量、数据安全规范等。
- b) 数据完整性：按照数据规则要求，评价数据元素被赋予数值的程度，包括数据元素完整性、数据记录完整性等。
- c) 数据准确性：评价业务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，包括数据内容正确性、数据格式合规性、数据重复率、数据唯一性等。
- d) 数据一致性：评价业务数据与其他特定上下文中使用的数据是否存在矛盾，包括相同数据一致性、关联数据一致性等。
- e) 数据时效性：评价业务数据在时间变化中的正确程度，包括基于时间段的正确性、基于时间点的及时性、时序性等。
- f) 数据可访问性：评价业务数据能被访问的程度，包括数据可访问、数据可用性等。

8.2.5 数据安全评价内容包括数据的安全要求、安全能力和安全等级，具体要求如下：

- a) 安全要求：包括通用管理安全、通用技术安全及数据处理活动安全三方面安全要求；
- b) 安全能力：包括组织、制度、人员、技术四方面安全能力；
- c) 安全等级：对于不同的安全等级选取相对应的安全要求进行评价。

### 8.3 评价工作要求

8.3.1 公共数据主管单位负责组织公共数据质量评价工作，可自行开展或委托社会机构开展评价工作。当选择委托社会机构组织评价时，公共数据主管单位对相关评价工作进行指导监督。

8.3.2 公共数据质量评价应组建评价小组，评价小组应由业务、计算机技术和数据等领域人员组成，成员至少3名的单数，并确定1名评审组长，评审组长应由具备相关业务领域评价工作经验的人员担任。

8.3.3 在正式开展评价前，数源单位应按照评价工作要求对本单位的公共数据进行自查，并将自查情况反馈给评价小组。

8.3.4 评价小组应制定评价方案，并提前把评价方案和计划通知数源单位确认，数源单位应配合评价小组的评价工作。

8.3.5 评价小组依据评价方案，对数源单位进行现场评价，现场评价应包括首次会议、过程评价、末次



会议，过程评价可采取实地查看、核实数据、座谈等方式。

8.3.6 评价小组依据评价过程记录及评价数据，根据指标权重计算评价得分，并按照优、良、中和差四个级别给出质量分级，形成评价报告。质量等级划分见表 1。

表 1 质量等级划分

质量等级	分数范围
优	大于90分（含）
良	80（含）~90分
中	70（含）~80分
差	小于70分

8.3.7 对于评价中发现的问题，数源单位应及时进行改进，改进数据质量。公共数据主管单位应跟踪改进情况。

8.3.8 评价结果可作为考核数源单位的依据。

附 录 A  
(资料性)  
市公共数据质检规则

A.1 市公共数据质检规则

市公共数据质检规则见表A.1。

表 A.1 市公共数据质检规则说明

规则范围	维度	规则定义模板名称	说明
库级	准确性	表行数检测	统计并输出指定数据库中所有数据库表的总行数，结果以数据库为维度呈现。
	完整性	空值检测	计算指定数据库中的所有数据库表的空值数量，结果以数据库为维度呈现。
		空字符串检测	统计指定数据库中所有数据库表的空字符串数量及占比，结果以数据库为维度呈现。
表级	准确性	表行数检测	统计并输出指定数据库表的行数，结果以数据库表为维度呈现。
		行重复检测	统计并输出数据库表中的重复行数量及占比。
	及时性	更新频率检测	检测数据更新时间是否符合预定的更新频率要求，包括每日、每周、每月、每季度、每半年、每年、每两年。
字段级	有效性	正则表达式检测	使用正则表达式匹配字段格式，统计符合与不符合正则表达式规则的记录数量及占比，结果以字段为维度呈现。
		长度一致性检查	检查字段内容是否符合预定义的长度约束，如“统一社会信用代码”长度必须为 18 位，“公民身份证号码”长度必须为 18 位或 15 位。
		范围校验	判断字段的值是否在预定义的范围，例如“年龄”应不小于零。
		值域范围检测	检测字段的代码值是否在预定义的值域范围内，例如性别代码、民族代码等。
		编码规范检测	判断统一社会信用代码、身份证号码等是否符合编码规范。例如，“统一社会信用代码”的内容必须满足以下规范：由 18 位数字或大写英文字母（不包括 I、O、Z、S、V）组成，第 18 位为校验码，需符合校验规则。
		标志字段的准确性检测	根据规定的信息要求，检测字段值是否符合正确的标志取值。
		范围校验	判断字段的值是否在预定义的范围，例如“年龄”应不小于零。
		日期格式检测	验证并统计指定字段中符合与不符合日期格式的记录数量及占比，结果以字段为维度呈现。
		数字格式检测	验证并统计指定字段中符合与不符合数字格式的记录数量及占比，结果以字段为维度呈现。
		手机号检测	验证并统计指定字段中符合与不符合手机格式的记录数量及占比，结果以字段为维度呈现。
		身份证检测	验证并统计指定字段中符合与不符合身份证格式的记录数量及占比，结果以字段为维度呈现。

表 A.1 市公共数据质检规则说明（续）

规则范围	维度	规则定义模板名称	说明
字段级	有效性	IP 地址校验	验证并统计指定字段中符合与不符合 IP 地址格式的记录数量及占比，结果以字段为维度呈现。
		电话格式校验	验证并统计指定字段中符合与不符合电话格式的记录数量及占比，结果以字段为维度呈现。
		SQL	执行自定义 SQL 语句进行数据检测，输出结果根据设定的字段级维度呈现。
		时空数据/流数据检测	通过将数据转化成 JSON 对象后，基于 XPATH 解析出字段内容，计算指定字段在不同的后置字段级规则产生的结果，结果以字段为维度呈现。
	完整性	必填字段检测	检测必填字段是否存在空值，例如统一社会信用代码、公民身份证号码等。
		外键检测	验证外键约束的完整性，输出不满足主键和外键关系的记录数量，结果以字段为维度呈现。
		空值检测	计算指定字段的空值数量，结果以字段为维度呈现。
		空字符串检测	统计指定字段的空字符串数量及占比，结果以字段为维度呈现。
	唯一性	主键唯一性检查	检测业务主键字段的唯一性，以判断数据是否存在冗余记录。
		唯一值检测	统计指定字段中的唯一值和非唯一值的数量及占比，结果以字段为维度呈现。
	准确性	字段类型验证	检查每个字段是否符合其定义要求，比如日期字段应为有效日期格式，数字字段应为数字
		平均值检测	计算并输出指定字段的数据平均值。
		总和检测	计算并输出指定字段的数据总和。
		最大值检测	计算并输出指定字段的最大值。
		最小值检测	计算并输出指定字段的最小值。
		枚举值检测	统计指定字段中各个枚举值的数量及占比，结果以字段为维度呈现。
		同比检测	根据设定时间范围，计算指定字段在不同时间段的前置规则结果的同比值，结果以字段为维度呈现。
		环比检测	根据设定时间范围，计算指定字段在不同时间段的前置规则结果的环比值，结果以字段为维度呈现。
	关联性	跨库逻辑一致性	通过计算和对比，确保数据值在相同数据库实例中的不同库下表中代表相同的信息。
		跨表逻辑一致性	通过计算和对比，确保数据值在不同表中代表相同的信息。例如，同一客户的不同交易记录中的身份信息应一致。
		跨字段逻辑一致性	判断表内字段的业务逻辑关系正确。例如一个订单的“发货日期”应早于“收货日期”。

附 录 B  
(资料性)  
省“一网共享”的质检规则

B.1 省“一网共享”的质检规则

省“一网共享”的质检规则说明见表B.1。

表 B.1 省“一网共享”的质检规则说明

评价维度	描述
及时性	判断数据的更新时间是否满足更新频率的要求。频率：每日、每周、每月、每季度、每半年、每年、每两年、实时。
完整性	关键业务字段非空检测。如操作类型标识、统一社会信用代码、公民身份证号码等。
	技术字段检测。检查库表挂接中技术字段是否符合挂接规范。
唯一性	业务主键字段的唯一性判断，检测数据是否存在冗余。
有效性	代码值检测。判断字段的代码值是否在值域范围内。如性别代码、民族代码等。
	长度校验检测。判断检测对象的长度是否满足长度约束。如“统一社会信用代码”长度必须为18位、“公民身份证号码”长度必须为18位或15位。
	内容规范检测。判断统一社会信用代码、身份证号码等是否符合编码规范。如“统一社会信用代码”的内容必须满足如下规范：由18位数字或大写英文字母（不使用I、O、Z、S、V）组成，第18位校验码符合校验规则。
	取值范围检测。判断字段是否在预定义的范围内。如“年龄”应不小于零。
	标志取值检测。通过相应规定的信息要求，检测值是否正确。
准确性	取值准确性检测。通过不同地方出现的信息判断取值是否准确。
一致性	逻辑一致性检测。判断表内字段的逻辑合理性。如“死亡日期”必须大于等于“出生日期”。
	存在一致性检测。检查相应的所需信息是否存在于相应库表中。
	等值一致性检测。通过一定的计算，检测数据值在信息含义上是否存在冲突。

参 考 文 献

[1] GB/T 42381.61—2023 数据质量 第61部分：数据质量管理：过程参考模型

[2] GB/T 42381.63—2023 数据质量 第63部分：数据质量管理：过程测量

[3] GB/T 25000.12—2017 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第12部分：数据质量模型

[4] GB/T 25000.24—2017 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第24部分：数据质量测量

[5] GB/T 33850—2017 信息技术服务 质量评价指标体系

[6] GB/T 43709—2024 资产管理信息化 数据质量管理要求

[7] GB/T 36344—2018 信息技术 数据质量评价指标

[8] GB/T 36073—2018 数据管理能力成熟度评估模型

[9] GB/T 42129—2022 数据管理能力成熟度评估方法

[10] SY/T 7005—2014 数据质量控制与评估原则

[11] DB4403/T 439—2024 公共数据安全评估方法

[12] GF 007—2022 广东省数据资源“一网共享”体系 数据质量检测管理规范（试行稿）

[13] 广东省人民政府. 广东省公共数据管理办法：广东省人民政府令第290号. 2021年

[14] 广东省人民政府. 广东省政务数据资源共享管理办法(试行)：粤府办〔2018〕50号, 2018年

[15] 深圳市第七届人民代表大会常务委员会. 深圳经济特区数据条例：2021年公告第十号. 2021年

[16] 深圳市人民政府. 深圳市政务信息资源共享管理办法：深府〔2015〕99号. 2015年

---