乐山市政务数据质量评估规范

# 说明

## 目的

为规范乐山市政务数据资源质量，跟踪数据处理流程闭环，保障政务数据高质量开放利用，特制定本规范。

## 内容概述

本规范主要包含政务数据质量评估指标、评估方法、评估流程、评估结果应用。数据质量评估反映的数据质量问题，及时反馈到数据提供方、数据处理方和数据使用方，保证数据质量持续稳步提升。

## 术语和定义

### 1.数据

对事实、概念或指令的一种形式化表示，适用于以人工或自动方式进行通信、解释或处理。

### 2.政务部门

各级地方党委、人大、政府、政协、法院、检察院及其直属各部门（单位），以及法律法规授权具有行政职能的事业单位和社会组织。

### 3.政务数据

各级政府部门及其技术支撑单位在履行职责过程中采集、生产、存储、管理的各类数据资源。根据可传播范围，政务数据一般包含可共享政务数据、可开放公共数据和不宜开放共享政务数据。

### 4.数据质量

在指定条件下使用时，数据的特性满足明确和隐含的要求的程度。

## 适用对象

本规范适用于共享交换平台归集到的所有政务数据，指政务部门在履行职责过程中产生或获取的，以一定形式记录、保存的文件、资料、图表和数据等各类信息资源数据，包括政务部门直接或通过第三方依法采集的、依法授权管理的和因履行职责需要依托政务信息系统形成的信息资源数据等。

## 其他说明

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。

本文件由乐山市数字经济发展局提出。

1. 质量评估规范

## 评估指标

### 1.指标框架

数据质量评估指标框架见图评估指标包括以下内容:

(1) 内容质量：包括数据规范性、准确性、完整性和可用性。

(2) 过程质量：包括数据处理过程的处理效果、一致性。

(3) 效用质量：包括数据可访问性、时效性。



图1数据质量评估指标框架图

### 2.指标说明

|  |  |  |  |
| --- | --- | --- | --- |
| **一级指标** | **二级指标** | **三级指标** | **指标描述** |
| 内容 质量 | 规范性 | 命名规范性 | 数据库、数据集、数据元的命名方式符合相关命名规范的情况。 |
| 元数据规范性 | 元数据描述符合相关规范的情况。 |
| 参考数据规范性 | 参考数据符合既有格式及规范的情况。 |
| 数据权限规范性 | 是否基于相关法律法规制定数据安全权限的规范性文件。 |
| 敏感字段脱敏占比 | 已脱敏字段占全部字段的比例。 |
| 准确性 | 数据格式合规性 | 数据格式（数据类型、数据范围、数据长度、精度、编码等） 是否满足预期要求，如手机号、身份证号、性别、统一社会信用代码等。具体体现为满足格式要求的数据集字段数占总数据集的字段总数的比例。 |
| 数据重复率 | 特定字段、记录、文件或数据集中存在重复数据的比例。 |
| 数据唯一性 | 特定数据字段、记录、文件或数据集唯一标识的程度，具体体现为满足唯一性的数据集个数与总数据集个数的比例。 |
| 脏数据岀现率 | 脏数据岀现率低于阈值的数据集的比例。 |
| 完整性 | 数据元素空值率 | 数据元符合数据元管理要求，不存在值为空的现象。 |
| 数据记录空值率 | 数据记录内容中存在空值的情况。 |
| 数据记录缺失率 | （融合应用时）数据记录内容覆盖所有数据，不存在缺失现象。 |
| 可用性 | 数据字段可用性 | 数据字段值有效、可用的比例。 可用的比例。 |
| 依赖字段可用性1 | 存在依赖关系的字段同时非空的比例。 |
| 数据集可用性 | 数据集由多个字段组成，根据字段的可用率伊算整个数据集的 可用率。 |
| 数据接口可用性 | 数据接口调用能够准确、正常的返回请求的比例。 |
| 过程质量 | 处理效果 | 数据合格率 | 校验后符合数据定义和规则的数据与总体数据的比例。 |
| 清洗保留率 | 清洗后的数据量占清洗前数据总量的比例。 |
| 一致性 | 相同数据一致性 | 同一数据在不同位置存储或被不同应用或用户使用时，数据的 一致情况。  |
| 关联数据一致性 | 符合基础数据元组合的数据集的比例。 |
| 效用 质量 | 可访问性 | 数据字段可访问率 | 可访问的数据字段的数量与总数的比例。 。 |
| 数据集 | 可访问的数据集的数量与数据集总数的比例。  |
| 数据接口有效性 | 可访问接口正确性。 |
| 时效性 | 时段数据正确性 | 日期范围的记录或者频率分布符合业务需求的情况。  |
| 时点数据正确性 | 特定时点的记录数、频率分布或延迟时间符合业务需求的情况。 |
| 数据时序正确性 | 数据时序的正确性。 |

## 评估方法

### 1.检核方法

数据质量评估按照评估指标对数据质量进行检核，检核方法包括：

（1）系统检核：使用数据质量管理系统等数据质量工具实现自动检核，如自定义 SQL 语句检核；

（2）人工检核：根据评估指标，结合个人专业判断进行数据检核，如资料对比、经验判断等。

### 2.检核方式

（1）全量检核

对涉及的所有数据进行逐一检核数据质量。

（2）增量检核

对涉及的数据，在特定的范围和时间段内新增的数据进行逐一检核数据质量。

（3）抽样检核

按照抽样方案，对抽取的数据进行逐一检核数据质量。

### 3.分值设计

评估指标分值设计用绝对权数表示，绝对权数等同于各指标满分分值。各层级评估指标分值设计见附录A。

### 4.结果计算

评估对象的最终评估总得分按公式（1）进行计算。一级指标得分按公式（2）计算，二级指标得分按公式（3）进行计算。

$P=\sum\_{i=1}^{3}F\_{i}$ （1）

$F\_{i}=\sum\_{j=a}^{n}S\_{ij}$ （2）

$S\_{ij}=\sum\_{k=b}^{m}T\_{ijk}$ （3）

式中：

$P$——评估对象最终总得分；

$i$——第 $i$ 个一级指标；

$Fi $——第 $i$ 个一级指标实际得分；

$j $——第 $j$ 个二级指标；

$a $——分解于第 $i$ 个一级指标的起始二级指标序号；

$n $——分解于第 $i$ 个一级指标的结尾二级指标序号；

$S\_{ij}$ ——第 $j$ 个二级指标（分解于第 $i$ 个一级指标）实际得分；

$k $——第 $k$ 个二级指标；

$b $——分解于第 $j$ 个一级指标的起始二级指标序号；

$m $——分解于第 $j$ 个一级指标的结尾二级指标序号；

$T\_{ijk} $——第$ k $个三级指标（分解于第 $j$ 个二级指标）实际得分。

## 评估流程

### 1.流程图

数据质量评估流程见图2，包括评估准备、指标选择、规则确定、评估实施、结果核验、报告编制。



图2 数据质量评估流程图

### 2.评估准备

评估前应了解具体业务对特定政务数据的需求，确定评估目的、对象及范围、评估方式，编制数据质量评估方案。

### 3.指标选择

根据明确的评估目的，按照第3章规定选择合适的评估指标，并确定评估指标分值。

### 4.规则确定

根据选定的评估指标、评估对象确定数据质量检核规则。

### 5.评估实施

（1）自评估

按照数据质量评估方案确定的评估指标对评估对象进行自评估。

（2）质量检核

按照数据质量检核规则对数据进行数据质量检核。

### 6.结果核验

结合自评估情况，对检核结果进行核验，根据不同数据使用场景和业务需求对数据质量评估得分进行分级评价，确定数据质量好坏。

### 7.报告编制

根据评估结果编制评估分析报告，内容宜包括但不限于：评估对象及范围、评估指标、计分规则、评估检核方法、评估实施过程、质量问题。

## 评估结果应用

1.数据质量评估报告反映的数据质量问题，及时反馈数据提供方、数据处理方和数据使用方的相关部门及人员进行处理。

2.跟踪数据质量问题的处理，形成数据质量问题的闭环处置。

3.通过数据质量问题及其相关处理经验的汇总、分析，逐步积累形成包含数据质量检核规则、质量问题描述、针对性解决方案的数据质量知识库。

附录A

**（资料性）**

**评估指标分值**

|  |
| --- |
| **表**A.1**评估指标分值** |
| 一级指标 | 一级指标分值 | 二级指标 | 二级指标分值 | 三级指标 | 三级指标分值 |
| 1.内容质量 | 56 | 1.1规范性 | 20 | 1.1.1命名规范性 | 4 |
| 1.1.2元数据规范性 | 4 |
| 1.1.3参考数据规范性 | 4 |
| 1.1.4数据权限规范性 | 4 |
| 1.1.5敏感字段脱敏占比 | 4 |
| 1.2准确性 | 15 | 1.2.1数据格式合规性 | 4 |
| 1.2.2数据重复率 | 4 |
| 1.2.3数据唯一性 | 4 |
| 1.2.4脏数据岀现率 | 3 |
| 1.3完整性 | 9 | 1.3. 1数据元素空值率 | 3 |
| 1.3.2数据记录空置率 | 3 |
| 1. 3. 3数据记录缺失率 | 3 |
| 1.4可用性 | 12 | 1.4. 1数据字段可用性 | 3 |
| 1.4.2依赖字段可用性 | 3 |
| 1.4.3数据集可用性 | 3 |
| 1.4.4数据接口可用性 | 3 |
| 2过程质量 | 20 | 2.1处理效果 | 10 | 2.1.1数据合格率 | 5 |
| 2.1.2清洗保留率 | 5 |
| 2.2一致性 | 10 | 2.2.1相同数据一致性 | 5 |
| 2.2.2关联数据一致性 | 5 |
| 3效用质量 | 24 | 3.1可访问性 | 12 | 3.1.1数据字段可访问率 | 4 |
| 3.1.2数据集可访问率 | 4 |
| 3.1.3数据接口有效性 | 4 |
| 3.2时效性 | 12 | 3.2.1时段数据正确性 | 4 |
| 3.2.4时点数据正确性 | 4 |
| 3.2.3数据时序正确性 | 4 |