

ICS

X XX

备案号: XXXXX-XXXX

**DB11**

北京市地方标准

DB 11/ X XXX—XXXX

## 政务数据质量评估规范

Management specification for government data quality evaluation

(征求意见稿)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

北京市市场监督管理局 发布

## 目 次

目 次 .....	I
前 言 .....	II
政务数据质量评估规范 .....	1
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
3.1 .....	1
3.2 .....	1
3.3 .....	1
4 数据质量评估模型 .....	1
4.1 数据质量六性评估模型 .....	1
4.2 数据质量五级评级模型 .....	2
5 数据质量六性评估 .....	2
5.1 指标编码规则 .....	2
5.2 数据及时性 .....	3
5.3 数据可用性 .....	3
5.4 数据完整性 .....	4
5.5 数据规范性 .....	4
5.6 数据准确性 .....	5
5.7 数据一致性 .....	5
6 数据质量五级评级 .....	6
6.1 数据质量评分 .....	6
6.2 质量等级判定 .....	8
7 数据质量评估程序 .....	8
参 考 文 献 .....	10

## 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。

本文件由北京市政务服务和数据管理局提出并归口。

本文件由北京市政务服务和数据管理局组织实施。

本文件起草单位：北京市政务服务和数据管理局、北京市大数据中心。

本文件主要起草人：

# 政务数据质量评估规范

## 1 范围

本文件规定了政务数据的质量评估管理规范，规定了数据的及时性、可用性、完整性、规范性、准确性和一致性六个维度的评估指标，描述了一种“五级六性”数据质量评估模型和数据质量评估指标，给出了数据质量评级方法。

本文件适用于政务数据汇聚、治理、共享过程中对数据质量的评估管理。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1 政务数据 government data

各级政务部门及其支撑单位在履行职责过程中依法采集、生成、存储、管理的各类数据资源。

[来源：GB/T 38664.1—2020,3.1，有修改]

### 3.2 数据质量 data quality

基于对数据使用的预期，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，有修改]

### 3.3 数据质量等级 data quality level

依据数据质量评估指标对数据进行质量评估的结果。

## 4 数据质量评估模型

质量评估模型包括数据质量六性评估和数据质量五级等级，其中数据质量六性评估主要从数据及时性、可用性、完整性、规范性、准确性和一致性6个维度进行了定义，数据质量5级评级用于评价数据的质量情况，具体如下：

### 4.1 数据质量六性评估模型

数据质量六性评估模型见图1。

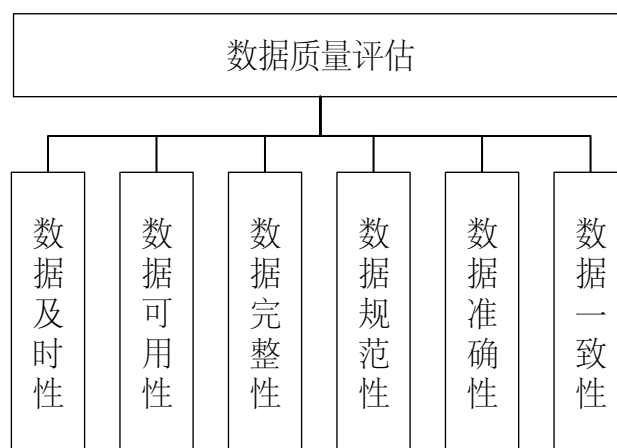


图1 数据质量六性评估模型

数据质量评估管理，包括数据及时性、可用性、完整性、规范性、准确性和一致性6个维度。

- a) 数据及时性：指数据更新周期是否符合业务逻辑及数据的获取是否及时，评估政务部门数据是否能够按照合理的更新周期及时更新数据。
- b) 数据可用性：指获取的数据是否能够有效的被使用。如结构化数据的数据描述文档是否存在，政务部门是否能够直接共享结构化数据等。
- c) 数据完整性：指数据是否完整，以及数据的描述性要素、要素属性及要素关系是否存在，主要包括实体缺失、属性缺失、记录缺失的内容。
- d) 数据规范性：主要用于评估数据内容与标准数据元的值域、格式符合度情况，衡量数据是否符合数据标准。
- e) 数据准确性：描述数据是否与其对应的客观实体的特征相一致。任何字段的数据都应该符合特定的值。准确性用于度量哪些数据和信息是不正确的，或者数据是没有可用含义的，准确性指标无法满足，那提供的数据会缺乏实际的业务使用价值。
- f) 数据一致性：主要用于描述数据类型、数据含义、以及数据内容是否一致，数据的值是否存在信息含义上的冲突。

#### 4.2 数据质量五级评级模型

数据质量评估等级，分为A、B、C、D、E五个级别，用于评价数据的质量情况，具体如下：

- a) A级：数据质量高。
- b) B级：数据质量较高。
- c) C级：数据质量存在轻微问题。
- d) D级：数据质量存在一般问题。
- e) E级：数据质量存在重大问题。

### 5 数据质量六性评估

#### 5.1 指标编码规则

指标编号是政务数据质量评估的唯一性编号，由一级指标、二级指标共4位数字组成，编码规则见图2。



图2 数据质量六性评估指标编码规则

一级指标由2位数字组成，01代表数据及时性指标、02代表数据可用性指标、03代表数据完整性指标、04代表数据规范性指标、05代表数据准确性指标、06代表数据一致性指标。二级指标由2位数字组成顺序码，范围为01-99。

## 5.2 数据及时性

数据及时性指标定义及评分方式见表1。

表1 数据质量六性评估数据及时性指标

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0101	更新周期合理性	描述数据资源的更新周期是否符合业务上数据实际产生周期情况。规范政务部门把数据汇聚周期设置在合理的范围。	示例：门诊就诊记录数据更新周期为1天。年报类数据的更新周期应为1年。	定性评分	根据数据的业务属性来判定数据更新周期是否更新 1、若数据更新周期合理，该数据资源得分为 $X=100 \times \text{权重}$ ； 2、若数据更新周期不合理，该数据资源得分 $X=0$ 。
0102	更新及时率	描述数据资源按照提交的更新周期进行更新。规范政务部门能够按时进行数据的更新。	示例：政务部门设置资源数据更新为1月，则需要每月更新数据	定性评分	根据数据是否及时按照更新周期更新，来判断数据是否及时更新。 1、若数据是按时更新或未更新时间，该数据资源得分 $X=100 \times \text{权重}$ ； 2、若数据未按时更新或无法判断是否更新，该数据资源得分 $X=0$ 。

## 5.3 数据可用性

数据可用性指标定义及评分方式见表2。

表2 数据质量六性评估数据可用性指标

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0201	可直接利用性	描述数据资源的可直接使用程度。如果源端是数据库类型，可直接通过数据库表的方式共享，增强数据的可用性。	示例：源端是数据库表直接通过数据库表对接实现。	定性评分	1、若原始数据为库表格式的结构化数据，则该数据资源得分 $X=100 \times \text{权重}$ ； 2、若原始数据为非结构化数据或者未入库的数据，则该数据资源得分 $X=0$ 。
0202	数据可理解性（数据描述文档）	指数据库表中的英文字段要有中文描述（如数据字典、码表等），描述要准确，	示例：数据库表字段CSRQ，有了中文描述才能理解是“出生日期”的含	定性评分	1、若数据资源表有字段的中文描述，则该数据资源得分 $X=100 \times \text{权重}$ ； 2、若数据资源表存在无注释的

		便于更好的理解数据。	义		字段导致无法理解该表的具体业务含义，则该资源得分 $X=0.6 \times 100 \times$ 权重。
--	--	------------	---	--	--

#### 5.4 数据完整性

数据完整性指标定义及评分方式见表3。

表3 数据质量六性评估数据完整性指标

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0301	字段完整性	描述实际汇聚数据资源的字段是否与资源目录注册的字段数量、描述保持一致。		定性评分	1、若保持一致，则该资源得分 $X=100 \times$ 权重； 2、若不一致，则该数据资源得分 $X=0$ 。
0302	空值校验	业务主键或重要字段等必填项的数据值不能为空值。	示例：涉人数据资源的姓名、身份证号等重要字段数据值不能为空。	定量评分	稽核表指定稽核列在限定范围内的空值数量为B，该稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times$ 权重。
0303	数据记录完整性校验	按照业务规则要求，验证当前稽核表中满足过滤条件的数据记录数范围是否满足预先设定规则，包括时间维度和空间维度。	示例：某数据资源实际业务产生时间为1966年，而实际汇聚数据只有2000年后的数据，那此数据资源不满足数据记录完整性要求	定性评分	1、若满足，则该资源得分 $X=100 \times$ 权重； 2、若不满足，则该数据资源得分 $X=0$ 。
0304	核心字段完整性	数据的核心关键字段内容应完整有效。	示例：涉人数据中证件号码、涉企数据中统一社会信用代码等唯一标识字段数据内容应完整。	定量评分	稽核表中指定稽核列在限定范围内核心字段内容缺失的数据数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times$ 权重。

#### 5.5 数据规范性

数据规范性指标定义及评分方式见表4。

表4 数据质量六性评估数据规范性指标

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0401	格式校验	数据属性值必须满足数据标准或数据字典中定义的展现格式约束	示例：日期字段必须定义为日期类的格式	定量评分	稽核表指定稽核列在限定范围内的不符合格式的数据数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times$ 权重。
0402	参照性校验-单项校验	稽核数据与参照数据进行比较校验，判断指定稽核列中	示例：在人口登记信息表中，稽核数据是人口登记性别数据，参照数据是性别码表，性别为：男性01、参照	定量评分	稽核表指定稽核列在限定范围内的不符合参照数据的数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100$

		不符合参照列内数据的情况	性校验，即人口登记数据的性别必须为01或02，不应出现其他代码。		×权重。
0403	参照性校验-双向校验	稽核数据与参照数据进行比较校验，判断指定稽核列中不符合参照列内数据的情况	示例：在人口登记信息表中，稽核数据是人口登记数据，参照数据是性别码表，性别为：男性01、女性02，要求人口登记数据的性别既有男又有女，若性别中只有01或只有02是错误的。	定性评分	1、结果正确，则该资源得分 $X=100 \times \text{权重}$ ； 2、结果错误，则该数据资源得分 $X=0$ 。
0404	业务规则校验	稽核数据符合业务规则的程度	示例：身份证号字段要符合身份证的编码规则。	定量评分	稽核表指定稽核列在限定范围内的不符合业务规则的数据质量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times \text{权重}$ 。

## 5.6 数据准确性

数据准确性指标定义及评分方式见表5。

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0501	重复校验	检查单列或者多列数据组合是否存在重复记录。	示例：每个人的身份证号码只能对应一个人。	定量评分	稽核表指定稽核列在限定范围内的重复数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times \text{权重}$ 。
0502	值域校验	描述数据是否与其对应的客观实体的特征相一致。	示例：年龄的取值范围应该在0-200岁之间。	定量评分	稽核表指定稽核列在限定范围内的不符合值域范围数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times \text{权重}$ 。
0502	数据内容可靠性校验	描述数据内容是否真实可靠，是否为无效的数据。	示例：企业注册资金为-100万，因注册资金不能为负值，此项数据内容无效。	定量评分	稽核表指定稽核列在限定范围内的数据内容为无效的数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times \text{权重}$ 。

表5 数据质量六性评估数据准确性指标

## 5.7 数据一致性

数据一致性指标定义及评分方式见表6。

表6 数据质量六性评估数据一致性指标

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0601	逻辑校验	检验数据的逻辑性或逻辑关系。	示例：入学时间应当早于毕业时间。	定性评分	1、结果正确，则该资源得分 $X=100 \times \text{权重}$ ； 2、结果不正确，则该数据资源得分 $X=0$ 。



0602	一致性校验	稽核表内数据，或表间数据是否相等。	<p>示例1：单表单行，身份证号码的7—10位要与出生日期一致；</p> <p>示例2：双表单行，同一个身份证号码在多个表中对应的姓名要一致。</p>	定量评分	稽核表指定稽核列中不符合一致性约束记录数量为B，稽核表限定范围内总条数为A，则该资源得分 $X=(1-B/A) \times 100 \times \text{权重}$ 。
------	-------	-------------------	---	------	--

## 6 数据质量五级评级

### 6.1 数据质量评分

政务数据质量评分，主要采用权重法计算质量评分。对数据质量评估的六性分配权重，通过加权求和的方式得到最终评分。

数据质量评估六性评分计算如下：

#### a) 数据及时性得分计算

$$X_1 = \frac{\sum_{i=1}^n (X_{1i} \times W_{1i})}{\sum_{i=1}^n W_{1i}}$$

式中：

$X_1$ ——数据及时性得分；

$X_{1i}$ ——数据及时性二级指标得分；

$W_{1i}$ ——数据及时性二级指标权重。

#### b) 数据可用性得分计算

$$X_2 = \frac{\sum_{i=1}^n (X_{2i} \times W_{2i})}{\sum_{i=1}^n W_{2i}}$$

式中：

$X_2$ ——数据可用性得分；

$X_{2i}$ ——数据可用性二级指标得分；

$W_{2i}$ ——数据可用性二级指标权重。

#### c) 数据完整性得分计算

$$X_3 = \frac{\sum_{i=1}^n (X_{3i} \times W_{3i})}{\sum_{i=1}^n W_{3i}}$$

式中：

$X_3$ ——数据完整性得分；

$X_{3i}$ ——数据完整性二级指标得分；

$W_{3i}$ ——数据完整性二级指标权重。

#### d) 数据规范性得分计算

$$X_4 = \frac{\sum_{i=1}^n (X_{4i} \times W_{4i})}{\sum_{i=1}^n W_{4i}}$$

式中：

$X_4$ ——数据规范性得分；

$X_{4i}$ ——数据规范性二级指标得分；

$W_{4i}$ ——数据规范性二级指标权重。

e) 数据准确性得分计算

$$X_5 = \frac{\sum_{i=1}^n (X_{5i} \times W_{5i})}{\sum_{i=1}^n W_{5i}}$$

式中：

$X_5$ ——数据准确性得分；

$X_{5i}$ ——数据准确性二级指标得分；

$W_{5i}$ ——数据准确性二级指标权重。

f) 数据一致性得分计算

$$X_6 = \frac{\sum_{i=1}^n (X_{6i} \times W_{6i})}{\sum_{i=1}^n W_{6i}}$$

式中：

$X_6$ ——数据一致性得分；

$X_{6i}$ ——数据一致性二级指标得分；

$W_{6i}$ ——数据一致性二级指标权重。

注 1：数据各维度的指标权重可根据数据实际使用场景自行设定。因对于不同的数据资源来说，并不是每个评估指标均适用，对不适用的指标权重应配置为 0。

对数据质量进行整体评分定级时，需根据数据实际应用情况对数据质量评估的六性评分设置不同的权重。数据是否准确、完整，能否被有效使用，是数据应用的基础，因此，数据的准确性评分和完整性评分需配置较高的权重。如若在某些特定场景需求下对数据时效性要求高，则应当将数据及时性评分配置较高的权重。

根据数据六性质量评分和配置的权重，最终得到数据质量评分：

$$S = \frac{X_1 \times W_1 + X_2 \times W_2 + X_3 \times W_3 + X_4 \times W_4 + X_5 \times W_5 + X_6 \times W_6}{W_1 + W_2 + W_3 + W_4 + W_5 + W_6}$$

式中：

$X_1$ ——数据及时性得分；

$X_2$ ——数据可用性得分；

$X_3$ ——数据完整性得分；

$X_4$ ——数据规范性得分；

- $X_5$ ——数据准确性得分；  
 $X_6$ ——数据一致性得分；  
 $W_1$ ——数据及时性权重；  
 $W_2$ ——数据可用性权重；  
 $W_3$ ——数据完整性权重；  
 $W_4$ ——数据规范性权重；  
 $W_5$ ——数据准确性权重；  
 $W_6$ ——数据一致性权重。

## 6.2 质量等级判定

数据质量评估分为A、B、C、D、E共5个等级，每个等级的描述及分值范围见表7。

**注2：**对数据质量进行等级评定时，分值范围可参考表7，不做强制要求，数据各维度的指标权重可根据各部门业务数据实际使用场景自行设定。

表7 数据质量5级评级等级描述及可应用情况

质量等级	数据质量等级描述及可应用情况	分值参考范围
A	数据质量高，可直接使用数据	$A < 85$
B	数据质量较高，经过基础处理后可使用	$80 < B \leq 84$
C	数据质量存在轻微问题，经过数据共性清洗后可使用	$75 < C \leq 80$
D	数据质量存在一般问题，进行针对性的数据清洗后可使用	$60 < D \leq 75$
E	数据质量存在重大问题，须从源头进行数据整改	$E \leq 60$

## 7 数据质量评估程序

数据质量评估程序详见图3，包括：确定质量评估对象、确定评估指标、配置质量评估规则、实施质量评估四大阶段。具体如下：

- 确定质量评估对象：根据具体业务对特定数据的需求，明确质量评估对象、范围等。
- 确定评估指标：根据明确的质量评估对象，按照第5章规定选择对应的评估指标。
- 配置质量评估规则：根据选定的质量评估对象和评估指标配置对应的质量评估规则。
- 实施质量评估：按照配置的质量评估规则对数据实施质量评估工作。

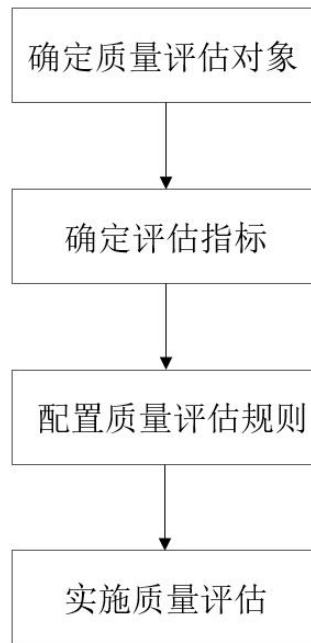


图3 数据质量评估程序

## 参 考 文 献

- [1] 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
- [2] DB52/T 1540.4-2002 政务数据 第4部分：数据质量评估规范
- [3] DB 3301/T 0364.1-2022 公共数据质量治理 第1部分：体系架构
- [4] GB/T 36344-2018 信息技术 数据质量评价指标
- [5] GB/T 38664.1-2020 信息技术 大数据 政务数据开放共享 第1部分：总则
-