

ICS 35.240.99
CCS A 00
备案号: 115203-2024

DB11

北京市地方标准

DB11/T 2317—2024

政务数据质量评估规范

Specification for government data quality evaluation

2024 - 09 - 23 发布

2025 - 01 - 01 实施

北京市市场监督管理局 发布

目 次

前言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
5 数据质量六性评估.....	2
6 数据质量五级评级.....	5
7 数据质量评估程序.....	7
参考文献.....	9

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。

本文件由北京市经济和信息化局提出。

本文件由北京市政务服务和数据管理局归口并组织实施。

本文件起草单位：北京市大数据中心、北京市政务服务和数据管理局、清华大学、北京大学、上海交通大学、北京交通大学、中国人民大学、北京邮电大学、北京工业大学、公安部第三研究所、中国工业互联网研究院、太极计算机股份有限公司、中电数据服务有限公司、华为技术有限公司。

本文件主要起草人：王睿宇、王宇航、骆亦飞、张琳、徐海琛、贾晓丰、张晰、高嵩、高文飞、张健枫、江茜、章敏、苗婕、马征、李凝云、崔鹏、徐葳、张久珍、吴帆、郑臻哲、李浥东、范举、高志鹏、林绍福、刘旭、赵琰昉、聂志锋、赵章界、赵莹、肖益、李宝东、穆显显、赵敏、杨燕、李世锋、盛小宝、田野、屈克。

政务数据质量评估规范

1 范围

本文件规定了政务数据质量“五级六性”评估模型、评估指标、评级方法、评估程序等内容。
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

数据

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

稽核表 audit from

通过数据质量稽核后存储的数据库表，表中有数据质量稽核的结果。

4 数据质量六性评估模型

数据质量六性评估模型主要从数据规范性、完整性、准确性、一致性、及时性和可用性6个维度进行了定义，评估模型见图1。

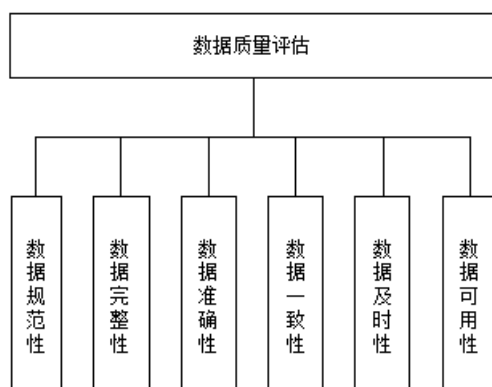


图1 数据质量六性评估模型

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

- a) 数据规范性：描述数据内容、值域、格式等符合标准数据元的情况，衡量数据标准、规范程度。
- b) 数据完整性：描述数据的描述性要素、要素属性及要素关系的完整情况，衡量数据在存储、传输和处理过程中的完整程度。
- c) 数据准确性：描述数据与其对应的客观实体的特征相一致的情况，衡量数据符合实际情况的程度。
- d) 数据一致性：描述数据类型、数据含义和数据内容的一致情况，衡量数据的一致程度。
- e) 数据及时性：描述数据获取和更新的情况，衡量数据的更新及时程度。
- f) 数据可用性：描述数据能够被访问、检索、使用和操作的水平，衡量数据的可用程度。

5 数据质量六性评估

5.1 指标编码规则

指标编码是政务数据质量评估的唯一性编码，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代表数据规范性指标、02代表数据完整性指标、03代表数据准确性指标、04代表数据一致性指标、05代表数据及时性指标、06代表数据可用性指标。二级指标由2位数字组成顺序码，范围为01-99。

5.2 数据规范性

数据规范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
0101	格式校验	数据属性值应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 $X = (1 - B/A) \times 100$ 。
0102	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性别数据，参照数据是性别码表，性别为：未知的性别0、男性1、女性2、未说明的性别9，参照性校验，即人口登记数据的性别必须为0、1、2、9，不应出现其他代码。	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 $X = (1 - B/A) \times 100$ 。
0103	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性别数据，参照数据是性别码表，性别为：未知的性别0、男性1、女性2、未说明的性别9，要求人口登记数据的性别既有男性又有女性，若性别中只有1或只有2是错误的。	定性评分	1、结果正确，则该资源得分 $X=100$ ； 2、结果错误，则该数据资源得分 $X=0$ 。
0104	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 $X = (1 - B/A) \times 100$ 。

5.3 数据完整性

数据完整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
0201	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 $X=100$ ； 2、若不一致，则该数据资源得分 $X=0$ 。
0202	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 $X = (1 - B/A) \times 100$ 。

表2 数据质量六性评估数据完整性指标（续）

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0203	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 $X=100$ ； 2、若不满足，则该数据资源得分 $X=0$ 。
0204	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 $X=(1-B/A) \times 100$ 。

5.4 数据准确性

数据准确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准确性指标

指标编号	指标名称	指标描述	示例	评分方式	评分标准
0301	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：一个身份证号码只能对应唯一的一个自然人。	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 $X=(1-B/A) \times 100$ 。
0302	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 $X=(1-B/A) \times 100$ 。
0303	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 $X=(1-B/A) \times 100$ 。

5.5 数据一致性

数据一致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
0401	逻辑校验	检验稽核表中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。	定量评分	稽核表指定稽核列中不符合逻辑关系的数量为B，稽核表限定范围内总条数为A，则该资源得分为 $X=(1-B/A) \times 100$ 。
0402	一致性校验	稽核表内数据，或表间数据是否相等。	示例1：单表单行，身份证号码的7—14位要与出生日期一致； 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。	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 $X=(1-B/A) \times 100$ 。

5.6 数据及时性

数据及时性指标定义及评分方式见表5。

表5 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
0501	更新周期合理性	描述数据资源的更新周期是否符合数据实际产生周期情况。规范政务部门把数据更新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合理。 1、若数据更新周期合理，该数据资源得分为 $X=100$ ； 2、若数据更新周期不合理，该数据资源得分 $X=0$ 。
0502	更新及时率	描述数据资源是否按照业务上数据实际产生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合理，来判断数据是否及时更新。 1、若数据是按时更新或未到更新时间，该数据资源得分 $X=100$ ； 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 $X=0$ 。

5.7 数据可用性

数据可用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
0601	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 $X=100$ ； 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 $X=0$ 。
0602	数据可理解性（数据描述文档）	指数据库表中的字段应有对应的描述文档（如数据字典、码表等），描述需准确，便于更好的理解数据。	示例：通过数据码表，将数据库表字段CSRQ与“出生日期”相对应。	定量评分	数据库表中字段对应的描述不准确或缺失的数量为 B ，数据库表中字段总数为 A ，则该资源得分 $X=(1-B/A) \times 100$ 。

6 数据质量五级评级

6.1 数据质量评分权重

6.1.1 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。

6.1.2 数据质量评估六性一级指标权重赋值占比参考，具体如下：

- a) 数据规范性权重：15%；
- b) 数据完整性权重：20%；
- c) 数据准确性权重：20%；

- d) 数据一致性权重：15%；
- e) 数据及时性权重：15%；
- f) 数据可用性权重：15%。

6.1.3 数据质量评估六性二级指标权重赋值占比参考，具体如下：

- a) 数据规范性二级指标权重赋值占比：
 - 1) 格式校验：30%；
 - 2) 照性校验-单向校验：20%；
 - 3) 参照性校验-双向校验：20%；
 - 4) 业务规则校验：30%。
- b) 数据完整性二级指标权重赋值占比：
 - 1) 字段完整性：25%；
 - 2) 空值校验：25%；
 - 3) 数据记录完整性校验：20%；
 - 4) 核心字段完整性：30%。
- c) 数据准确性二级指标权重赋值占比：
 - 1) 重复校验：40%；
 - 2) 值域校验：30%；
 - 3) 数据内容可靠性校验：30%。
- d) 数据一致性二级指标权重赋值占比：
 - 1) 逻辑校验：50%；
 - 2) 一致性校验：50%。
- e) 数据及时性二级指标权重赋值占比：
 - 1) 更新周期合理性：50%；
 - 2) 更新及时率：50%。
- f) 数据可用性二级指标权重赋值占比：
 - 1) 可直接利用性：50%；
 - 2) 数据可理解性（数据描述文档）：50%。

6.1.4 一级指标、二级指标权重总和均应是 100%。各部门可根据实际情况对二级指标进行变更或扩展。

6.2 数据质量六性评分计算

6.2.1 数据质量评估六性评分计算如下：

$$X = \sum_{i=1}^n (x_i \times w_i) \dots\dots\dots (1)$$

式中：

- X——数据一级指标得分；
- x_i ——数据二级指标得分；
- w_i ——数据二级指标权重。

注1：不同数据资源适用评估指标不同，各部门根据业务数据实际使用场景设定数据指标权重。

6.2.2 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。

6.2.3 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：

$$S = X_1 \times W_1 + X_2 \times W_2 + X_3 \times W_3 + X_4 \times W_4 + X_5 \times W_5 + X_6 \times W_6 \dots \dots \dots \quad (2)$$

式中：

- X₁——数据规范性得分；
- X₂——数据完整性得分；
- X₃——数据准确性得分；
- X₄——数据一致性得分；
- X₅——数据及时性得分；
- X₆——数据可用性得分；
- W₁——数据规范性权重；
- W₂——数据完整性权重；
- W₃——数据准确性权重；
- W₄——数据一致性权重；
- W₅——数据及时性权重；
- W₆——数据可用性权重。

6.3 数据质量等级判定

6.3.1 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。

6.3.2 每个等级的描述及分值范围见表 7。

表7 数据质量五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	S>85
B	数据质量较高，经过基础处理后可使用	80<S≤85
C	数据质量存在轻微问题，经过数据共性清洗后可使用	75<S≤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<S≤75
E	数据质量存在重大问题，须从源头进行数据整改	S≤60

7 数据质量评估程序

数据质量评估程序详见图3，包括：构建质量评估组织、确定质量评估对象、确定评估指标、配置指标权重、实施质量评估使用五阶段。具体如下：

- a) 构建质量评估组织：构建质量评估组织，明确每个部门及个人在组织中的质量评估职责。
- b) 确定质量评估对象：根据政务部门的具体业务需求，确定开展数据质量评估的数据资源（数据表级）。
- c) 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。

- d) 配置指标权重：根据选定的质量评估对象，配置一级指标和二级指标权重。
- e) 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。

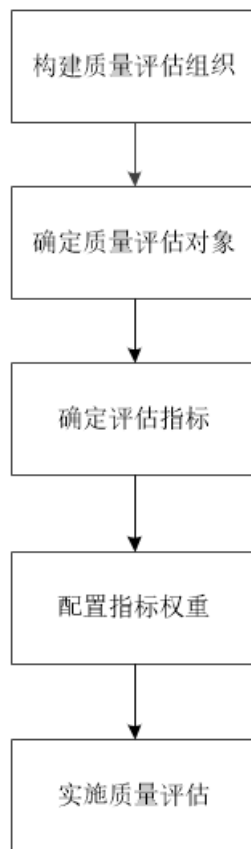


图3 数据质量评估程序

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] GB/T 36344—2018 信息技术 数据质量评价指标
- [3] 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
-