

UDC

MH

中华人民共和国行业标准

P

MH/T xxxx—xxxx

# 智慧民航数据治理 数据质量规范

Smart Civil Aviation Data Governance

Data Quality Specification

(征求意见稿)

2021-xx-xx 发布

202x-xx-xx 施行

中国民用航空局

发布

中华人民共和国行业标准

# 智慧民航数据治理 数据质量规范

Smart Civil Aviation Data Governance  
Data Quality Specification

MH/T××××—202×

(征求意见稿)

主编单位：中国民用航空局发展计划司

中国民航管理干部学院

批准部门：中国民用航空局

施行日期：202×年××月××日

中国民航出版社有限公司

2021 北京

## 前 言

数据质量是数据价值得以发挥的前提，随着民航行业数据量的飞速增长、数据类型和数据来源的不断丰富，民航各单位面临的数据质量问题也相应增加。按照《推动新型基础设施建设促进民航高质量发展实施意见》、《推动民航新型基础设施建设五年行动方案》的要求，为完善行业数据治理，规范数据质量要求，指导民航业单位进行高效的数据质量管理，实现行业数据质量的提升，特制定本规范。

在本规范的编制过程中，编写组深入调研和总结了民航行业的数据质量现状，认真吸收总结各单位数据质量管理经验，借鉴行业内外数据质量的有关标准和技术成果，经广泛征集行业意见和多次专家论证审查，最终形成本规范。

本规范共 7 章。主要内容包括：总则、术语、民航数据质量管理基本框架、民航数据质量评价规则、民航数据质量控制、民航数据质量改进、民航数据质量组织保障。

本规范的日常维护工作由中国民航管理干部学院大数据与信息管理研究中心负责，执行过程中如有意见或建议，请函告本规范日常维护组（联系人：XX；地址：北京市朝阳区花家地东路 3 号；电话：XXXX，邮箱：XXXX），以便修订时参考。

主编单位：中国民用航空局发展计划司

中国民航管理干部学院

主 编：\*\*\*\*

参编人员：\*\*\*\*

主 审：\*\*\*\*

参审人员：\*\*\*\*

## 目 次

1 总 则.....	1
2 术 语.....	2
3 民航数据质量管理基本框架.....	3
4 民航数据质量评价规则.....	5
4.1 制定步骤.....	5
4.2 评价维度.....	5
4.3 评价指标.....	6
5 民航数据质量控制.....	10
5.1 事前控制.....	10
5.2 事中控制.....	10
5.3 事后控制.....	10
6 民航数据质量改进.....	12
6.1 改进流程.....	12
6.2 问题分析.....	12
6.3 改进措施.....	13
7 民航数据质量组织保障.....	15
7.1 支撑组织分工.....	15
7.2 数据责任人分工.....	15
标准用词说明.....	17
引用标准名录.....	18

## 1 总 则

**1.0.1** 为提升民航数据质量，指导行业数据质量管理工作，实现数据质量的有效评价、控制和改进，制定本规范。

**1.0.2** 本规范适用于航空公司、机场、空管、运行保障单位、行业监管单位等民航业单位的数据质量管理工作。

**1.0.3** 民航业单位应在本规范内容的框架与指导下，结合自身发展现状及目标，进一步细化研究具体实施方案与细则，确保符合实际、具备可操作性。

**1.0.4** 民航行业数据质量管理工作，除应满足本规范的规定外，尚应符合国家、行业现行有关标准的规定。

## 2 术语

下列术语适用于本规范。

### 2.0.1 数据质量 data quality

数据质量是指数据满足应用的程度。

### 2.0.2 数据质量管理 data quality management

数据质量管理是指对数据全生命周期的每个阶段里可能引发的数据质量问题,进行识别、度量、监控、改进的一系列管理活动。包括数据质量规则构建,数据质量检测控制,数据质量改进等。

### 2.0.3 数据质量规则 data quality rule

数据质量规则是判断数据是否符合数据质量要求的逻辑约束。

### 2.0.4 数据质量控制 data quality control

数据质量控制是通过实施手段发现并反馈数据质量风险,消除或减少异常数据,满足数据质量要求。

### 2.0.5 数据质量改进 data quality improve

数据质量改进指针对数据质量问题采取措施,对现有的数据质量水平加以提高。

### 2.0.6 数据结构 data structure

数据结构是指相互之间存在着一种或多种关系的数据元素的集合和该集合中数据元素之间的关系。

### 2.0.7 数据维护 data maintenance

数据维护是指修改数据元素或数据记录定义、维护数据逻辑一致性和数据更新等。

### 3 民航数据质量管理基本框架

**3.0.1** 民航数据质量管理应根据业务问题和数据需求，着眼于关键数据和痛点问题，阶段性划定范围。

**3.0.2** 数据质量管理的基本流程为数据质量评价规则制定、数据质量控制、数据质量问题分析、改进方案制定与实施。数据质量管理基本流程如图 3.0.1 所示。

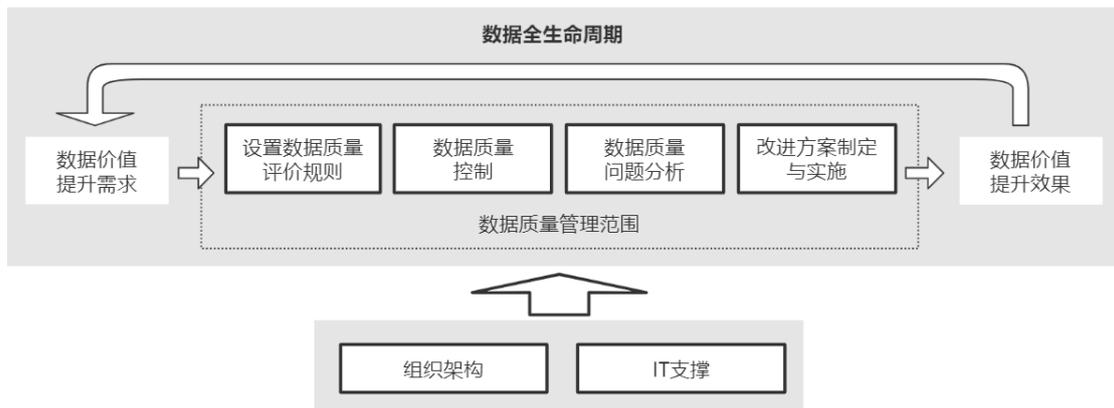


图 3.0.1 数据质量管理基本流程

1 数据质量评价规则制定：根据业务需求及数据标准要求制定用来衡量数据质量的规则。

2 数据质量控制：根据数据质量评价规则实施数据质量检查、监控等控制手段，及时发现并反馈数据质量风险。包括事前、事中和事后控制三个部分。

3 数据质量问题分析：分析数据实际质量和目标质量差距的原因，可从信息、流程、技术，人员等方面分析。

4 改进方案制定与实施：进行数据质量问题的处理与改进。

**3.0.3** 数据质量管理应贯穿数据的全生命周期，在数据的设计、产生、汇聚、应用阶段均应落实质量管理措施。

1 设计阶段：推动数据标准化的制定和执行，在数据架构建设中融入数据质量的评估。

1) 数据资产目录方面：数据资产目录应完整，业务对象应有明确的数据责任人对其全流程数据质量负责，并定义数据质量目标及数据质量工作规划。

2) 数据标准方面：数据标准应唯一，业务用途及定义应准确，数据标准在信息系统及其对应的业务流程中应得到应用。

3) 数据模型方面：物理数据模型设计应遵从逻辑数据模型设计，数据库中物理表的落

地应遵循物理模型。

4) 元数据管理方面：业务对象的元数据质量应达标，如数据分类完整性、业务定义准确性等。

2 产生阶段：应厘清各个业务源系统的数据结构和定义，利用数据模型保证数据结构完整、一致，执行数据标准、规范数据维护过程，进行数据质量检查，从源头系统保证数据的正确性、完整性、唯一性。

3 汇聚阶段：应加强数据质量问题的检查，跟踪数据质量问题的解决情况。

4 应用阶段：应根据数据的应用目标和现实情况，对现行的数据质量评估规则给予及时反馈，敦促上下游系统之间数据服务水平的提高。

## 4 民航数据质量评价规则

### 4.1 制定步骤

4.1.1 数据质量评价规则的制定包括以下步骤，数据质量评价规则制定流程如图 4.1.1 所示。



图 4.1.1 数据质量评价规则制定流程

- 1 梳理本单位数据资产概况，确定关键数据项；
- 2 分析数据质量业务需求，明确数据标准及业务规则；
- 3 设计数据质量评价的关键维度，制定数据质量评价指标；
- 4 根据业务发展需求及数据质量评价分析结果对数据质量评价规则进行验证；
- 5 将数据质量评价规则进行技术实现，实现数据质量的自动化评价及监控。

### 4.2 评价维度

4.2.1 数据质量评价规则应考虑数据标准与业务规则，宜从以下六个维度对数据质量评价规则进行描述。数据质量评价指标维度如图 4.2.1 所示。



图 4.2.1 数据质量评价指标维度

- 1 完整性：衡量数据按规则要求被完整赋值的程度。
- 2 时效性：衡量数据在时间变化中的正确程度，包括数据获取的及时程度，数据符合当前业务时序逻辑的程度。
- 3 准确性：衡量数据表示其所描述的真实实体（实际对象）真实值的程度。

4 一致性：衡量数据与其他特定上下文（不同业务、不同信息系统等）中使用的数据无矛盾的程度。

5 规范性：衡量数据与所定义的数据标准、数据模型、业务规则、元数据等规范的符合程度。

6 可访问性：衡量数据能被访问和使用的程度。

## 4.3 评价指标

### 4.3.1 完整性评价指标

完整性评价指标定义如表 4.3.1 所示。

表 4.3.1 完整性评价指标

指标名称	指标描述	计算方法
数据元素完整性	按照业务规则要求，数据集中数据元素被完整赋值的程度。	$X=A/B$ 式中： A=数据集中被完整赋值的元素个数； B=数据集中预期被完整赋值的数据元素个数
数据记录完整性	按照业务规则要求，数据集中数据记录被完整赋值的程度。	$X=A/B$ 式中： A=数据集中被完整赋值的数据记录个数； B=数据集中预期被完整赋值的数据记录个数

【条文说明】以航班地面服务保障节点为例，当“航班”作为一条数据记录时，可包括“保洁开始时间”、“保洁完成时间”、“机务放行时间”等多项数据元素。当某航班“保洁开始时间”缺失时，该航班属于不完整的数据记录，该航班的保洁开始时间属于不完整的数据元素。

### 4.3.2 时效性评价指标

时效性评价指标定义如表 4.3.2 所示。

表 4.3.2 时效性评价指标

指标名称	指标描述	计算方法
及时性	基于时间戳的记录数、频率分布或延迟时间符合业务需求的程度。	$X=A/B$ 式中： A=数据集中满足及时性要求的元素

		的个数; B=被评价的数据集中数据元素个数
时序性	数据集中同一实体的数据元素满足相对时序关系的程度。	$X=A/B$ 式中: A=数据集中满足时序性要求的元素的个数; B=被评价的数据集中数据元素个数

【条文说明】及时性示例：以“机场已占有有机位数量”为例，若规定的更新频率为1分钟/次，更新频率低于要求的数据项为不满足及时性要求的元素。

时序性示例：以航班飞行动态为例，包含“抬前轮时间”、“着陆时间”等多元素，同一航班的各时间节点应符合正确的时序关系，当“抬前轮时间”晚于“着陆时间”时，属于不满足时序性要求的元素。

### 4.3.3 准确性评价指标

准确性评价指标定义如表 4.3.3 所示。

表 4.3.3 准确性评价指标

指标名称	指标描述	计算方法
数据内容正确性	数据内容符合预期的程度	$X=A/B$ 式中: A=数据集中满足数据正确性要求的元素个数; B=被评价的数据集中数据元素个数
数据格式准确性	数据元素格式（包括数据类型、数值范围、数据长度、精度等）满足约定要求的程度。	$X=A/B$ 式中: A=数据集中满足格式要求的元素个数; B=被评价的数据集中数据元素个数
数据唯一性	衡量特定字段、记录、文件或数据集唯一存在的程度。	$X=A/B$ 式中: A=数据集中满足唯一性要求的元素个数; B=被评价的数据集中数据元素个数

【条文说明】数据内容正确性示例：以旅客订座信息为例，包含“旅客名”、“始发站航站楼”、“目的站航站楼”、“证件类型”、“证件号码”等多元素，当某元素与真实发生情况不一致时，该元素属于不满足数据内容正确性要求的元素。

数据格式正确性示例：以航班离港时间为例，若约定为十二位时间格式，则不符合该格

式要求的数据为不满足数据格式正确性的数据。

数据唯一性示例：以空管计算撤轮挡时间为例，当多套系统对特定航班生成计算撤轮挡时间时，同一时刻记录并对外公布的该航班有效计算撤轮挡时间应唯一。

#### 4.3.4 一致性评价指标

一致性评价指标定义如表 4.3.4 所示。

表 4.3.4 一致性评价指标

指标名称	指标描述	计算方法
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，数据保持一致的程度；数据发生变化时，存储在不同位置的同一数据被同步修改的程度。	$X=A/B$ 式中： A=数据集中满足相同数据一致性要求的元素个数； B=被评价的数据集中数据元素个数
关联数据一致性	关联数据符合逻辑关系约束的程度	$X=A/B$ 式中： A=数据集中满足关联数据一致性要求的元素个数； B=被评价的数据集中数据元素个数

【条文说明】相同数据一致性示例：以航班“计划取消状态”为例，在航空公司营销部门、运行控制部门的不同数据库内分别存储的同一航班的“计划取消状态”应保持一致。

关联数据一致性示例：以机场停机位数量为例，“停机位数量”、“已使用停机位数量”、“剩余停机位数量”三者之间应满足“停机位数量=已使用停机位数量+剩余停机位数量”的逻辑关系。

#### 4.3.5 规范性评价指标

规范性评价指标定义如表 4.3.5 所示。

表 4.3.5 规范性评价指标

指标名称	指标描述	计算方法
数据标准	数据符合数据标准的程度。 注：评价时应收集数据在命名、创建、定义、更新和归档时遵循的标准，包括国际标准、国家标准、行业标准、地方标准或相关规定等。	$X=A/B$ 式中： A=数据集中满足数据标准要求的数据元素个数； B=被评价的数据集中数据元素个数
数据模型	数据符合数据模型规范的程度。	$X=A/B$ 式中： A=数据集中满足数据模型要求的元

		素个数; B=被评价的数据集中数据元素个数
元数据	数据符合元数据定义的程度。	$X=A/B$ 式中: A=数据集中被元数据描述的元素个数; B=被评价的数据集中数据元素个数
业务规则	数据符合业务规则的程度。	$X=A/B$ 式中: A=数据集中满足业务规则的元素个数; B=被评价的数据集中数据元素个数
安全规范	数据满足安全和隐私方面规则的程度, 包括数据权限管理, 数据脱敏处理等。	$X=A/B$ 式中: A=数据集中满足安全规范的元素个数; B=被评价的数据集中数据元素个数

【条文说明】数据标准、数据模型、元数据规范参照《智慧民航数据治理数据架构规范》(MH/T\*\*\*).

#### 4.3.6 可访问性评价指标

可访问性评价指标定义如表 4.3.6 所示。

表 4.3.6 可访问性评价指标

指标名称	指标描述	计算方法
可访问性	数据能够被正常获取的程度	$X=A/B$ 式中: A=数据集中满足可访问性要求的元素个数; B=被评价的数据集中数据元素个数
可用性	数据在有效生命周期内能够被正常使用的程度	$X=A/B$ 式中: A=数据集中满足可用性要求的元素个数; B=被评价的数据集中数据元素个数

【条文说明】数据由于外部权限设置、网络中断等原因, 无法被正常获取, 则不满足可访问要求; 数据由于损坏, 无法在系统中打开使用, 则不满足可用要求。

## 5 民航数据质量控制

### 5.1 事前控制

5.1.1 数据质量事前控制包括对质量规则的检查和数据源系统的检查,包括但不限于以下内容:

- 1 数据上线前基于元数据对象质量规则进行上线数据的质量检查。
- 2 数据业务规则合理性检查。
- 3 核对跨系统数据的一致性。
- 4 检查源系统变更情况,包括表结构变更情况、字段值变更情况。

### 5.2 事中控制

5.2.1 数据质量事中控制应在数据流转和应用过程中设置关键质量控制节点,针对潜在数据质量风险制定数据质量检测计划,基于数据的质量规则对计划范围内的数据进行稽核。针对质量未达标的数据应发出告警,针对质量问题严重的数据宜设定熔断机制终止数据流转。数据质量事中监控流程如图 5.2.1 所示。

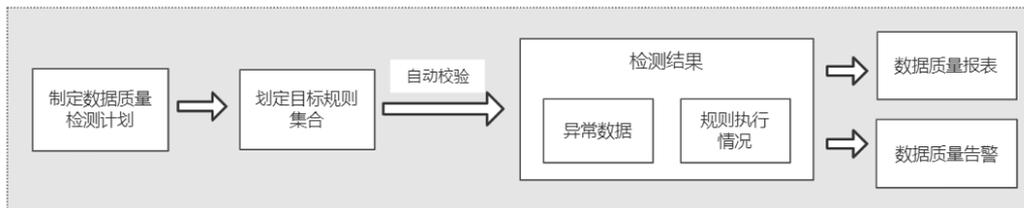


图 5.2.1 数据质量事中监控流程

5.2.2 数据质量事中控制包括但不限于以下内容:

- 1 对运行过程中的数据进行质量监控,基于数据质量评价规则对运行结果稽核并进行错误告警。
- 2 对质量问题告警进行定位分析、相似问题分析,利用元数据的血缘分析定位问题数据路径接口节点及关联数据。
- 3 对数据质量问题进行初步的分类、统计,形成数据质量报表。
- 4 应支持对问题数据的增、删、改、查、熔断。

### 5.3 事后控制

**5.3.1** 数据质量事后控制的重点是对事前和事中控制阶段暴露的质量问题进行归纳总结，包括但不限于以下内容：

1 将生产运行中系统监控到的数据质量问题、人员报告的数据质量问题、维护发现的数据质量问题统一汇聚到数据质量问题库，对问题的原因和解决方法进行分类。

2 按周期形成质量分析报告，包括质量问题和影响，以及问题处理的情况等内容。

3 针对数据完整性、数据唯一性、数据合法性、编码检查、统计口径等形成数据质量评估报告。

## 6 民航数据质量改进

### 6.1 改进流程

6.1.1 数据质量改进应包括以下步骤，数据质量改进流程如图 6.1.1 所示。

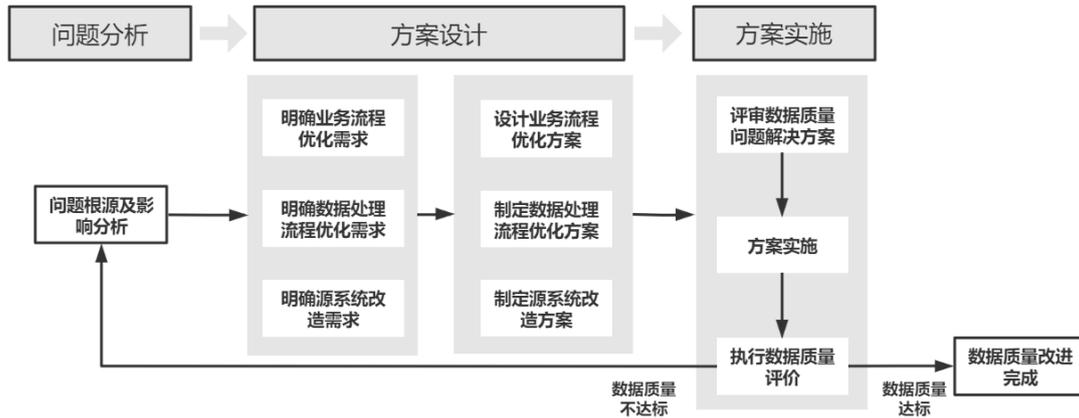


图 6.1.1 数据质量改进流程

- 1 问题分析：根据数据质量评价分析结果，进行数据质量问题原因分析和影响分析。
- 2 方案设计：针对数据质量问题分析结果，明确业务流程、数据处理流程和源系统等方面的优化和改造需求，制定数据质量问题解决方案。
- 3 方案实施：数据质量问题解决方案经审核通过后，进行推广执行，并对执行结果进行数据质量评价，对不满足质量要求的数据进行持续优化，形成数据质量改进闭环。

### 6.2 问题分析

6.2.1 数据质量问题分析应识别导致数据质量问题的根本原因，包括但不限于人员、流程、技术和信息等方面原因，如图 6.2.1 所示。

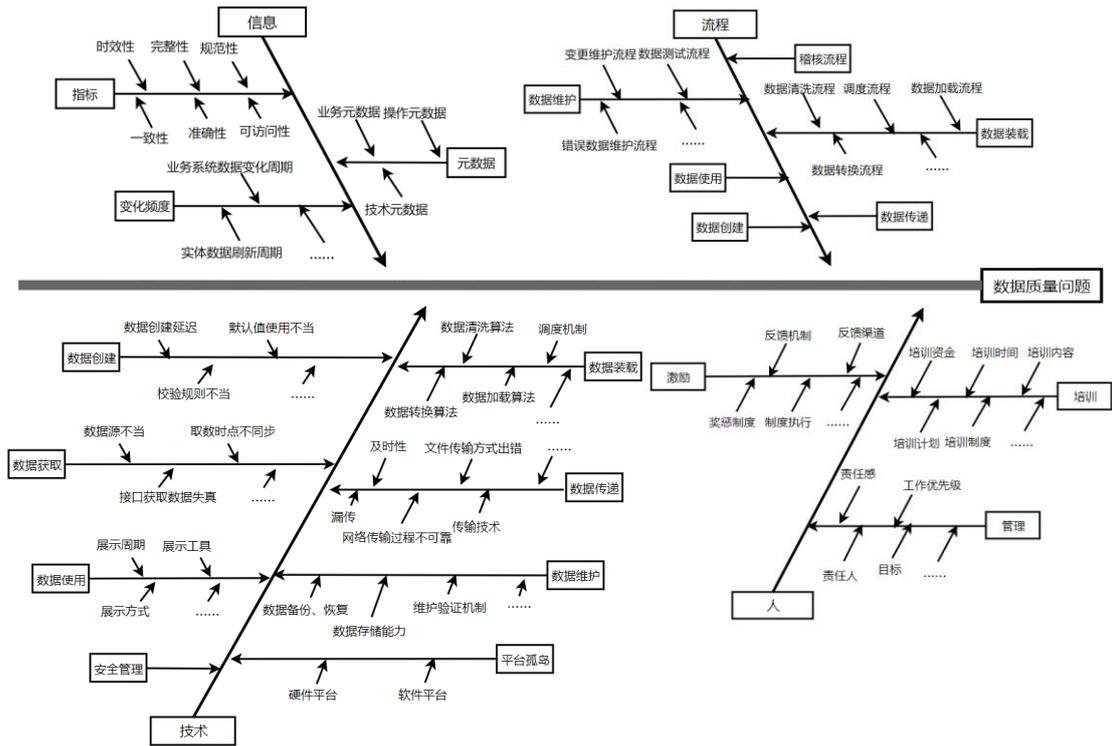


图 6.2.1 数据质量原因鱼骨图

- 1 人员原因：由于人员素质及管理机制方面的原因造成的数据质量问题。
- 2 流程原因：由于信息系统作业流程和人工操作流程设置不当造成的数据质量问题，主要来源于信息系统内数据的创建、传递、装载、使用、维护 and 稽核等流程。
- 3 技术原因：由于数据处理各技术环节的异常造成的数据质量问题。如数据校验规则设计不当、数据源接入不当、数据存储能力有限等，主要涉及数据创建、数据获取、数据传输、数据装载、数据使用、数据维护等环节。
- 4 信息原因：由于数据规范性异常造成的数据质量问题。包括元数据标准异常、数据质量规则异常、变化频度异常等因素。

## 6.3 改进措施

**6.3.1** 应针对由人员、流程、技术和信息原因产生的质量问题，分别采取相应改进措施。

- 1 人员原因：应明确工作目标、角色和职责，加强人员培训，建立激励机制，提升人员的积极性、责任心和业务能力，在源头上预防和控制人员原因导致的数据质量问题。
- 2 流程原因：应在数据流转过程中建立监控数据转换质量的流程和体系，控制问题数据在系统中的流转。
- 3 技术原因：应针对数据质量问题所在环节，从数据接口、数据清洗、数据转换、数

据装载等方面提升技术配置与技术能力。

4 信息原因：应在源头建立数据标准或接入标准，规范数据定义，确定数据来源及范围，并在信息系统实施中严格遵循相关标准。

## 7 民航数据质量组织保障

### 7.1 支撑组织分工

**7.1.1** 单位数据管理组织负责以下数据质量管理工作：

- 1 建立本单位数据质量工作机制，建立内部沟通协调机制；
- 2 制定本单位数据质量管理工作规划，制定数据质量目标；
- 3 组织数据质量管理工作开展，发布本单位数据质量管理报告；
- 4 组织评审数据质量标准及评价规则，验收数据质量问题情况；
- 5 确保配备数据质量管理体系所需的适宜资源，建立健全保障机制；
- 6 建立数据质量监控体系，覆盖数据全生命周期，对数据质量持续监测、分析、反馈和纠正；
- 7 建立数据质量考核评价体系，组织数据质量的考核评价，考核结果纳入本单位绩效考核体系。

**7.1.2** 各业务数据管理组织负责以下数据质量管理工作：

- 1 基于所辖业务的数据质量管理工作要求，拟定数据质量标准并设计数据质量评价规则，执行数据质量评价；
- 2 组织各业务领域专家，分析数据质量问题根因，制定改进举措及闭环管理；
- 3 执行适用的法规、标准及其他要求。

**7.1.3** 数据质量工作组负责监督和评价单位各业务领域的的数据质量，负责以下数据质量管理的工作：

- 1 评审各业务领域数据质量标准及评价规则；
- 2 考核评价各业务领域数据质量。

### 7.2 数据责任人分工

**7.2.1** 各单位最高数据责任人是本单位数据质量管理工作的最高决策者，负责以下工作：

- 1 下达数据质量目标；
- 2 签发数据质量评价及考核结果；
- 3 基于数据质量结果及改进状况，对相应领域数据责任人进行奖励及问责。

**7.2.2** 各领域数据责任人是所辖领域数据质量的负责人，负责以下工作：

- 1 承接本单位最高数据责任人设定的数据质量目标；
- 2 建立、实施、保持和持续改进所辖领域数据质量管理体系；
- 3 向最高数据责任人报告数据质量管理体系的绩效和改进需求；
- 4 基于所辖领域的的数据质量结果及改进状况，对相应的业务数据责任人进行奖励及问责；
- 5 负责与所辖领域数据质量管理的相关外部联系工作。

**7.2.3** 各业务数据责任人是所辖业务数据质量的负责人，负责以下工作：

- 1 承接所属领域数据责任人设定的数据质量目标；
- 2 建立、实施、保持和持续改进所辖业务数据质量管理体系；
- 3 向所属领域数据责任人报告数据质量管理体系的绩效和改进需求；
- 4 明确所辖业务数据质量问题改进责任人，并推动问题闭环管理；
- 5 负责与所辖业务数据质量管理工作的相关外部联系工作。

## 标准用词说明

1 为了便于在执行本规范条文时区别对待，对要求严格程度不同的用词，说明如下：

1) 表示很严格，非这样做不可的用词：

正面词采用“必须”，反面词采用“严禁”。

2) 表示严格，在正常情况下均应这样做的用词：

正面词采用“应”；反面词采用“不应”或“不得”。

3) 表示允许稍有选择，在条件许可时首先这样做的用词：

正面词采用“宜”；反面词采用“不宜”。

4) 表示有选择，在一定条件下可以这样做的，采用“可”。

2 本规范中指定应按其他有关标准、规范执行时，写法为“应符合……的规定”或“应按……的规定执行”。

## 引用标准名录

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用。于本文件凡是不注日期的引用文件，其最新版本（包含所有修改单）适用于本文件。

- [1] 新时代民航强国建设行动纲要
- [2] 推动新型基础设施建设促进民航高质量发展实施意见
- [3] 推动民航新型基础设施建设五年行动方案
- [4] 信息技术服务 治理 第 5 部分：数据治理规范（GB/T 34960.5）
- [5] 信息技术 数据质量评价指标（GB/T 36344）
- [6] 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（GB/T 25000.12）
- [7] 工业数据质量 通用技术规范（GB/T 39400）
- [8] 数据质量控制与评估原则（SY/T 7005）