
上海教育信息化规范

编号：SJX-DAT/T 002-2021

上海教育数据质量管理规范 (试行)

(本发布版最后修订日期： 2021-8-10)

上海市教育委员会信息中心

目 录

1. 总则	2
2. 评估维度	4
3. 职责分工	4
4. 数据采集和归集	6
5. 数据检查	6
6. 数据维护	7
7. 数据质量考核	8
8. 保障机制	8
9. 附则	9
附件一 数据归集的质量管理流程	10
附件二 数据检查实施示例	11
附件三 数据质量反馈表	14
附件四 数据质量处理流程	15

1. 总则

1.1 目的

为提高上海教育数据质量，明确各部门数据质量管理的责权利和流程规范，确保数据的准确性、完整性、及时性、一致性、规范性和唯一性，根据有关法律、法规和《教育部机关及直属事业单位教育数据管理办法》（教发厅〔2018〕1号）、《上海市公共数据和一网通办管理办法》（上海市人民政府令 第9号）、《上海市加快推进数据治理促进公共数据应用实施方案》（沪委办〔2019〕8号）、《上海教育系统加快推进数据治理和“一网通办”工作实施方案》（沪教委信息〔2019〕31号）、《上海教育数据管理办法（试行）》等文件规定，结合本市教育工作实际，制定本规范。

1.2 适用范围

本规范适用于上海教育数据质量管理工作。本规范所指上海教育数据包括上海市教育委员会（以下简称“市教委”）、各区教育局及其辖区内的公办中小学和幼儿园、各市属公办高等学校（以下简称“各高等学校”）和各市属公办中等职业学校（以下简称“各中职校”）等教育单位（部门）（以下简称“各级各类教育单位”）在履行职责过程中产生、采集和使用的各类非涉密数据。

1.3 参考资料

《上海教育数据管理办法（试行）》

《上海教育管理基础信息分类与代码(试行)》

《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》

《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》

《GB/T 36344-2018 信息技术 数据质量评价指标》

1.4 基本原则

（一）权责明确原则

各部门之间的职责和权限边界清晰，各司其职，各尽其力，共同推动数据质量管理有序进行。

（二）统一规范原则

各部门的数据质量管理过程，应符合上海教育数据标准体系框架内的相关标准和规范。

（三）全程监管原则

建立涵盖数据采集和归集、数据检查、数据维护的全过程监管体系，确保数据全生命周期的准确、完整、及时、一致、唯一。

（四）持续改进原则

数据质量管理是一个动态长期过程，遵循PDCA（策划-实施-检查-处理）理念，逐步、持续改进数据质量。

1.5 术语和定义

下列术语和定义用于本规范。

1.5.1

数据质量

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。

1.5.2

数据质量管理

对数据全生命周期各个阶段里可能引发的各类数据质量问题，进行识别、度量、监控、预警等一系列管理活动，并通过改善和提高组织的管理水平使数据质量维持在较高水平。

1.5.3

数据采集

指数据提供部门根据业务管理职能需要，通过信息系统或其他手段进行数据获取的行为。

1.5.4

数据归集

指数据提供部门将依法履职过程中采集和产生的数据根据数据管理要求集中传输到本级数据资源管理技术平台（下文简称“数据平台”）的行为。

1.5.5

数据检查

对数据的准确性、完整性、及时性、一致性、唯一性等特性进行质量检查的行为。

1.5.6

数据维护

对已经存储于数据库中质量不合格数据进行整改的行为，如字段状态或属性更正。

2. 评估维度

数据质量评估维度是数据质量的评估标准和定义约束规则的依据。根据业务需求，参考GB/T 25000.12-2017国家标准，本规范选取以下教育数据评估维度。

2.1 准确性

准确性（Accuracy）是在特定的使用环境中，数据具有正确地表示一个概念或事件的相关属性真实值的属性的程度。数据的准确性需要将数据与其所描述的实际对象比较，数据应符合业务规则和统计口径。

2.2 完整性

完整性（Completeness）指在特定使用环境中，与一个实体相关联的主题数据具有所有预期属性值和相关实例值的程度。该维度用于度量哪些数据是缺失的或者哪些数据不可用。

2.3 及时性

及时性（Currentness）指在特定的使用环境中，数据具有表征其正确寿命的属性的程度。由于数据的数值随时间而不断变化，因此该维度的评估检查需要贯穿数据全生命周期。

2.4 一致性

一致性（Consistency）指在特定的使用环境中，数据具有表征其与其它数据无矛盾且连贯的属性的程度。该维度是对不同数据仓库、应用和系统中所存储或使用信息等价程度的测量，以及使数据等价的处理流程的测量标准。满足一致性是不同系统间数据集成的基础。

2.5 唯一性

唯一性（Uniqueness）指数据或者属性的重复程度。保证唯一性能避免数据冗余，同时也减少了出现数据不一致的潜在可能性。

3. 职责分工

3.1 数据管理协调部门

数据管理协调部门是统筹规划、协调推进、监督指导各级教育数据管理工作的部门。上海市教育委员会信息化工作处是负责市级教育数据管理协调和监督审核工作的责任部门。

教育数据管理协调部门的主要职责如下：

- （一）制定本市教育数据质量管理的相关规章制度；
- （二）对各部门的履职情况进行监督、考核；
- （三）对违反数据质量管理规定，造成数据错误、失真、延误、遗漏等后果的行为责令改正。

3.2 数据提供部门

数据提供部门根据业务职能产生某类数据，是该类数据的质量管理责任部门，主要职责如下：

- （一）遵守本市教育数据质量管理的相关规章制度；
- （二）负责职责范围内所采集和归集数据的更新、维护等工作；
- （三）负责提供归集数据的质量检查规则，为技术管理部门提供业务支持；
- （四）当数据发生变更时，应及时通知数据管理协调部门，数据技术管理部门和数据使用部门。涉及数据结构变化时，应对变更后所涉及到的业务情况提供解释；
- （五）应及时整改数据质量报告中指出的、以及数据使用部门反馈的数据质量问题；
- （六）定期开展本部门数据质量自查工作，及时发现、解决数据质量管理中出现的问题。

3.3 数据技术管理部门

数据技术管理部门是指为提供数据服务而实施教育数据质量监管、承担数据质量管理的技术支撑保障工作的部门，主要职责如下：

- （一）遵守本市教育数据质量管理的相关规章制度；
- （二）协助数据提供部门和使用部门完成数据采集、归集及维护工作；
- （三）负责与数据管理协调部门共同建立数据质量考核体系，积极开展数据质量考核工作，督促各部门的质量管理工作；
- （四）实施数据质量检查、实时监测和全面评价，负责数据平台的数据质量校核和质量约束规则库维护；
- （五）定期发布数据质量管理及考核报告；
- （六）及时向数据提供部门通报出现的数据质量问题，并督促其解决；

(七) 定期向数据管理协调部门报告数据质量管理执行情况。

3.4 数据使用部门

数据使用部门主要职责如下：

(一) 遵守本市教育数据质量管理的相关规章制度；

(二) 校核共享数据的数据质量。如在法定履职过程中使用数据时，发现数据不满足原需求或存在数据质量问题，及时向数据提供部门反馈，配合数据提供部门进行数据维护。

4. 数据采集和归集

4.1 质量管理流程

数据采集和归集过程中的数据质量管理由数据提供部门负责，数据技术管理部门协助，其质量管理流程详见附件一。

4.2 实施细则

(一) 数据提供部门根据数据管理协调部门制定的标准规范，严格执行数据采集工作，加强对采集人员的管理、工作情况监督和培训，对数据进行清洗加工以及标准化之后再行采集工作，确保数据采集符合各项评估维度；

(二) 当数据发生变更（包括但不限于以下情况：数据结构、数据长度、数据类型、字典表、验证逻辑变化等情况），数据提供部门应提前通知数据管理协调部门解释变更后的业务情况，再由数据管理协调部门进一步依据数据使用情况同步通知到数据技术管理部门和数据使用部门。数据提供部门还应配合数据技术管理部门进行原数据集成接口调试，保障数据集成服务能继续正常运行；

(三) 当数据源结构变化对《上海教育管理基础信息分类与代码（试行）》产生影响时，且数据提供部门有意愿将变更结构加入到《上海教育管理基础信息分类与代码（试行）》，则应遵照《上海教育数据管理办法（试行）》的指导，提出数据标准变更申请；

(四) 数据技术管理部门负责接入数据平台数据的审核，做好数据检查、情况反馈、督促整改、采集入库等工作。

5. 数据检查

5.1 检查方法

- (一) 应用数据质量管理工具检查；
- (二) 数据库及系统程序的自动化控制措施，实时或定期监测关键数据的质量；
- (三) 手工复料，抽取原始档案材料与数据库中数据进行对比；
- (四) 其它数据检查方法。

数据检查实施示例详见附件二

5.2 实施细则

(一) 数据技术管理部门按照“多源校核、动态更新”的原则，依据数据质量评估维度制定约束规则库，并配置入数据质量管理工具；数据提供部门有义务配合数据技术管理部门制定数据质量检查的业务规则；

(二) 数据技术管理部门采用全检或抽验的方式对接入数据进行质量检查，形成数据质量报告，发送至数据提供部门，并督促及时整改维护质量不合格数据；

(三) 建立数据质量可追溯机制，并保持上述质量检查的全部证实性记录；

(四) 当数据发生与数据质量有关的业务或技术变更时，数据提供部门需要通知数据技术管理部门进行对应的数据质量检查规则变更。

6. 数据维护

6.1 维护内容

(一) 数据更正，包括但不限于对数据的错误、失真、延误、遗漏等情况的补全与纠错；数据补全，根据客观事实按照通用标准明确业务流程、数据描述主体和补充缺值信息；数据纠错，根据“一数一源”要求明确数据来源，校准数据内容；

(二) 数据标准变更，包括但不限于数据长度、数据类型、数据范围、字典表、验证逻辑等；

(三) 数据结构或接口变更；

(四) 数据质量约束规则库的维护。

6.2 实施细则

(一) 数据维护前应做好数据和系统的备份工作；

(二) 数据使用部门在数据使用过程中, 应建立相应措施和机制, 对所用数据质量进行评价, 及时向数据提供部门反馈质量问题, 反馈表详见附件三;

(三) 数据提供部门应在收到问题通知后的5个工作日内, 对数据进行维护, 并向数据技术管理部门反馈, 数据质量问题处理流程详见附件四;

(四) 数据维护人员在进行数据维护时, 必须认真负责, 避免在数据维护过程中产生新的错误数据;

(五) 数据维护工作应严格备案, 数据技术管理部门对各部门报送的数据维护的时间、内容、原因、责任人等记录进行备案, 涉及的书面材料必须登记存档。

7. 数据质量考核

(一) 由数据管理协调部门、数据技术管理部门共同建立数据质量考核体系, 积极开展数据质量考核工作, 督促各部门的质量管理工作;

(二) 数据技术管理部门应从准确性、完整性、及时性、一致性、唯一性五个评估维度对各部门数据质量进行考核;

(三) 数据技术管理部门应根据约定的质量检查规则, 保持每月一次的频率对各数据提供部门的数据质量进行考核并出具数据质量报告, 报告内容包括考核对象、考核结果、问题种类以及考核时间等;

(四) 考核发现的数据质量问题应及时通知数据提供部门及时进行整改。

8. 保障机制

8.1 资源保障机制

各级各类教育单位应为实施、保持并持续改进质量管理体系提供适宜的资源保障, 并建立保障机制。

(一) 明确数据质量管理组织、配备具有专业知识和能力的责任人及数据质量管理专员;

(二) 开展数据质量管理工作所涉及的数据标准制定与优化、系统建设与改造、数据质量自查与校核、数据质量监管以及考核评估等相关经费保障, 应当根据实际情况纳入财政资金预算;

(三) 配备所需的设备、设施及相应的环境条件;

(四) 配备所需的测试工具和监测装置;

(五) 具备质量管理工作的经验、技术和方法。

8.2 能力保障机制

为使相关人员能够胜任数据质量管理工作，保证所采集数据的质量，各级各类教育单位应：

（一）确保本部门人员认识到数据质量工作的业务相关性和重要性，具备为实现质量目标作贡献的责任意识；

（二）确保从事数据质量管理工作的人员具备相应的能力；

（三）对从事数据采集、检查、考核的人员进行岗位专业技能培训；

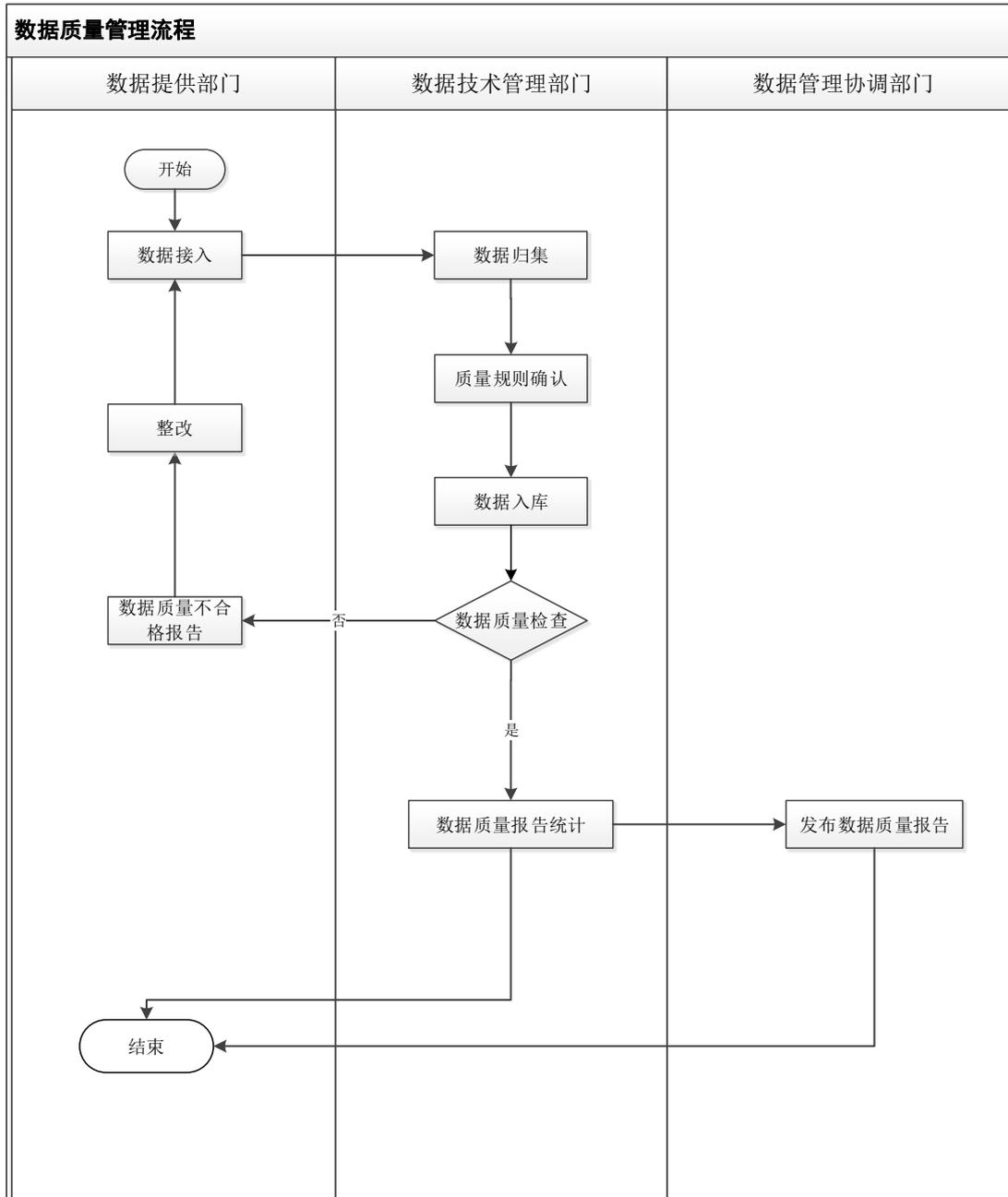
（四）保持人员教育、培训、技能和经验的适当记录。

9. 附则

本办法由上海市教育委员会信息中心负责解释。

本办法自印发之日起执行。

附件一 数据归集的质量管理流程



附件二 数据检查实施示例

1. 数据质量约束规则

一级规则维度	二级规则维度	说明
准确性	实体准确性约束	描述数据信息与客观实体一致，如数据所描述的学生学籍信息与其实际情况完全相符。
	数据倾斜的合理性约束	描述字段的取值分布主要集中在某个特定类别或者特定区间的逻辑合理性，如某区幼儿园评级记录均为一级，则可能与实际不符。
	长度约束	描述检核对象的长度是否满足长度约束，如身份证位数18位。
	取值范围约束 (离散型)	描述检核对象的值是否满足取值约束，如性别码只能取1(男性)，2(女性)。
	取值范围约束 (连续型)	描述检核对象的取值是否在预定义的范围，如身份证只能包含0-9的数字或X。
	数据类型约束	描述检核对象的数据类型是否正确，如学生的出生年月应当为日期型数据类型。
	数据精度约束	描述检核对象属性值与精度要求(小数点位数等)的符合程度，如学生单科成绩一般为整数，无小数点。
	来源约束	描述数据由有资质组织提供的程度，如“一数一源”原则。
完整性	非空约束	描述是否存在该检核对象信息、描述检核对象是否存在空的情况。
	非缺省值约束	描述检核对象是否取缺省值的情况，如日期类的空值赋值为9999，那么取值为9999的数据信息是不完整的。
	数据规模的合理性约束	描述是否存在数据记录缺失的情况。如一个班级只有几个学生，那么很可能记录缺失。
及时性	更新频率约束	描述数据按所需频率被更新的程度，如从学生信息采集到可以查看学生信息的时间差不得大于某一临界值。
	时序约束	描述数据集中同一实体的数据元素之间的相对时序关系。
一致性	存在一致性约束	描述同一实体或者属性值在关联数据集中是否一致，如同一学生在不同数据集中性别应当一致。
	逻辑一致性约束	描述同一实体在关联数据集中逻辑关系是否一致，如第一名的总分应当大于等于第二名的总分。
	等值一致性约束	描述某一属性值与一个或多个属性值在通过函数或计算公式计算得出的结果是否一致，如单个学生总分应当等于各科成绩合计。
	变更一致性约束	描述某一属性值发生变更时，相关属性值的变化情况，如单科成绩变更时，相应的平均分及总分应当随之改变。

	代码一致性约束	描述检核对象的代码值是否在对应的代码表内，如身份证件类型的值域为GB/T 29808—2013中7.100。
唯一性	关键字段唯一性约束	描述同一客观实体在系统是否出现超过一条的记录，如主索引字段重复，说明同一学生出现了超过一条的记录，且记录之间可能信息冲突。
	非所有字段重复约束	描述同一客观实体是否重复记录，如所有字段重复，说明同一学生重复记录了多次。

(一) 准确性符合：

- 1、应严格依据所描述的实际对象准确录入相关数据，符合相关标准规定的逻辑校验关系，符合教育管理活动规定的语境关系，符合数据项名称表征的客观实体，如实反映，不得随意修改、增减；
- 2、可借助可靠的外部数据源进行交叉验证；
- 3、可依据内部的原始资料进行手工检查。

(二) 完整性符合：

- 1、按照业务系统的有关要求数据进行采集，不得缺省，保证数据齐全，避免数据缺失；
- 2、业务系统的数据采集界面应设置必要的填写约束，如必填项提示和缺失情况下的禁止跳转规则等。

(三) 及时性符合：

- 1、数据要在规定的时间内完成采集，不得无故拖延或推迟数据采集时间，确保业务系统数据及时反映教育管理实际；
- 2、按照规定的更新时间要求对数据进行更新。

(四) 一致性符合：

按照多源校核、动态更新的原则，对数据平台的数据进行校核、确认，确保同一数据在同一业务系统不同表格或不同系统间属性及逻辑保持一致。

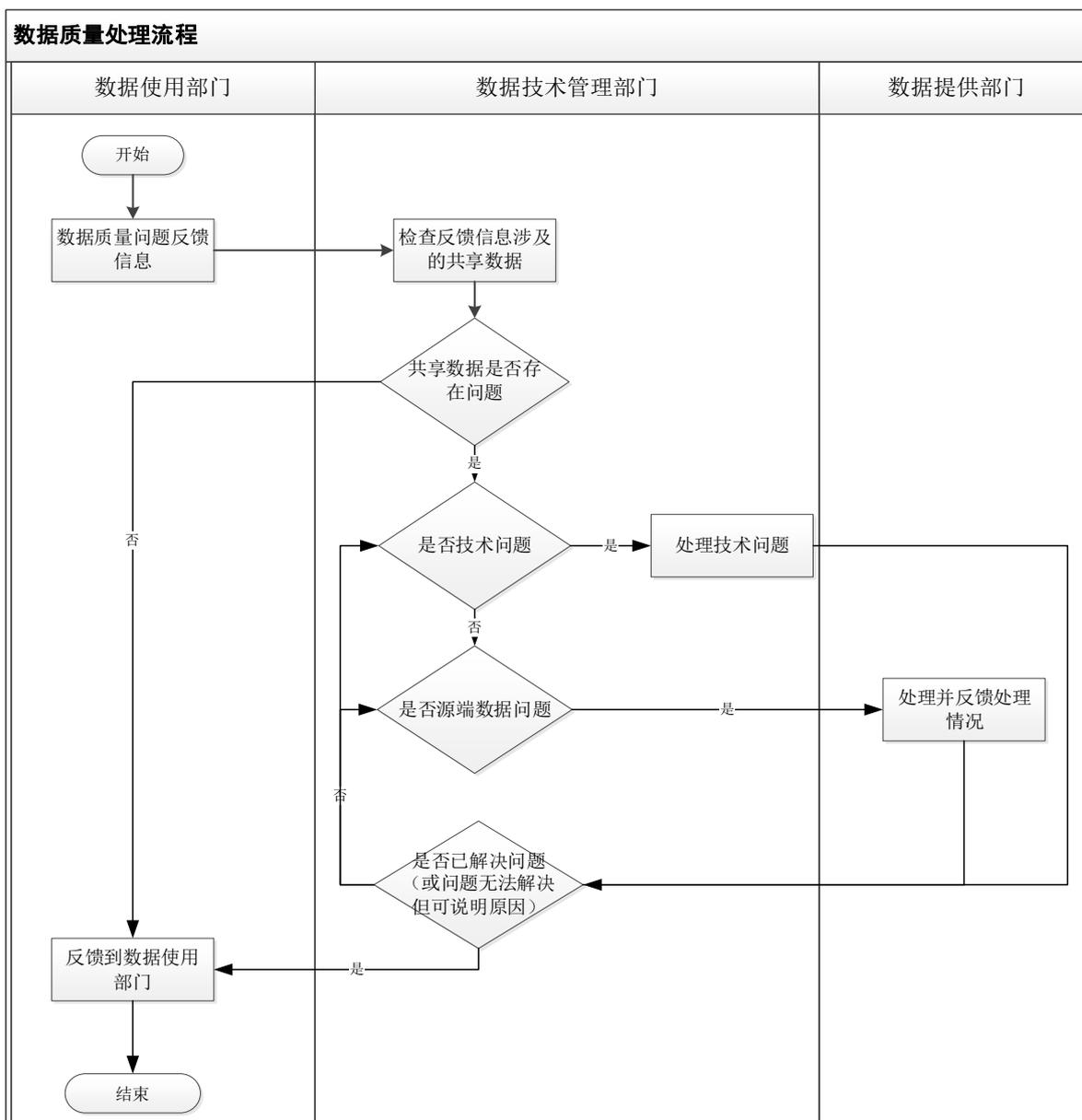
(五) 唯一性符合：

- 1、按照一数一源要求采集数据，明确各类数据的权威数据源；
- 2、严格按照规定进行岗位设置和授权，严格按照岗位和权限操作。严禁在未授权的情况下委托他人以本人的账户和口令进行数据录入和修改，定期更改口令，确保数据录入节点的唯一性。

2. 数据质量约束规则模板

数据项名称	字段名称	准确性	完整性	及时性约束	一致性约束	唯一性约束
更新时间	updateTime	数据类型 (timestamp) 数据精度 (毫秒)	非空	更新频率 (≤ 24 小时) 时序 ($\geq createTime$)	无	无

附件四 数据质量处理流程



注：数据质量问题反馈信息是指各数据使用部门在数据使用过程中对数据质量问题的反馈，例如大数据中心数据纠错工单等。