**基于大数据的数据治理**

**建设方案**

目 录

[一 数据管理的现状 3](#_Toc409713748)

[二 数据治理的概述 4](#_Toc409713749)

[（一） 数据治理概念 4](#_Toc409713750)

[（二） 数据治理目标 5](#_Toc409713751)

[三 数据治理体系 5](#_Toc409713752)

[四 数据治理核心领域 6](#_Toc409713753)

[（一） 数据模型 6](#_Toc409713754)

[（二） 数据生命周期 6](#_Toc409713755)

[（三） 数据标准 8](#_Toc409713756)

[（四） 主数据 9](#_Toc409713757)

[（五） 数据质量 10](#_Toc409713758)

[（六） 数据服务 12](#_Toc409713759)

[（七） 数据安全 12](#_Toc409713760)

[五 数据治理保障机制 13](#_Toc409713761)

[（一） 制度章程 13](#_Toc409713762)

[(1) 规章制度 13](#_Toc409713763)

[(2) 管控办法 13](#_Toc409713764)

[(3) 考核机制 13](#_Toc409713765)

[（二） 数据治理组织 15](#_Toc409713766)

[(1) 组织架构 15](#_Toc409713767)

[(2) 组织层次 16](#_Toc409713768)

[(3) 组织职责 17](#_Toc409713769)

[（三） 流程管理 19](#_Toc409713770)

[（四） IT技术应用 19](#_Toc409713771)

[(1) 支撑平台 19](#_Toc409713772)

[(2) 技术规范 22](#_Toc409713773)

[附件A 数据管理规范 23](#_Toc409713774)

[附件B 数据质量评估办法 38](#_Toc409713775)

[附件C 数据质量管理流程 42](#_Toc409713776)

关于数据治理的理解

# 数据管理的现状

根据行业信息化发展的现状，结合当今行业数据治理的要求，大型集团或政务管理部门现阶段数据管理方面存在以下的不足：

（1） 数据多头管理，缺少专门对数据管理进行监督和控制的组织。信息系统的建设和管理职能分散在各部门，致使数据管理的职责分散，权责不明确。组织机构各部门关注数据的角度不一样，缺少一个组织从全局的视角对数据进行管理，导致无法建立统一的数据管理规程、标准等，相应的数据管理监督措施无法得到落实。组织机构的数据考核体系也尚未建立，无法保障数据管理标准和规程的有效执行。

（2） 多系统分散建设，没有规范统一的省级数据标准和数据模型。组织机构为应对迅速变化的市场和社会需求，逐步建立了各自的信息系统，各部门站在各自的立场生产、使用和管理数据，使得数据分散在不同的部门和信息系统中，缺乏统一的数据规划、可信的数据来源和数据标准，导致数据不规范、不一致、冗余、无法共享等问题出现，组织机构各部门对数据的理解难以应用一致的语言来描述，导致理解不一致。

（3） 缺少统一的主数据，组织机构核心系统间的人员等主要信息并不是存储在一个独立的系统中，或者不是通过统一的业务管理流程在系统间维护。缺乏对集团公司或政务单位主数据的管理，就无法保障主数据在整个业务范围内保持一致、完整和可控，导致业务数据正确性无法得到保障。

（4） 缺乏统一的集团型数据质量管理流程体系。当前现状中数据质量管理主要由各组织部门分头进行；跨局跨部门的数据质量沟通机制不完善；缺乏清晰的跨局跨部门的数据质量管控规范与标准，数据分析随机性强，存在业务需求不清的现象，影响数据质量；数据的自动采集尚未全面实现，处理过程存在人为干预问题，很多部门存在数据质量管理人员不足、知识与经验不够、监管方式不全面等问题；缺乏完善的数据质量管控流程和系统支撑能力。

（5） 数据全生命周期管理不完整。目前，大型集团或政务单位，数据的产生、使用、维护、备份到过时被销毁的数据生命周期管理规范和流程还不完善，不能确定过期和无效数据的识别条件，且非结构化数据未纳入数据生命周期的管理范畴；无信息化工具支撑数据生命周期状态的查询，未有效利用元数据管理。

# 数据治理的概述

## 数据治理概念

数据治理是指将数据作为组织资产而展开的一系列的具体化工作,是对数据的全生命周期管理。

数据治理体系是指从组织架构、管理制度、操作规范、IT应用技术、绩效考核支持等多个维度对组织的数据模型、数据架构、数据质量、数据安全、数据生命周期等各方面进行全面的梳理、建设以及持续改进的体系。

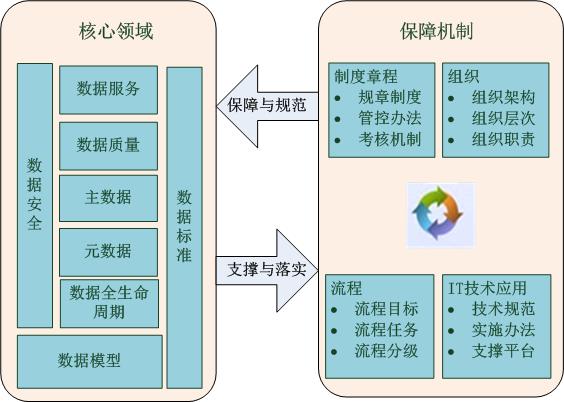
## 数据治理目标

数据治理的目标是提高数据的质量（准确性和完整性），保证数据的安全性（保密性、完整性及可用性），实现数据资源在各组织机构部门的共享；推进信息资源的整合、对接和共享，从而提升集团公司或政务单位信息化水平，充分发挥信息化作用。

# 数据治理体系

数据治理体系包含两个方面，一是数据质量核心领域，二是数据质量保障机制。

具体两者内容及相互关系可以参见下图：



# 数据治理核心领域

为了有效管理信息资源，必须构集团级数据治理体系。数据治理体系包含数据治理组织、数据构架管理、主数据管理、数据质量管理、数据服务管理及数据安全管理内容，这些内容既有机结合，又相互支撑。

## 数据模型

数据模型是数据构架中重要一部分，包括概念数据模型和逻辑数据模型，是数据治理的关键、重点。理想的数据模型应该具有非冗余、稳定、一致、易用等特征。逻辑数据模型能涵盖整个集团的业务范围，以一种清晰的表达方式记录跟踪集团单位的重要数据元素及其变动，并利用它们之间各种可能的限制条件和关系来表达重要的业务规则。数据模型必须在设计过程中保持统一的业务定义。为了满足将来不同的应用分析需要，逻辑数据模型的设计应该能够支持最小粒度的详细数据的存储，以支持各种可能的分析查询。同时保障逻辑数据模型能够最大程度上减少冗余，并保障结构具有足够的灵活性和扩展性

## 数据生命周期

一般包括数据生成及传输、数据存储、数据处理及应用、数据销毁四个方面。

（1）数据生成及传输

数据应该能够按照数据质量标准和发展需要产生，应采取措施保证数据的准确性和完整性，业务系统上线前应该进行必要的安全测试，以保证上述措施的有效性。对于手工流程中产生的数据在相关制度中明确要求，并通过事中复核、事后检查等手段保证其准确性和完整性。数据传输过程中需要考虑保密性和完整性的问题，对不同种类的数据分别采取不同的措施防止数据泄漏或数据被篡改。

（2）数据存储

这个阶段除了关注保密性、完整性之外，更要关心数据的可用性，对于大部分数据应采取分级存储的方式，不仅存储在本地磁盘上，还应该在磁带上，甚至远程复制到磁盘阵列中，或者采用光盘库进行存储。对于存储备份的数据要定期进行测试，确保其可访问其数据完整。数据的备份恢复策略应该由数据的责任部门或责任人负责制定，信息化管理部门可以给予相应的支持。同时还需要注意因为部门需要或故障处理的需要，可能对数据进行修改，必须在数据管理办法中明确数据修改的申请审批流程，审慎对待后台数据修改。

（3）数据处理和应用

信息化相关部门需要对数据进行分析处理，以挖掘出对于管理及业务开展有价值的信息，为保证过程中数据的安全性，一般应采用联机处理，系统只输出分析处理的结果。但是实际中，因为相关数据分析系统建设不到位，需要从数据库中提取数据后再对数据进行必要的分析处理，在这个过程中就需要关注数据提取操作是否可能对数据库造成破坏、提取出的数据在交付给分析处理人员的过程中其安全性是否会降低、数据分析处理的环境安全性等等。

（4）数据销毁

这个阶段主要涉及数据的保密性。应明确数据销毁的流程，采用必要的工具，数据的销毁应该有完整的记录。尤其是对于需要送出外部修理的存储设备，送修之前应该对数据进行可靠的销毁。

## 数据标准

数据标准是集团单位建立的一套符合自身实际，涵盖定义、操作、应用多层次数据的标准化体系。

数据标准的建立是集团单位信息化、数字化建设的一项重要工作，行业的各类数据必须遵循一个统一的标准进行组织，才能构成一个可流通、可共享的信息平台。

数据治理对标准的需求可以划分为两类，即基础性标准和应用性标准。前者主要用于在不同系统间，形成信息的一致理解和统一的坐标参照系统，是信息汇集、交换以及应用的基础，包括数据分类与编码、数据字典、数字地图标准；后者是为平台功能发挥所涉及的各个环节，提供一定的标准规范，以保证信息的高效汇集和交换，包括元数据标准、数据交换技术规范、数据传输协议、数据质量标准等。

（1）数据分类与编码

数据分类与编码标准是信息化建设中标准化的一项基础工作，该类标准规定平台汇集、交换相关信息统一的分类系统和排列顺序以及编码规则，目的是在不同系统和用户之间建立交通数据的一致参照，对提高数据采集、处理和数据交换效率具有重要作用。数据分类与编码标准的制定将有力推进平台标准化及交通信息化建设标准化的进程。

（2）数据字典

针对实际需求，定义数据集，建立各个领域的数据字典，规范数据概念和数据定义。在此基础上，形成完备的集团单位数据集和数据字典。

（3）元数据标准

元数据标准是描述数据资源的具体对象时所有规则的集合，它包括了完整描述一个具体数据对象时所需要的数据项集合。针对各种信息资源分别制定适当的元数据标准，可为信息的管理、发现和获取提供一种实际而简便的方法，从而提高数据交换效率。

（4）数据交换标准

为了保证数据共享和交换的顺利实现，必须明确定义和规范数据交换的相关标准。数据交换的标准规范是集团单位综合信息平台的核心标准。其中应当包括数据交换内容、数据交换格式、数据传输方式、各类中心间数据接口的标准化等方面。

（5）数据质量标准

由于数据采集任务通常由其他二级平台完成，数据治理平台的标准方法主要集中在数据的加工和管理上。应该重点开发的一个领域是数据质量控制方法。应当从三个方面对数据质量方法进行研究：“ 坏数据”或“不可靠数据”的识别，错误数据的编辑方法，以及缺少值的处理。

## 主数据

主数据管理要做的就是从各部门的多个业务系统中整合最核心的、最需要共享的数据（主数据），集中进行数据的清洗和丰富，并且以服务的方式把统一的、完整的、准确的、具有权威性的主数据传送给集团单位范围内需要使用这些数据的操作型应用系统和分析型应用系统。

主数据管理的信息流应为：

1) 某个业务系统触发对主数据的改动；

2) 主数据管理系统将整合之后完整、准确的主数据传送给所有有关的应用系统

3) 主数据管理系统为决策支持和数据仓库系统提供准确的数据源。

因此对于主数据管理要考虑运用主数据管理系统实现，主数据管理系统的建设，要从建设初期就考虑整体的平台框架和技术实现。

## 数据质量

数据质量不高将影响数据仓库应用程度不高。低下的数据质量往往造成开发出来的系统与用户的预期大相径庭，数据质量关系建设有关分析型信息系统成败，同时数据资源是集团单位的战略资源，合理有效的使用正确的数据能指导集团单位做出正确的决策，提高省综合竞争力。不合理的使用不正确的数据（即差的数据质量）可导致决策的失败，正可谓差之毫厘、谬以千里。

数据质量管理包含对数据的绝对质量管理、过程质量管理。绝对质量即数据的真实性、完备性、自治性是数据本身应具有的属性。过程质量即使用质量、存储质量和传输质量，数据的使用质量是指数据被正确的使用。再正确的数据，如果被错误的使用，就不可能得出正确的结论。数据的存贮质量指数据被安全的存贮在适当的介质上。所谓存贮在适当的介质上是指当需要数据的时候能及时方便的取出。数据的传输质量是指数据在传输过程中的效率和正确性。

高质量的交通运输行业数据至少有如下几项要求：

一是正确性，在转换、分析、存储、传输、应用流程中不存在错误；

二是完整性，数据库应用或要求的所有记录、字段都存在；

三是一致性，体现在整个数据库的定义和维护方面，确保数据在使用的整个过程中是一致的；

四是时效性，衡量指标是在指定的数据与真实的业务情况同步的时间容忍度内，即指定的更新频度内，及时被刷新的数据的百分比；

五是可靠性，提供数据的数据源必须能够可靠稳定地提供数据。

数据质量管理的规划和实施包括以下内容：

一是数据质量管控体系的建立，包括数据质量的评估体系，定期评估数据质量状况；

二是在部门各个应用系统中的落实，包括每个应用系统中的数据质量检查等；

三是在最开始建立数据质量管理系统的时候，借助数据治理平台上，通过建立数据质量管理的规则来集中化地建立数据质量管理系统，发现问题并持续改进；

四是数据质量管理与业务稽核的结合，通过业务规则的稽核来发现数据质量深层次的问题，将数据质量与业务一线结合起来，使业务人员对数据质量问题有更加清晰和明确的认识。

完善的数据质量管理是保障各项数据治理工作能够得到有效落实，达到数据准确、完整的目标，并能够提供有效的增值服务的重要基础。

## 数据服务

数据整理最终目的就是要服务于各部门单位、人员等，能更准确更快更方便的服务是数据服务管理的目标。

数据服务管理是指针对内部积累多年的数据，研究如何能够充分利用这些数据，分析行业业务流程优化业务流程。数据使用的方式通常包括对数据的深度加工和分析，包括通过各种报表、工具来分析运营层面的问题，还包括通过数据挖掘等工具对数据进行深度加工，从而更好的管理者服务。通过建立统一的数据服务平台来满足针对跨部门、跨系统的数据应用。通过统一的数据服务平台来统一数据源，变多源为单源，加快数据流转速度，提升数据服务的效率。

## 数据安全

由于集团单位的重要且敏感信息大部分集中在应用系统中，数据安全更是至关重要。如何保障数据不被泄露和非法访问，是非常关键的问题。数据安全管理主要解决的就是数据在保存、使用和交换过程中的安全问题。

数据安全管理主要体现在以下六个方面：

一是数据使用的安全性，包括基础数据的保存、访问和权限管理；

二是数据隐私问题，系统中采集的证件号码、银行账号等信息在下游分析系统和内部管理系统中，是否要进行加密，以避免数据被非法访问；

三是访问权限统一管理，包括单点登录问题及用户名、数据和应用的访问授权统一管理；

四是数据安全审计，为数据修改、使用等环节设置审计方法，事后进行审计和责任追究；

五是制度及流程建立，逐步建立数据安全性的管理办法、系统开发规范、数据隐私管理办法及相应的应用系统规范、在管理决策和分析类系统中的审计管理办法等；

六是应用系统权限的访问控制，建立集团级权限管理系统，增加数字水印等技术在应用系统中的使用**。**

# 数据治理保障机制

## 制度章程

### 规章制度

数据治理章程类似于企业的公司条例。该章程阐明数据治理的主要目标、相关工作人员、职责、决策权利和度量标准。

具体可参见 附件 数据管理规范

### 管控办法

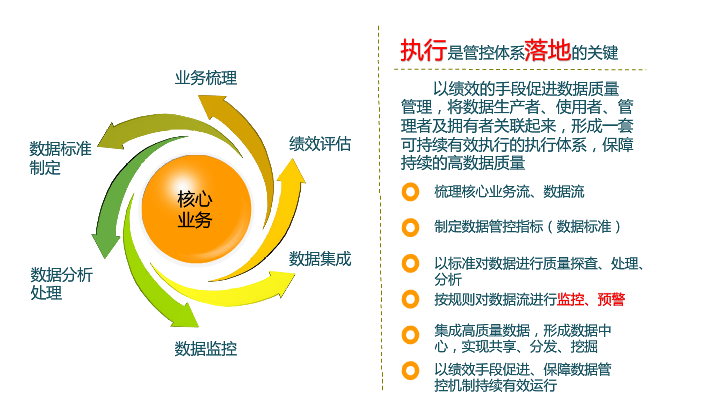
管控办法是基于规章制度与工具的结合，可落地的操作的办法。

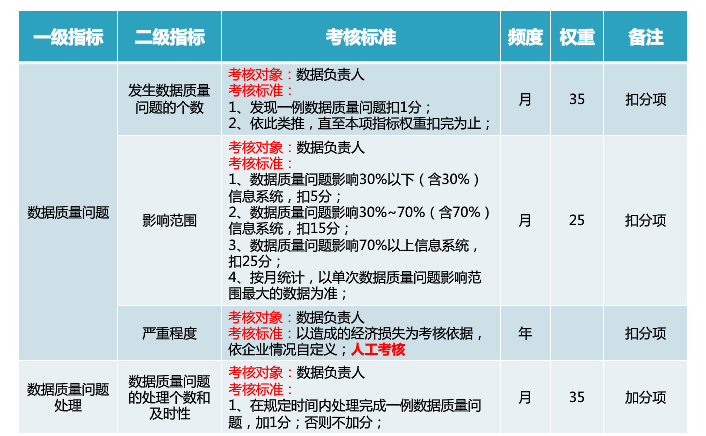
具体可参见 附件 数据质量评估办法

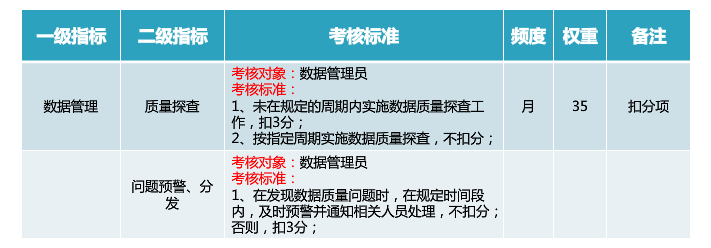
### 考核机制

考核是是保障制度落实的根本，建立明确的考核制度，实际操作中可根据集团单位情况，建立相应的针对数据治理方面的考核办法，并与个人绩效相关联。可参考管理学中相关考核、绩效管理相关部分。

对于数据治理的考核，可见下图进行理解：







## 数据治理组织

### 组织架构

有效的组织机构是项目成功的有力保证，为了达到项目预期目标，在项目开始之前对于组织机构及其责任分工做出规划是非常必要的，数据治理项目管理组织建议宜采用如图所示的组织结构：



### 组织层次

数据治理委员会由集团公司的高层领导者组成。委员会定义数据治理愿景和目标；组织内跨业务部门和 IT部门进行协调；设置数据治理计划的总体方向；在发生策略分歧时进行协调。此委员会也将包含来自部门或子公司的领导代表，以及来自各单位视数据为机构资产的信息科技部门的代表。这些高层管理人员是数据治理计划的所有拥护者，确保在整个组织内获得支持。

数据治理工作组是组织内委员会下面的下一个级别。工作组执行数据治理计划。工作组负责监督数据管理员工作。数据治理工作组由数据治理委员会中各局领导主持。

每各业务部门有至少一位业务分析员，信息科技部门设置数据质量分析员、数据管理员、集成开发人员。各工作人员负责本部门数据的质量，履行职责，解决具体的问题。

### 组织职责

根据数据管理工作的实际需要，在业务管理部门、技术管理部门和业务应用部门确定各工作人员的职责。

**数据治理委员会的职责范围：**

1）从战略角度来统筹和规划，对数据资产和系统进行清理，确定数据治理的范围；明确数据源的出处、使用和管理的流程及职责；

2）明确数据治理的组织、功能、角色和职责；

3）负责各工作组成员的培训工作；

4）负责审查各工作小组的目标、原则，批准数据管理的相关制度、标准及流程；

5）负责确定数据治理的工具、技术和平台；

6）负责制定数据治理的评估指标、方法。

**数据治理工作小组，其主要工作职责是：**

负责数据治理的牵头，组织、指导和协调本单位的数据治理工作；

综合数据治理管控办法、数据治理考核机制等有关规章制度的牵头制定、修改等；

负责数据的分析整理并出具数据指标报告；

负责数据的监测预测工作；

建立数据冲突的处理流程和数据变更控制流程。

负责对基础数据质量的检测、发布、考核和清理完善工作。

**工作组成员：业务分析员、数据质量分析员、数据管理员、集成开发员**

这些不同的角色在数据治理过程中承担着彼此不同，而又相辅相成的职责。其中集成开发人员在数据治理流程中需要肩负起数据访问、验证数据结构、验证数据、交付数据以及数据库/知识库的构建等角色，因此他们的工作包括：

* 访问及交付相应数据给业务用户
* 提高生产力和性能
* 最大化减少异常/出错的影响
* 开发和完善技术最佳实践

数据质量分析员在数据治理流程中负责数据的剖析、清洗匹配合并等。工作包括：

* 为开发人员定义数据规格及标准
* 为机构有效的追踪数据质量问题
* 实施被业务人员和数据管理员定义正确的数据质量规则
* 不间断的监控数据质量水平及问题

业务分析人员在数据治理流程中负责定义数据的转换规则，工作包括：

* 与需求开发人员协作，正确捕获和解析业务需求
* 与开发人员和数据管理员协作，为业务用户缩短数据产生价值的时间

数据管理员需要定义引证数据，并管理元数据，工作包括：

* 保证数据的质量、正确、完整、一致、审计及安全性
* 定义引证/参考 数据
* 为组织机构数据实体给出正确业务定义
* 为组织机构解决混淆和有争论的数据定义

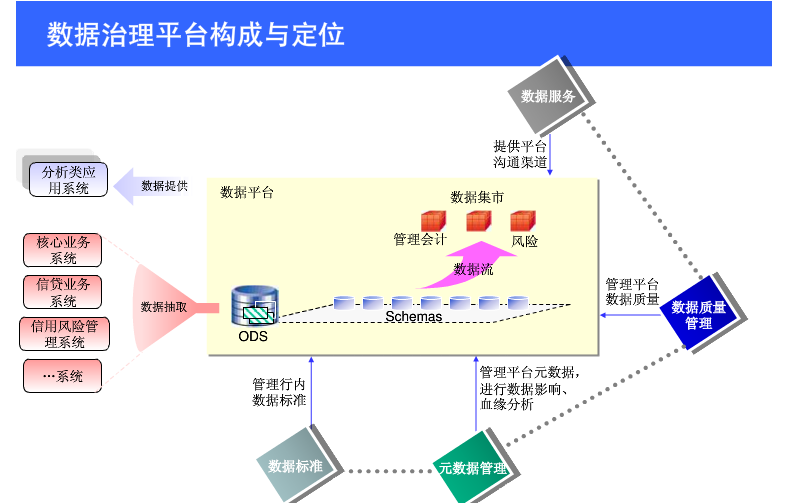
## 流程管理

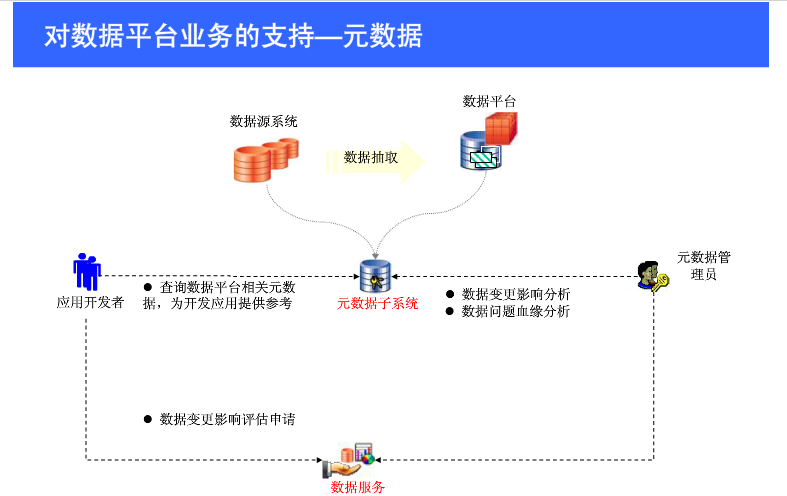
流程管理包括流程目标、流程任务、流程分级，根据数据治理的内容，建立相应流程，且遵循本单位数据治理的规则制度。实际操作中可结合所使用的数据治理工具，与数据治理工具供应商进行协商，建立符合集团公司的流程管理。

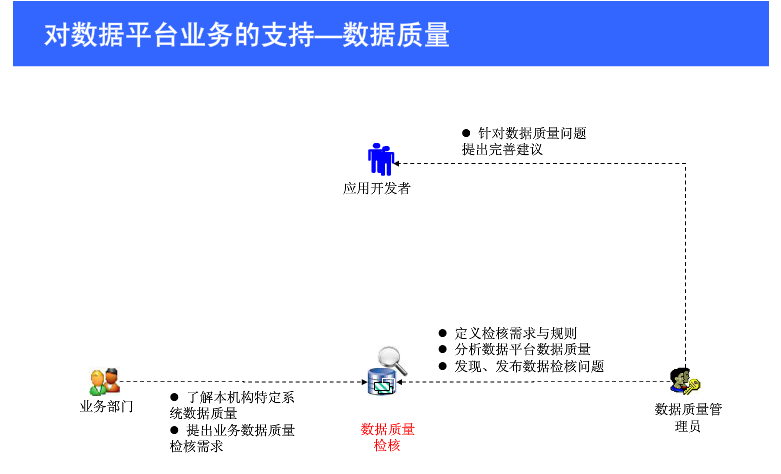
## IT技术应用

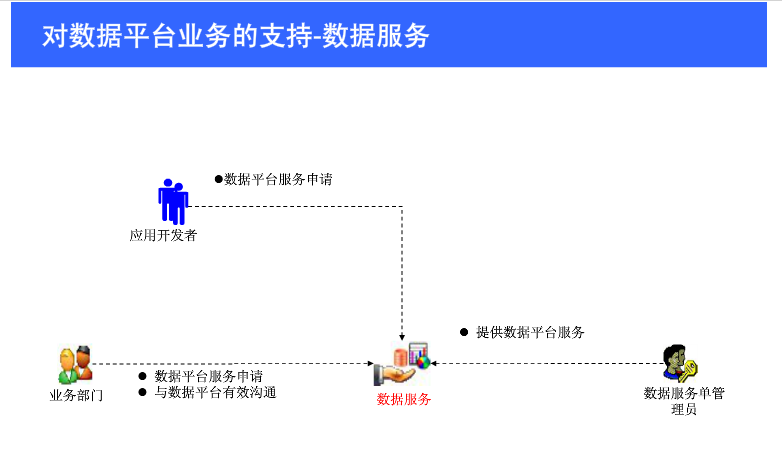
### 支撑平台

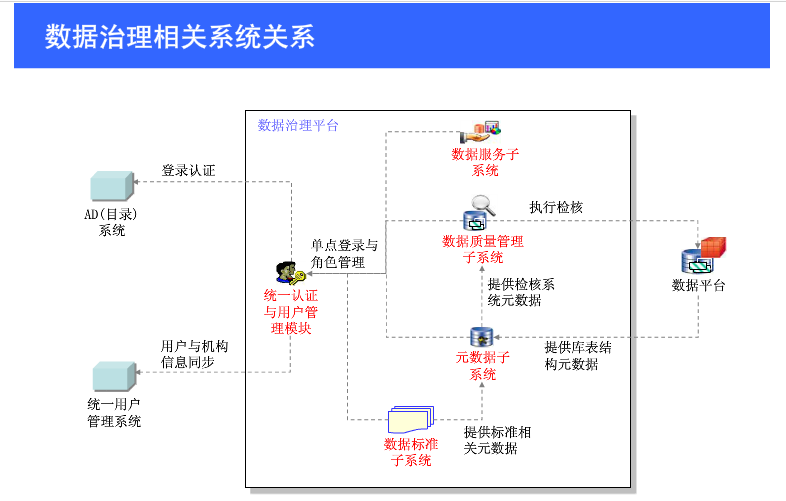
现市场上关于数据治理平台存在不同的成熟产品，但在功能实现上大致相同，具体可参加下图了解。











### 技术规范

技术规范是保障数据治理平台可持续管理的基础，随着数据量的增长、技术水平的发展，为更好、可持续的实现数据资产的管理、应用，需建立明确的技术规范。

# 附件A 数据管理规范

为逐步提高数据资源利用效果，推动信息化建设管理向标准化、信息化和数字化方向发展，数据管理应遵循以下原则：

（1）统一规范。数据标准要严格执行省交通运输厅的统一标准。

（2）分级管理。实行分层级的数据管理模式，明确职责分工，层层落实责任。

（3）过程控制。建立数据从采集、报送、审核到应用、维护全过程的控制规范，保证数据质量，提高应用效果。

（4）保障安全。建立数据访问的身份验证、权限管理及定期备份等安全制度，规范操作，做好病毒预防、入侵检测和数据保密工作。

（5）数据共享。整合应用系统，做到入口唯一，实现数据一次采集，集中存储，共享使用。

数据管理的规范工作包括：数据标准、数据采集、数据审核、数据维护、数据分析、数据应用、数据发布、数据传输、数据存储（备份、恢复）、数据安全管理、数据质量监控、数据管理考核等。

**（一）数据标准**

第一条 安徽省交通运输行业数据必须按国家交通部制定的数据标准和代码规范执行。国家交通部标准中未包括的内容执行安徽省交通运输厅的统一规范。

第二条 在数据采集工作中，各级应严格依据省局制定的标准进行数据的采集录入、审核修正、数据质量监控和管理考核，确保数据质量。

第三条 数据指标的设定和修改必须严格执行软件的管理规定和程序。凡使用省局统一软件的，市局对有关数据指标无权制定和修改，必须执行全省统一规范。使用非省局统一软件的，市局业务管理部门应根据工作需要和所管软件的需求变化，及时提出设定（修改）数据指标的具体内容、质量标准和时间要求，并明确数据录入部门、录入岗位和录入时限，报分管领导批准后提交安徽省交通运输联网运行管理中心。安徽省交通运输联网运行管理中心应根据业务管理部门的需求，及时制定（修改）数据采集的技术规范，包括数据采集格式、权限设定和方法步骤等，并按时向业务管理部门通报所采集数据的指标内容变化情况，以便业务部门根据采集到的指标提出数据加工需求。

第四条 数据指标的确定应以有效实用和优化服务为目标，不断提高数据集中度和信息共享度，科学归并各项业务的同类、同属性指标，避免业务部门重复上报数据。

**（二） 数据采集**

第一条 数据采集重点是原始数据的采集，原始数据主要包括业务部门通过纸质资料报送的数据，通过电子传输报送的数据（含电磁介质报送的数据和网络传输报送的数据，下同）；相关部门通过纸质、电子传输提供的数据；管理业务处理结果（检查、调查、核实、认定、审批等）产生的数据。

第二条 数据采集主要依靠应用系统进行录入（导入），现用应用系统不能满足数据采集需要的，通过相关系统的补录模块进行数据录入（导入）。

第三条 数据采集方式分为手工采集、数据电文导入和光学字符识别（OCR）、图形扫描等其他方式。

为保障数据采集的质量，提高数据采集的效率，应充分应用信息化技术，最大限度地减少手工录入。

第四条 为提高数据录入质量，各市交通部门有关单位要认真执行资料数据的核对制度，严格以原始资料为依据，确保录入数据的一致性和准确性。数据采集录入前，操作人员应先将准备录入的原始资料或有关表证单书进行完整性、逻辑性、真实性审核，发现漏填、错填和逻辑关系不符的，应于做出修正或补正；

第五条 对通过电子传输方式报送数据的，数据必须通过接收软件相应设置的完整性、逻辑性审核。

第六条 各级各单位应建立纸质数据和电子数据的核对制度，明确职责分工和工作程序，落实岗位责任，发现问题及时处理。

第七条 对于交通部、省厅没有规定的数据采集标准规范，各局市交通部门及信息化责任部门应根据不同的数据采集岗位和软件覆盖面，细化工作职责，量化作业标准，结合业务操作规程，建立统一的数据采集录入操作规范和管理制度。数据采集规范应包括数据来源、采集时间、操作岗位、操作步骤、操作内容和采集内容的逻辑校验等要素。

第八条 各级数据采集人员应强化数据质量意识, 熟练掌握正确的采集和审核方法，按照部门职责分工，严格按照相关操作规范采集各类数据。

第九条 任何人不准擅自委托他人以本人用户名录入数据。确因工作需要委托他人以自己名义录入的，必须报经主管领导批准，并及时更改口令。

第十条 安徽省交通运输联网运行管理中心应简化、归并各级地单位报送的报表，统一采集，集中处理，自动生成，分别使用，提高工作效率和质量，减轻各级单位的工作量。

**（三）数据审核**

第一条 数据审核是确保数据质量的重要环节。审核的重点是数据的及时性、完整性、真实性、准确性和规范性。

（1）及时性是指数据要在规定的时间内采集，确保应用系统数据与实际交通信息业务同步；

（2）完整性是指数据采集应按照有关规定及相应表、证、单、书采集的要求进行，不得缺表或漏项，杜绝数据的机外操作和循环；

（3）真实性是指数据采集要如实反映交通行业管理实际，以合法、真实的原始资料为依据，不得随意捏造;

（4）准确性是指数据采集应准确反映交通行业管理实际，与纸质资料数据一致，数据之间逻辑相符，不得出现运算错误或逻辑错误；

（5）规范性是指数据采集应按照统一的业务标准及技术规范进行。

第二条 各市交通部门及相关岗位人员按照自身工作职责范围，负责相关数据的日常审核。

第三条 为加强数据质量管理，省厅建立数据审核制度，数据管理员定期或不定期对已采集数据进行审核。对审核中发现的数据质量问题应及时反馈至具体的原始数据采集岗，限期补录修正。

第四条 各级可采取人工抽样、数据软件检测等方式进行数据审核。对检索出不符合标准的数据，反馈至原数据采集岗进行补录修正。在数据应用环节发现的数据采集差错，应通过相关业务管理部门的数据管理员进行审核确认，然后反馈至原数据采集岗进行变更修正。

第五条 审核中发现的差错应及时通知相关部门。

第六条 数据审核岗人员与数据采集岗人员对数据问题有异议的，由上级业务主管部门负责对争议事项做出裁定。

第七条　对各级数据审核发现的问题数据基层分局已无权限修正的，严格按照数据维护规定的程序开展修正工作，任何人不得擅自修改。

**（四） 数据维护**

第一条 数据维护包括对数据中错误数据的修正、不完整数据的补充、垃圾数据的清理及历史数据的迁移等。

第二条 各市交通业务管理部门应结合各应用系统制定详细的数据维护工作制度，明确数据维护的权限、职责，严格按照工作制度进行数据维护。已经进入应用系统的数据，不得擅自修改、删除。

第三条 对错误数据的修正和不完整数据的补充，省厅统一按以下程序处理：

（1）对尚未录入应用系统或已录入尚保留修改权限的数据，需修改的，可由前台原数据采集岗根据审核意见纠正后重新录入（或直接纠正）；

（2）对已录入应用系统的数据，经审核发现错误，前台原数据采集岗已无权限修改的，应根据具体业务工作流程和要求实行层级审批修正。

第四条  对已录入应用系统需层级审批修正的数据，市级信息部门有权限处理的，应由数据修正申请单位及时提出数据修正申请，报数据所属系统的责任业务管理科室审核审批。市信息管理部门同意修正或在接到同意修正的审批后办理。

对经确认不能予以修正的，有权修正单位应及时向申请单位或相关人员说明原因。

第五条 对需由省数据中心办理的数据修正事项，应由申请单位及时提出数据修正申请，分别报经区市信息管理部门及责任业务管理科室审批。省级数据中心在接到同意修正的申批后办理。

对经确认不能予以修正的，有关单位应及时向数据修正申请单位及有关人员说明原因。

第六条 对不按规定程序上报审批的数据修正申请，各级数据管理的技术管理部门不予以修改。

第七条　对经查实，属有关单位弄虚作假、人为调整指标等主观因素造成其无法自行恢复的数据错误，各级数据管理的技术管理部门有权不予修正，并向省交通运输单位报告。

第八条 各级数据管理的技术管理部门应按照系统初始维护设置和上级部门分配的权限实施数据修正业务。执行过程中如遇权限调整，各级数据管理的技术管理部门应在权限变动后十五日内向辖区内本级和下一级业务部门公告通知。

第九条 数据维护前应做好相应备份工作。数据维护工作应严格备案，对每项数据维护的内容、时间、维护原因、责任人等进行详细记录，涉及的书面材料必须登记存档。

第十条 应用系统因数据平台转换、系统升级等原因需对历史数据转储、迁移的，由信息技术部门会同业务管理部门确定历史数据的处理方法，保证新旧数据的衔接和系统的平稳过渡。

第十一条 各级信息技术部门应加强对数据的监控，定期检测数据的存储，分析数据的构成，提出数据清理优化的方案，经业务管理部门确认后，定期进行垃圾数据的清理。垃圾数据主要指过期的、重复的、没有使用价值的数据，不包括按规定应保留的历史数据。

**（五）数据分析**

第一条 数据分析是采取科学合理的方法，利用现代信息技术手段，对计算机应用系统生成的数据进行分析，充分发掘数据中蕴涵的信息，形成有用指标数据值的过程。

第二条 数据分析一般按照选题—分析—应用—反馈—再选题的方式进行，积极研究、探索科学实用的分析方法，为交通运输工作服务。同时依据数据应用过程中暴露出的问题，改进数据管理方法和手段，促进数据管理质量不断提高。

第三条 数据分析分为固定分析和专项分析两种。固定分析指标在一定时期内相对固定，通过编制应用软件自动生成。专项分析的选题指标根据交通运输管理工作的新形势和新问题需要专门设计，临时从前台或后台抽取相关数据进行分析处理。

第四条 各级各单位应充分利用数据资源，规范数据分析程序，积极探索数据获取、加工、计算、展现等过程的规律，不断探索和创新数据分析方法，提高数据分析质量。

**（六） 数据应用**

第一条 数据应用分为直接使用和加工后使用。数据加工是指根据工作需要，用数理统计、多维分析等科学方法对数据进行抽取、汇集、归类、挖掘、比对，并以报表、图形、文字等形式展现数据处理结果。

第二条 各级各单位应充分利用数据资源，深化数据应用，用数据描述现状，预测趋势，规范执法，加强管理，优化服务，提供决策支持。数据应用的任务一般包括：

（1）整合应用系统的数据，建立数据关联关系，满足各级、各部门、各岗位业务管理的需要。

（2）分析行业管理现状，研究存在问题，提出改进措施，提高管理水平，规避管理风险。

（3）监控执法过程，跟踪执法结果，检查执法质量。

（4）建立合理的指标体系，全面考核和监控交通各部门管理的工作质量和效率。

（5）统筹规划、合理分布数据，实现各类数据的全面共享，简化交通运输行业审批程序，优化交通服务。

第三条 省数据加工统一由各级数据管理中心承担，信息部门提供技术支持。业务管理部门有特殊需求时，报分管局长批准后及时向同级数据管理中心提出，各级数据管理中心应及时提供已加工数据列表和数据使用授权，以保证各级、各部门业务管理需要。

第四条 数据加工分两个层次，一个层次是为满足日常业务需要，利用数据加工工具进行数据加工，供各部门日常管理使用，可直接展示数据处理结果；另一层次是满足评估分析和辅助决策等数据应用系统的需要，利用数据加工工具或专用软件对数据加工或进行系统间的数据衔接。

第五条 全省各级数据中心负责数据加工软件的日常管理工作。

第六条 数据使用单位和人员必须严格按照授权使用数据，负责管理本单位、本人口令，不得越权使用数据；不得采取任何方法破坏数据；对所使用的涉密数据负有保密责任。

**（七） 数据发布**

第一条 数据发布是数据管理的重要内容。各级各单位必须建立数据发布制度，明确数据发布职责，依法发布数据：

（1）可向交通运输从业人员本人提供的信息主要包括：除稽查在查案件以及其他涉密信息之外的交通运输从业人员本人的信息；

（2）可向社会公众发布的信息主要包括：非正常户、违法违章信息, 停业、企业质量信誉等级、政务公开信息等信息,驾驶证、从业资格证、经营许可证信息及真伪信息等；

（3）可在交通管理部门之间传递、共享的信息主要包括：企业登记信息、核定认定信息、申报信息、审批信息、行政许可信息、委托机构信息、违法违章信息、稽查信息、救济信息等。

（4）其他单位或个人要求交通部门提供相关信息的，符合法律法规规定的，交通部门应当提供。

第二条 各级各单位对外公布数据必须报领导或分管领导批准后对外公布，否则不得擅自对外公布和分布数据。

**（八） 数据传输**

第一条 各级各单位应针对各项数据传输工作，设置专门岗位，明确职责分工，制定相应考核制度。

第二条 数据传输工作严格遵照相应的操作规程和时间要求，不得延误。由于特殊原因，数据发送方不能按时完成数据传输任务时，应及时通知数据接收方，双方应积极采取措施，恢复正常传输。数据传输完成，双方应及时进行数据对账。

第三条 数据传输应当使用交通运输机构内部计算机网络完成，未经批准不得借助其它公共计算机网络平台进行数据传输。使用可移动载体进行数据传输的，传输完毕后，必须从载体上完全清除数据。

第四条 全省各级交通部门内部信息的传递，如任务下达、信息交换、调查统计、报表报送、成果报告、情况通报等，应采取电子数据传输方式。

能够通过应用系统查询得到的信息或生成的报表，上级交通部门不得要求下级机关报送。

第五条 各级信息技术部门要定期检测网络运行状况，及时发现解决网络传输中的问题，确保数据传输网络畅通。

**（九）数据存储、备份和恢复**

第一条 各级信息技术部门应加强对各类数据存储和备份的管理，以保障应用系统的正常运行，保存完整的历史数据。

第二条 各级信息技术部门应定期对存储和备份的数据进行整理优化，以提高系统运行和数据处理的效率。

第三条 各类数据由信息技术部门统一集中存储和备份。

第四条 各级信息技术部门应根据不同类型数据的更新频率、数据量、重要程度、保存期限，制定相应备份、恢复策略和操作规范。

第五条 已做备份的数据任何部门不得擅自更改。

第六条 数据备份文件必须存储在非本机磁盘的其它介质中，建立登记制度，由专人保管，备份介质必须保存在符合条件的环境中，对应用系统中存储时间长、使用频度低的历史数据，可按有关规定予以结转、存储，长期妥善保管，并随时可供调取应用。重要数据应异地存放。

第七条 数据备份文件应定期进行恢复测试，以确保所备份的数据能够及时、准确、完整地恢复。

**（十） 数据安全管理**

第一条 省各级信息管理中心应建立健全数据安全管理制度和办法。数据安全管理的内容包括：数据访问的身份验证、权限管理及数据的加密、保密、日志管理、网络安全等。

第二条  为统一规范操作权限，省交通行业各级各单位操作人员的录入权限、访问权限、维护权限应先由各业务管理部门按照部门、岗位的职责分工，提出权限设定规则，报数据管理领导小组批准后由信息技术部门予以维护，任何人不得擅自设立、变更和注销。

第三条 各类应用系统的使用必须实行用户身份验证。应用系统应按规定设置相应的用户名、密码，并按不同权限级别，对用户口令加密保护。操作人员应注意自己用户名和口令的保密，并定期或不定期修改口令。

第四条  数据安全实行专人负责制。信息技术部门要指定专人负责系统数据及磁介质资料的安全管理工作。对数据库的管理实行数据库管理员制度，制定和明确管理员用户和数据查询用户的操作权限及规程，关键数据库管理岗位应设两人或两人以上。

第五条 对涉密数据的传输、存储，应严格按照相关规定进行加密处理。

第六条 对各类数据应严格执行保密制度，不得泄漏。对涉及从业人员隐私而未征得本人同意或按有关保密规定不能公开的数据，不得对外公开，也不得用于业务以外的其它目的。

第七条 对数据的各项操作应实行日志管理，严格监控操作过程，对发现的数据安全问题，要及时处理和上报。

第八条 各级各单位要加强网络安全管理，采取严格措施,做好计算机病毒的预防、检测、清除工作，防止各类针对网络的攻击，保证数据传输和存储安全。管理员用户应掌握和运用数据库访问跟踪布控技术，加强对数据库的核查与监控。

第九条 对需要长期保存的数据磁带、磁盘，应在质量保证期内（一般为一年）进行转储，以防止数据失效造成损失。

第十条 系统维护和数据管理责任单位应当采取以下措施，保证信息系统应用数据的安全：

（1）明确岗位职责，严格操作规范，完善内部控制；

（2）加强应用系统的运行维护，定期进行数据备份；

（3）实时监控数据库的访问情况，定期审核和更新数据库的口令等

**（十一）数据质量监控**

第一条 数据质量监控是全省各级数据管理中心以交通行业业务规程为基础，通过建立数据质量指标体系和设置过错类别，运用一定的手段和方法，对一定时期的数据质量进行检查、认证、鉴定与评价等一系列管理活动。

第二条 数据质量监控应按照分途采集、集中比对、管住增量、强化责任的原则，对数据采集和处理的真实性、准确性、完整性、合法性、逻辑性、及时性进行监控。

第三条 全省交通系统建立省、市、县三级数据质量监控联动机制，实行省局数据管理中心、市级数据管理中心、基层分局数据管理员三级监控联动。

（1）省局数据管理中心负责全省范围内的数据质量监控，定期或不定期发布全省数据质量监控检测结果。

（2）市级数据管理中心负责全市、区范围内的数据质量监控，定期或不定期发布全市、区数据质量监控检测结果，指导基层分局落实县级数据质量监控检测结果。

（3）基层分局数据管理员负责督导落实辖区范围内的数据质量监控检测结果。

第四条 各级数据管理中心应建立相应的数据质量监控机制，采用人机结合等方式，以一定形式对本地区数据质量实施监控审核，并及时发布监控结果。

第五条 各级数据管理中心应于年度开始10日内制定年度数据质量监控检测计划，经分管局长批准后有计划开展全年数据质量监控检测工作，并报上级数据管理中心备案。

第六条　 各级数据管理中心负责对数据质量监控发现的问题数据限期督导基层分局完成修正，一般最长要求5日内完成，并将修正结果上报发现问题的数据管理中心审核。

第七条 对数据质量监控发现的问题数据严格按照本办法数据维护规定的程序开展修正工作，任何人不得擅自修改。

**（十二） 数据管理考核**

第一条 数据管理是交通行业信息化建设的重要内容，数据质量直接关系到交通行业信息化的应用效果。各级各单位应根据本办法所涉及的内容制定具体的数据管理考核办法，建立切实有效的数据管理考核奖惩机制。

第二条 数据管理考核应遵循客观公正、公开透明的原则。各级可采取日常考核和定期考核相结合、人工抽检考核和机器全面检测相结合的方法，强化数据管理考核。省厅将通过数据检测软件等定期或不定期考核通报各市区（开发区）交通局数据采集质量情况，并对各县市区（开发区）交通局数据管理工作实施百分制考核，主要考核项目包括：

（1）组织机构是否健全，岗位职责是否明确；

（2）数据录入是否及时，数据采集是否完整；

（3）数据内容是否准确，抽取展示是否按时；

（4）审核差错是否修正，数据维护是否及时；

（5）管理机制是否建立，安全措施是否到位。

第三条 市交通局数据管理中心根据上级要求，针对各应用系统和业务管理的实际，制定全市数据管理考核办法并于年度结束20日内完成对各县市区（开发区）交通局的数据管理考核工作。

第四条 数据管理考核办法应突出数据录入、审核、维护、备份、安全等重点环节，进行指标量化。依据标准，对必录项的考核指标，在采集录入环节，其及时率、准确率、完整率必须达到95％以上；经过审核、修正后，其“三率”的考核指标应达到99％以上。数据管理考核结果纳入全市目标管理考核内容。

第五条  交通部门和部门人员发生下述行为之一的，上级机关可根据其危害程度、过错大小、情节轻重等，依据执法责任制和目标管理责任制追究相应责任：

（1）数据录入不及时，造成下一环节工作延误的；

（2）数据录入不规范、差错率连续居高的；

（3）未认真履行数据审核职责，造成大量数据差错不能及时纠正的；

（4）未及时进行数据维护，造成应用环节数据严重失真的；

（5）未按规定存储和备份数据，造成数据丢失的；

（6）未按规定进行系统维护，或因过失引发信息系统运行障碍造成数据丢失的；

（7）未按规定实施安全管理防范措施，造成计算机系统瘫痪或病毒侵害严重，严重影响数据传输和存储的；

（8）未按规定执行保密制度，发生重大泄密事件的；

（9）擅自对数据、权限进行修改、删除，或擅自使用他人的用户名、密码进行操作访问的；

（10）其他过错行为。

对有意破坏、恶意攻击税收应用信息系统，造成重大损失，涉嫌刑事责任的，移送司法机关处理。

# 附件B 数据质量评估办法

下面将从数据质量评估核心指标、数据质量评估模式、数据质量评估管理流程三个方面介绍数据质量评估方法。

数据质量评估的核心指标

|  |
| --- |
| **数据质量问题频率**  指标定义：数据质量问题频率＝数据质量问题发生次数/存储的总数据量  指标单位：次/GB |

根据数据质量评估指标将各数据仓库中的主数据及其历史行为划分为三个等级：

|  |  |  |
| --- | --- | --- |
| **数据质量等级** | **描述** | **统计口径** |
| 一级 | 数据质量差，需要重点监控 | 数据质量问题频率大于等于1次/GB |
| 二级 | 数据质量一般 | 数据质量问题频率大于等于0.5次/GB，小于1次/GB |
| 三级 | 数据质量好 | 数据质量问题频率小于0.5次/GB |

通过对数据质量问题频率的考评和等级划分，就可以从数据仓库众多的数据中解放出来，集中精力把有限的资源投入到需要重点关注的主题数据。因此数据质量可信等级是数据质量提高的有效途径。与之相配套的，必须建立了一套相关的管理制度，管理制度主要包括：

（1）可信等级初始值确立流程

数据中心平台各主题域的数据质量可信等级初始值一般设定为一级，由数据质量管理小组牵头，执行全面检查后报数据中心主管部门批准后确立。

（2）可信等级变更流程

数据质量管理小组每季度、每年组织定时抽查、全面检查时，每半年根据检查结果对相关主题域的可信等级提出调整意见，报数据中心主管部门批准后确立。

数据质量评估具体方法：

对于具体数据的质量检查模式采用记录数检查法、关键指标总量验证法、历史数据对比法、值域判断法、经验审核法及匹配判断法。通过这些方法方法，可以对单个数据点的数据准确性进行检查，及时发现数据质量问题。

（1）记录数检查法

通过比较记录条数，对数据情况进行概括性验证。主要是检查数据表的记录数是否为确定的数值或在确定的范围内。

适用范围：

对于数据表中按日期进行增量加载的数据，每个加载周期递增的记录数为常数值或可以确定的范围时，必须进行记录条数检验。

（2）关键指标总量验证法

对于关键指标，对比数据总量是否一致。主要是指具有相同业务含义，从不同维度统计的汇总逻辑的检查。

适用范围：

同表内对同个字段从不同的维度进行统计，存在汇总关系时，必须进行总量检验。

本表的字段与其它表中的字段具有相同的业务含义，从不同的维度统计，存在汇总关系，且两张表的数据不是经同一数据源加工得到。满足此条件时必须进行总量检验。

（3）历史数据对比法

通过历史数据观察数据变化规律，从而验证数据质量。通常以同比发展速度进行判断。评估时应根据各种指标发展特点，重点对同比发展速度增幅（或降幅）较大的数据进行审核。历史数据对比法包括同比和环比两种方式。

适用范围：

不能进行记录数检查法、关键指标总量验证法，且事实表的记录数小于1000万条时必须进行历史数据对比法。

（4）值域判断法

确定一定时期内指标数据合理的变动区间，对区间外的数据进行重点审核。其中数据的合理变动区间范围是直接根据业务经验来确定的。

适用范围：

事实表中的字段可以确定取值范围，同时可以判定不在此范围内的数据必定是错误的。满足此条件必须进行值域判断法。

（5）经验审核法

针对报表中指标间逻辑关系仅靠计算机程序审核无法确认、量化，或有些审核虽设定数量界限，但界限较宽不好判定的情况，需要增加人工经验审核。

适用范围：

以上方法都不适用的情况下，可以使用经验审核法。

（6）匹配判断法

与相关部门提供或发布的有关数据进行对比验证。

适用范围：

与有相关部门提供或发布的有关数据口径一致的，可以使用匹配判断法。

# 附件C 数据质量管理流程

（一）数据质量校验流程

数据仓库每天都有很多ETL任务定时执行加载数据，确保ETL加载数据的完整性、准确性是数据质量管理的基本要求。  
 1）日常数据校验  
 数据质量管理人员每天要对ETL加载任务执行情况进行检查。

数据校验方法选择

ETL任务数据质量校验要求必须采用以下三类方法中的至少一种来进行判断：记录数检查法；关键指标总量验证法；值域判断法。

数据校验周期

每天ETL加载任务比较多，如果全部执行数据校验需要的时间过长，因此根据每个主题数据的可信等级确定校验频率。

可信等级与校验频率的对应关系如下：

一级：每次加载都必须执行数据校验

二级：每三次加载执行一次数据校验

三级：每六次加载执行一次数据校验

对于需要特别保障的主题数据，可调整校验频率并额外增加经验审核法。

2） 定时数据抽查

数据校验确保每天加载的增量数据的完整性、准确性，在此基础上，数据质量管理小组必须每季度组织一次数据仓库的定期抽查。

定期抽查的范围必须包括可信等级为一级的所有主题数据，可信等级为二级的二个主题的数据，可信等级为三级的一个主题的数据。

定期抽查必须采用数据质量评估方法中定义的所有方法。

3）全面数据检查

数据质量管理小组必须每年组织一次数据仓库的全面检查。

全面检查的范围包括企业数据中心平台所有主题的数据。

全面检查必须采用数据质量评估方法中定义的所有方法。

（二） 数据异常处理流程

（1）数据质量管理人员发现数据差错应及时核对核实，根据核对核实情况填写数据问题处理单，描述数据质量问题的现状、原因和改正、预防措施。

（2）数据质量管理小组组长审批后，报数据中心主管部门批准后执行数据修正任务。

（三）数据质量的评估报告

最终根据数据质量检查的情况，数据质量管理小组会定期或不定期的生成相关的数据质量评估报告，数据质量报告分为两类：

每月定期提交的数据质量报告，即每月数据质量问题的陈述。

数据质量抽查或全面检查后提交的数据质量报告。

1. 每月数据质量评估报告

每月的数据质量报告是在每月月末或者下月月初时编制的月度执行报告，数据质量管理小组会对当月数据质量情况进行汇总统计，并根据“数据质量问题频率”的变化情况对各个主题的等级进行检查策略的调整，报告格式如下：

A、非系统问题

（1）数据质量问题的现象

（2）业务数据范围

（3）问题陈述

B、系统问题

（1）数据质量问题的现象

（2）业务数据范围

（3）问题陈述

2. 数据质量抽查或全面检查后提交的数据质量报告

数据质量抽查或全面检查报告是在每次进行完整个企业数据中心抽查或全面检查后编制的质量报告，同每月数据质量报告相比，除了要统计数据质量情况并重新划分主题等级外，还需要对整个质量体系运作的情况进行评估和改进，报告格式如下：

A、基本概况

包括：参与数据质量检查的相关组织与人员；数据质量检查的时间；数据质量检查的地点；数据质量检查的形式；数据质量检查的范围等方面的内容。

B、数据质量的检查与评价

（1）评价过程与步骤

（2）数据检查方式

主要分为全部检查和抽查。全检必须说明检查的范围、内容及方法。抽查必须说明抽样方案、过程及数据检查的范围、内容及方法。

（3）数据质量评价方法

C、数据质量评述与结论

（1）评述

对数据质量进行的综合描述（包括存在问题）。

（2）结论与建议

包括检查结果、可信等级调整建议等。