

文章编号: 1005-8451 (2017) 01-0031-05

## 铁路主数据管理平台关键技术研究

邹丹

(中国铁道科学研究院 铁路大数据研究与应用创新中心, 北京 100081)

**摘要:** 针对铁路信息化建设过程中出现的分散建设, 基础核心数据不统一、系统间无法互联互通的问题, 建设了铁路主数据管理平台。对平台建设过程中的主数据规范建立、主数据全生命周期管理、数据综合分析以及对铁路主数据的质量管理等关键技术给出了详细的设计思路及实现方法, 对于铁路主数据平台建设的逐步完善有借鉴作用。

**关键词:** 主数据规范; 数据全生命周期管理; 数据质量管理

**中图分类号:** U29 : TP39 **文献标识码:** A

### Key technologies of railway master data management platform

ZOU Dan

(Research and Application Innovation Center for Big Data Technology in Railway, China Academy of Railway Sciences, Beijing 100081, China)

**Abstract:** The construction of the railway master data management platform was built to solve the problem that the railway system was built with independent and repeat period, the basic core data was not unified, the system could not be interconnected. This article provided the design ideas and implementation methods in detail for the master data specification, lifecycle management, data comprehensive analysis, quality management of railway master data and other key technologies during the process of platform construction. This had a reference for the gradual improvement railway master data platform construction.

**Keywords:** master data specification; data whole lifecycle management; data quality management

随着大数据时代的到来, 数据已经成为企业一种重要的资产。铁路在信息化建设过程中积累了大量历史数据, 但数据来源不统一, 各系统分散建设、独立维护, 信息不对称等问题普遍存在。在进行系统整合与数据集成的过程中, 由于基础核心数据不统一, 系统与系统间无法进行互联互通和交互共享的问题日益突出, 铁路核心基础数据整合, 统一维护管理的问题亟待解决。一个可对铁路核心基础数据进行综合维护管理及统一集中发布的铁路主数据管理平台应运而生。

铁路主数据管理平台将铁路多个业务系统共享性强、需要统一的数据, 通过先进的技术和流程, 对数据进行准确、及时的维护管理与共享分发。通过建立数据标准规范、面向服务的架构和系统集成应用等关键技术, 将统一、完整、规范、准确的数据及时提供给各信息系统使用。平台建设的关键技术

主要包括: 铁路主数据规范体系的建立; 应用数据全生命周期管理的方法对主数据进行维护管理; 主数据变更管理以及数据活跃度分析和对主数据进行数据质量的监控与管理。

### 1 铁路主数据规范建立

铁路主数据的维护管理, 需要建立一个完整完善的主数据规范。规范内容包括: 主数据名称、各主数据中包括的字段、字段内容的详细描述、主数据的业务主管部门以及数据的维护方式, 以此作为主数据平台数据维护的依据。

建立铁路主数据规范, 要明确铁路主数据的含义, 主数据是指具有高业务价值的、可以在企业内跨越各个业务部门被重复使用的数据, 是单一、准确、权威的数据来源, 以此为原则确定铁路主数据的数据需求。对数据质量进行考核, 确定主数据平台的数据范围。明确主数据的采集方案, 建立数据的维护机制, 保证平台平稳运行。

收稿日期: 2016-06-25

基金项目: 中国铁道科学研究院科研专项课题(研发中心)(J2016X006)。

作者简介: 邹丹, 助理研究员。

### 1.1 明确数据需求

铁路主数据的需求建立在深入了解全路各业务领域信息系统数据需求的基础上。为此，需要对全国铁路运输组织、客货营销、经营管理、安全管理、建设管理、综合管理 6 大业务板块投入运营的重要业务应用系统进行详细调研。内容涉及各业务应用系统在用的基础数据情况；在用基础数据的维护流程和相关管理制度办法；以及对铁路主数据管理平台主要功能、接口方式、接口位置等方面需求。

### 1.2 确定数据范围

铁路主数据平台将铁路数据按照面向实体的分类方法划分为 6 大主题域，包括固定设施主题域、移动设备主题域、物资设备主题域、运输产品主题域、组织机构主题域和其他相关类主题域。根据对数据需求分析的结果以及对数据质量的评估确定了目前 6 大主题域中所包括的主数据。

- (1) 固定设施主题域：车站主数据、线路主数据等。
- (2) 移动设备主题域：客车车种主数据、客车车型主数据、客车车号主数据等。
- (3) 铁路物资设备主题域：物资分类主数据和物资主数据。
- (4) 铁路运输产品主题域：集装箱箱型主数据等。
- (5) 铁路组织机构主题域：铁路局主数据。
- (6) 铁路其他相关类主题域：国民经济行业分类主数据、性别主数据等。

### 1.3 主数据采集方案

铁路主数据管理平台的数据采集方案依照以下原则进行：对于已经形成国际标准或行业标准的主数据，采用相应标准进行采集，保证数据与发布标准完全一致；对于虽未形成相应标准，但铁路总公司有正式成文规定或在各系统中运行良好，也形成领域事实标准的主数据，选择依据标准性技术文件，并实时保持与业务系统同步的方式进行采集，采集方案示例如图 1 所示。

## 2 铁路主数据全生命周期管理

每个数据都需要经历从申请到审批，然后变更、

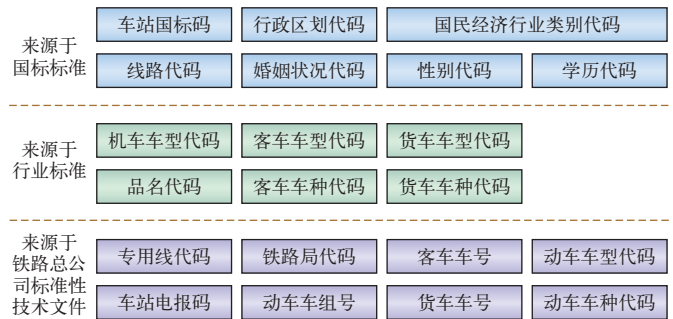


图1 铁路主数据采集方案

查询、分析，经过集成和分发，最终到归档这样一个生命周期。数据在生命周期的每个环节都需要进行科学管理，以提升数据价值。铁路主数据的全生命周期管理有 3 个重要环节：数据获取的方式与原则、数据清洗的规则制定和数据管理流程的控制。铁路主数据全生命周期管理如图 2 所示。

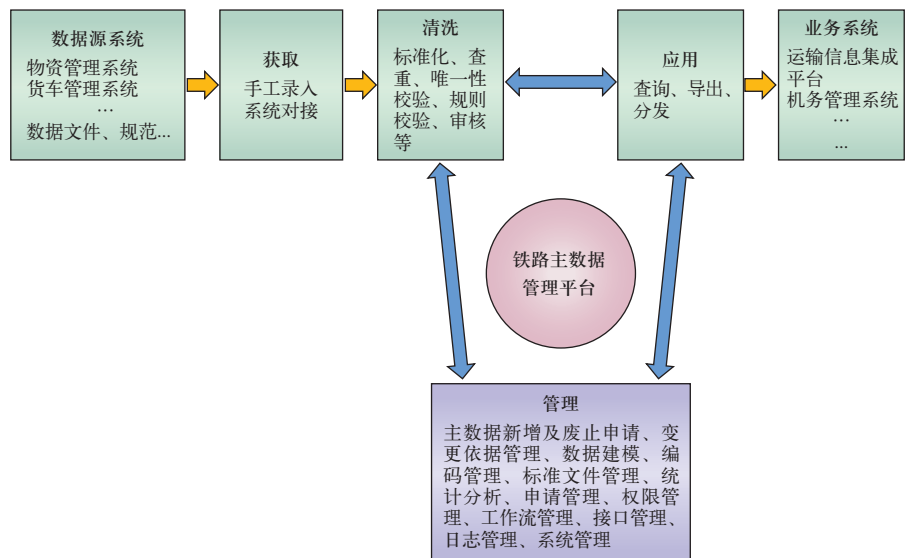


图2 主数据全生命周期管理

### 2.1 铁路主数据的获取

根据铁路主数据的数据需求，确定需要创建的数据内容。根据已出台的标准将创建的数据分类管理，入库的方式包括手工创建入库和系统接入。

#### 2.1.1 手工录入

运营维护部门在铁路主数据平台上按照平台主管部门与主数据业务主管部门确定的主数据规范进行数据建模、制定数据维护流程、建立数据校验规则。录入数据内容及相关标准、规范，数据质量由业务主管部门负责。

### 2.1.2 系统接入

运营维护部门按照对接系统与主数据管理平台相互协商好的接口方案进行对接，并在主数据平台上进行数据结构的建模，数据源系统负责数据的实时性，业务主管部门对数据质量负责。

## 2.2 铁路主数据的清洗

平台获取数据后，原始数据保存在临时库中，系统将创建的数据统一进行清洗后，导入平台生产库成为标准数据保存，并为下一步的发布做准备。

数据的清洗包括平台对数据的校验和管理层面上对数据的审核。初始化的数据经过查重与格式内容的多重校验后，由主数据业务主管部门确认该主数据符合准确、完整、唯一的质量要求，正式导入公用基础编码及主数据平台。

### 2.2.1 批量导入时数据查重原则

来源于多个系统的源数据在进行查重时，须按照以下先后顺序确定数据：有确切第三方资质依据的数据，如行政区划代码、国家及地区代码等，与第三方资质保持一致的数据作为主数据；符合少数服从多数原则，保持数据来源一致性。

### 2.2.2 批量导入时数据清洗规则

检查非空属性和编码的统一格式；在数据导入的过程中，需要对导入过程或导入结果进行监控；通过重复录入信息的方式检验系统的重复性判断是否准确；编码规则的检验。

## 2.3 铁路主数据的管理

铁路主数据的管理包括主数据新增/结构变更/废止的流程审批，数据变更依据的管理、数据标准文件管理、数据建模、数据管理、流程管理、权限管理、系统管理等多项功能。每个主数据的变更过程都需要根据业务主管部门的实际需求制定维护细则和对应的工作流控制。其中涉及工作流控制的功能主要包括以下几项。

### 2.3.1 主数据新增/结构变更/废止流程审批

主数据使用部门提出主数据新增、结构变更及废止的申请。平台根据主数据申请流程，以不同用户角色对流程中的环节进行操作及审批，保证主数据管理的每个环节都有对应主管部门负责。

### 2.3.2 主数据的更新管理

主数据内容的更新由业务主管部门提出，针对不同数据的操作方式有：(1) 业务部门直接在平台上进行操作，数据按照事先定义的流程进行操作及审批等各种环节的流转，并在主数据平台上进行新版本发布；(2) 由业务部门提交数据变更工单，由平台运维部门按照工单内容对数据进行变更并发布新版本，业务主管部门需对数据变更结果进行确认；(3) 直接以系统对接的方式进行数据更新，由数据源提供系统保证将变更数据及时同步到铁路主数据管理平台，数据质量由业务主管部门保证。

## 3 铁路主数据管理平台数据分析

铁路主数据管理平台的数据分析主要包括：主数据活跃度分析、用户活跃度分析、数据变更轨迹统计、数据下载情况统计。分别从数据、用户、数据与用户的关联关系这3个角度进行分析。其中，活跃度分析和变更轨迹分析对于优化平台主数据结构和提高统计分析的准确性起到了比较重要的作用。

### 3.1 数据及用户活跃度分析

数据及用户活跃度分析是主数据进行分析整合以及考核主数据管理工作的重要指标。通过活跃度分析，能够看出哪类主数据是各系统最为迫切需要的，用户活跃度能够分析哪些系统对于主数据的使用需求最为强烈。

铁路主数据的活跃度分析是基于时间段，分析不同主题域下各个主数据数据量的变更情况，图3和图4分别是统计某一时间段内各主题域数据的增长情况和用户登录情况。

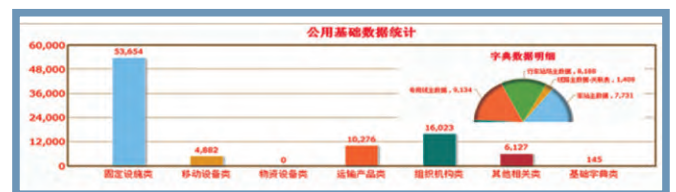


图3 按主题域统计主数据活跃度

### 3.2 主数据变更管理

在主数据管理过程中，对数据的变更管理将决定在以后的统计分析过程中，不会存在因数据的名称、编码发生了变化而导致最终统计结果出现较大误差的情况。铁路主数据管理平台中，为实现对主



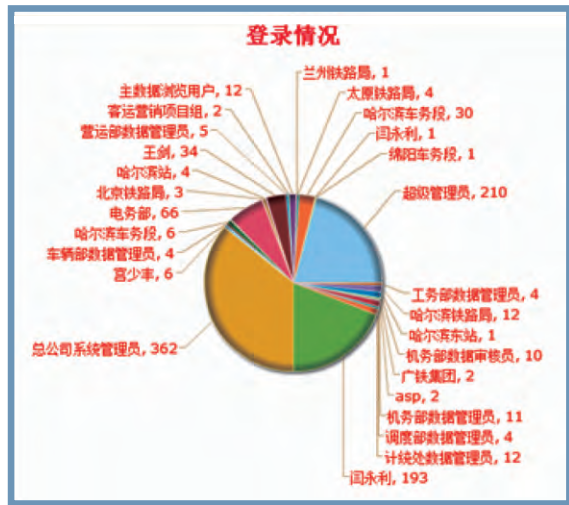


图4 用户活跃度统计

数据的变更管理，为每一条数据增加唯一标识ID并同时携带变更时间，此ID为该数据的基因标识，随机产生，不变且保证唯一。当数据由A变更到B再到C的过程中，无论数据发生了何种变化，此ID均不做任何变化，在数据统计分析、数据迁移等过程中，通过唯一标识ID和变更时间排序，即可得到该数据的整个变换过程，具体如图5所示。

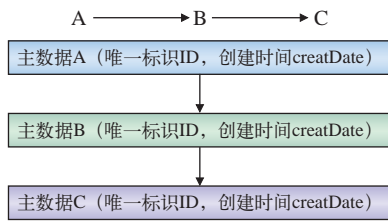


图5 主数据变化轨迹

### 4 铁路主数据质量管理

铁路主数据管理平台的数据质量管理，保证了铁路主数据在平台进行全生命周期管理的全过程顺利进行。数据按照识别、度量、监控3个主要的管理活动，对数据质量进行评估，分析发现问题，提高数据质量。铁路主数据管理平台的数据质量管理主要内容包括建立数据质量的评估维度，分析影响数据质量的因素，并按照分阶段的管理方法对平台

数据进行全方位的监控与改进。

#### 4.1 数据质量评估维度

铁路主数据评估的维度主要包括数据的完整性、规范性、一致性、准确性、唯一性和关联性。完整性评估指主数据中没有丢失或不可用的数据；规范性评估指主数据中的数据内容符合主数据规范定义的数据结构、数据类型、数据长度，并符合该数据定义的规则；一致性评估指数据内容含义在各个业务中保持一致，不存在冲突；准确性指主数据内容是符合业务逻辑的，不存在超出预期的数据；关联性指各主数据间不存在缺失或无效的索引。

#### 4.2 影响数据质量的因素

信息因素：产生这部分数据质量问题的原因主要有元数据描述及理解错误、数据度量的各种性质得不到保证和变化频度不恰当等。

技术因素：由于具体数据处理各技术环节的异常造成的数据质量问题。数据质量问题的产生环节包括数据创建、数据获取、数据传输、数据装载、数据使用、数据维护等。

流程因素：由于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，主要来源于系统数据的创建流程、传递流程、装载流程、使用流程、维护流程和稽核流程等各环节。

管理因素：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。如人员培训、人员管理、培训或者奖惩措施不当导致的管理缺失或者管理缺陷。

#### 4.3 数据质量管理方法

铁路主数据质量管理主要分为3个主要步骤：评估、认知和行动，如图6所示。

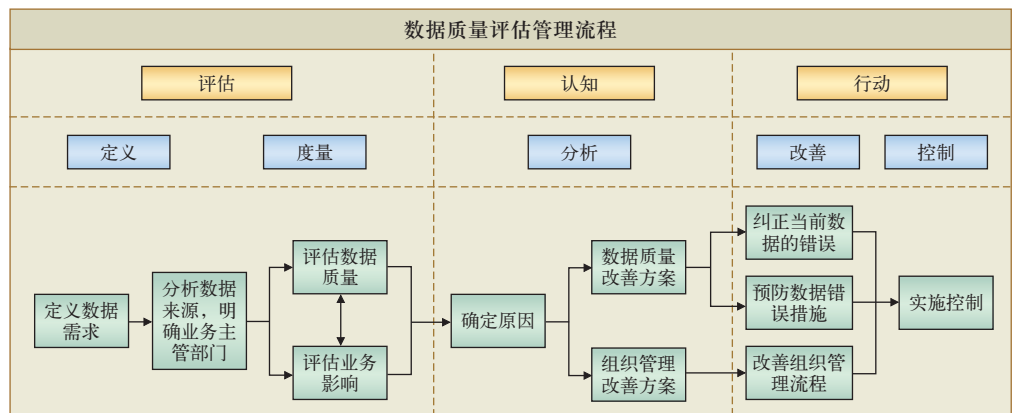


图6 数据质量评估管理流程

在数据质量评估阶段，需要定义数据需求，明确数据的业务主管部门；管理部门与业务主管部门共同按照数据需求及实际数据内容协商确定主数据的结构及初始数据源；对数据质量以及劣质数据对业务的影响程度进行评估。

在数据质量认知阶段，分析影响业务的因素并制定改善数据质量及组织管理的方法。

在改善数据质量的行动阶段，对数据的错误进行纠正，不断进行数据的清洗与验证，并对组织管理流程实施改进。对数据和管理实施监控，维护已改善的效果。

在对数据进行管理的过程中，业务主管部门、数据质量监督部门、运营维护部门、技术部门等都需要将沟通贯穿管理的始终，循环的评估组织管理流程，以确保数据质量改善的成果得到有效保持。

### 5 结束语

铁路主数据管理平台的建设目的是对铁路核心

基础数据进行集中、有效的管理，并作为统一的数据源头对各业务应用系统进行发布。本文对主数据从内容规范的确立到数据全生命周期过程管控，再到数据质量的管控与综合分析过程中所涉及到的关键技术进行了研究，结合平台建设过程中与各业务部门协调、沟通的实际情况，理论与实践相互修正，对进一步优化铁路主数据管理平台提供有力支撑。铁路主数据管理平台的逐步完善对于推进铁路信息化建设，实现全国铁路公共基础数据的统一、系统互联互通，提升资源共享的整体水平具有借鉴意义。

### 参考文献：

[1] 中国铁道科学研究院. 铁路公用基础编码及主数据管理平台总体方案 [R]. 北京：中国铁道科学研究院，2014.  
 [2] 赵飞. 基于全生命周期的主数据管理MDM 详解与实践 [M]. 北京：清华大学出版社，2015.  
 [3] 俞荣华. 数据质量和数据清洗关键技术研究 [D]. 上海：复旦大学，2002.

责任编辑 陈蓉

(上接 P30)

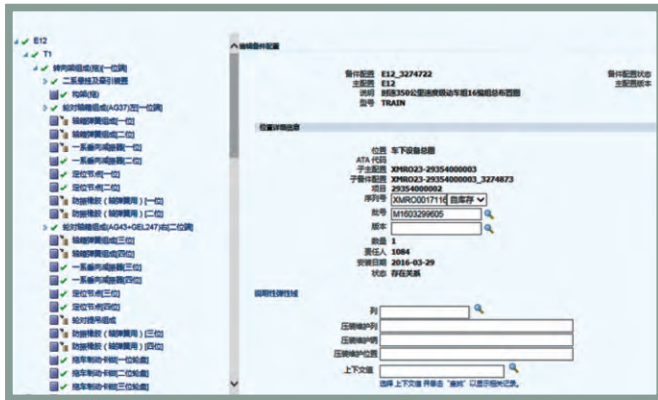


图5 基于序列号的静态履历查询

理需求，提出了面向全生命周期高速动车组履历系统功能，建立了面向全生命周期的高速动车组履历系统架构。通过基于序列号的高速动车组全生命周期履历信息系统，可以有效地利用零部件生命，降低库存水平，同时实现对关键部件的全生命周期管理，发生问题时，按序列号实现在检修、运用、生产及设计各阶段信息的反向追溯。

### 参考文献：

[1] 管江旗,王钰,贾志凯. 新造动车组履历信息同步方案的研

究与应用 [J]. 铁路计算机应用, 2012, 21 (12) : 19-21.  
 [2] 王建民,任良全,张力,等. MRO 支持技术研究 [J]. 计算机集成制造系统, 2010, 16 (10) : 2017-2024.  
 [3] 郝伟,曹炳欣,郭宗鹏,等. 面向全生命周期的高速列车 MRO 系统应用 [J]. 铁路计算机应用, 2016, 25 (2) : 40-42.  
 [4] V.V. Pantelev, V.A. Kamaev, A.V. Kizim.: Developing a Model of Equipment Maintenance and Repair Process at Service Repair Company Using Agent-based Approach[J].Procedia Technology,2014,16(1): 1072-1079.  
 [5] 王建民. MRO2：制造业新的增长点 [J]. 中国制造业信息化, 2010 (5) : 13-15.  
 [6] EckartUhlmann, Martin Bilz, Jeannette Baumgarten.: MRO – Challenge and Chance for Sustainable Enterprises[J].Procedia CIRP, 2013(11): 239-244.  
 [7] 程曜安,张力,刘英博,等. 大型复杂装备 MRO 系统解决方案 [J]. 计算机集成制造系统, 2010, 16 (10) : 2026-2037.  
 [8] 胡阳,夏榆滨,张力,等. 面向 MRO 领域的分布式系统性能测试方法 [J]. 计算机集成制造系统, 2013, 19 (12) : 3026-2034.  
 [9] 莫易敏,田斌,袁新宇,等. 机车履历管理信息系统的研制与开发 [J]. 武汉理工大学学报, 2005, 27 (4) : 106-108.

责任编辑 陈蓉