

文章编号: 1005-8451 (2021) 03-0023-04

基于铁路数据服务平台的铁路数据资产管理研究

王 喆, 马小宁, 邹 丹, 王沛然, 孙思齐

(中国铁道科学研究院集团有限公司 铁路大数据研究与应用创新中心, 北京 100081)

摘要: 随着铁路企业公司化改革的不断深入, 铁路数据的资产化以及服务于“强基达标、提质增效”的工作主题成为铁路企业越来越紧迫的需求。当前铁路数据资产管理还面临着资产盘点不清、共享困难、缺乏安全分级以及主数据不规范等一系列问题。文章明确了铁路数据资产管理相关重要环节, 并基于铁路数据服务平台的服务能力和开发引擎, 构建了一系列服务于数据资产管理应用, 实现了数据从盘点、接入到安全分级和共享的全流程管理, 在实践中取得了很好的应用效果。

关键词: 数据资产管理; 铁路数据服务平台; 数据登记; 数据汇聚

中图分类号: U29 : F530.67 : TP39 **文献标识码:** A

Research on railway data asset management based on railway data service platform

WANG Zhe, MA Xiaoning, ZOU Dan, WANG Peiran, SUN Siqi

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

Abstract: With the deepening of corporate reform of railway enterprises, the capitalization of railway data and serving for the work theme of “strengthening the foundation, reaching the standard, improving quality and efficiency” have become more and more urgent needs of railway enterprises. At present, railway data asset management is also facing some problems, such as unclear asset inventory, difficulty in data sharing, lack of security classification and non-standard master data. This paper defines the important links of railway data asset management. Based on the service capability and development engines of railway data service platform, a suite of data asset management applications are constructed, which realizes whole-process data asset management involving inventory, access, security classification and sharing, and achieves good application effect in practice.

Keywords: data asset management; railway data service platform; data registration; data aggregation

随着企业信息化不断深入发展, 数据对企业的价值愈发重要。瑞士达沃斯经济论坛的经济报告曾称, 数据是一种类似于货币或基金的新型资产^[1], 尤其对于互联网企业, 其数据资产的价值已远超企业其它资产价值之和。在“互联网+”背景下, 传统企业汲取互联网企业的经验, 重视大数据对企业数字化转型带来的推动作用, 建立起数据采集、积累、分析、共享等管理流程, 努力使数据资产成为企业重要的生产要素, 以服务于企业整体战略目标。

1 铁路数据资源及数据资产管理现状

1.1 铁路数据资源现状

收稿日期: 2020-07-09

基金项目: 国铁集团科技研究开发计划课题 (K2019B005)

作者简介: 王 喆, 副研究员; 马小宁, 研究员。

中国国家铁路集团有限公司(简称: 国铁集团)自2019年挂牌成立后, 作为市场经营主体, 实施货运增量行动、客运提质计划、复兴号品牌战略3项举措, 重视经济效益和投入产出, 在规划建设、运营生产中积累了大量得天独厚的优质数据资源^[2]。如何充分利用这些数据资源, 挖掘数据潜在的价值, 以提升客货运服务质量、提高经济效益成为铁路当前重要课题。同时, 海量数据资源也成为管理难题, 如何有效地管理铁路数据资源, 将其转化为对企业有价值的数据资产并服务于运营和决策是亟需研究解决的问题。

1.2 铁路数据资产管理现状

(1) 数据资产全貌尚未掌握: 铁路数据资产庞杂, 分散在200多个国铁集团统一建设的业务系统及

各铁路局、客专公司自建的上千个业务系统中，数据类型复杂多样，包括结构化数据及图像、音频、视频、文本等非结构化数据。虽已开展多次铁路数据资产调研，但调研范围不一致，调研资料缺乏梳理，利用率不高，无法形成完整的铁路数据资产视图，不利于企业制定数字化转型相关战略。

(2) 数据内部共享难题：铁路信息系统繁多，系统之间数据共享场景多样^[3]；此外，铁路企业内部数据共享审批流程复杂，一事一议，数据共享技术标准化较差、接口不统一等问题成为企业内部数据共享的障碍，妨碍了数据价值的发挥。

(3) 与外部数据交换不足：与其它行业（如银行、电信运营商、电网等）相比，目前铁路数据与外部数据的交换明显不足；对外部数据的引入还比较有限，尤其亟需将气象、地质、社交媒体等数据与铁路自有数据结合，以更好地支持辅助决策；另外，囿于对数据安全的担忧和数据价值的认识，对外提供铁路数据的场景也极为有限。

(4) 主数据规范推进困难：铁路主数据是描述铁路基础设施、客运服务、货运服务、人员及机构基本标识或状态属性的数据^[4]。国铁集团于2015年建设了铁路主数据平台，已发布车站主数据、线路主数据、路局主数据、动车组主数据等主题数据，但目前仍有很多铁路信息系统尚未接入铁路主数据平台，导致数据融合受限。

2 铁路数据服务平台的数据资产管理

铁路数据资产管理包括资产盘点、资产汇聚、价值提升、资产流通等重要环节。铁路数据服务平台提供了基础的数据管理服务，为数据应用的开发提供多种接口和引擎。目前，基于平台提供的数据存储引擎、标签管理引擎、数据访问接口、数据开发引擎、工作流引擎等平台层技术工具，已实现数据登记、数据汇聚、数据质量管控、数据分级管理、数据共享等一系列应用，构建起铁路数据资产管理应用集合，其总体架构如图1所示。

2.1 铁路数据登记

铁路数据登记是实现铁路数据资产管理的首要环节，包括资产调研表管理、数据资产内容检索、

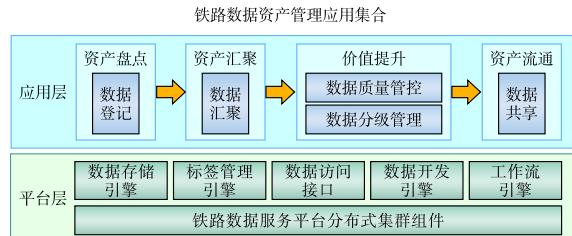


图1 铁路数据资产管理应用总体架构

数据资产统计等功能，解决铁路历次数据调研获取的调查研表的数字化管理问题。随着全面调查和阶段性日常调查工作的开展，可逐步完善铁路数据资产全貌。同时，数据登记也给每个铁路应用系统建立了数据资产卡片，是后续开展各应用系统数据汇聚的基础。

在数据资产登记的基础上，可对资产情况进行统计分析，如结构化数据量（存量和增量）、非结构化数据量、网络部署情况、数据库类型等，从多个角度掌握铁路数据资产概况，为制定相关政策提供决策依据。

2.2 铁路数据汇聚

铁路数据汇聚是在铁路数据资产登记的基础上，将业务系统中存储的数据传输至铁路数据服务平台的过程，包括接口方案制定、接口开发和接口调试等环节；汇聚至铁路数据服务平台的数据是开展铁路数据资产管理的对象。

制定数据汇聚方案要综合考虑数据的类型（结构化和非结构化）和数据传输量的大小。铁路数据服务平台提供3种标准的数据接入接口：RESTful接口、FTP传输和Kafka接口，这3种接口的特点见表1。

表1 铁路数据服务平台提供的数据接入接口及其特点

接口方式	使用场景	身份认证	开发难度	适用的数据类型
RESTful	日增量小于500 M，单次传输小于50 M的数据	平台分配访问令牌	一般	结构化数据
FTP/SFTP	实时性要求不高，增量数据量较大，易于导出为文件的数据表	平台提供FTP/SFTP账户密码	容易	结构化/非结构化数据
消息队列	实时性要求较高、并发量和吞吐量较大的数据	用户名/密码	较复杂	结构化数据

铁路数据服务平台的存储架构设计采用“数据湖”理念，即以所汇聚数据的原始格式存储数据，对外提供数据服务时，根据应用需求对数据进行转

换。平台中数据存储方案要考虑数据的类型和数据应用场景需求(实时和非实时)。铁路数据服务平台的存储架构集成多种存储组件,如分布式文件系统、数据仓库、内存数据库、对象存储等,能够为铁路数据资产管理提供稳定可靠的存储设施,表2概括了各种存储组件的特点及其适用场景。

表2 铁路数据服务平台各类存储组件的特点及其适用场景

存储组件类型	特点和适用场景
分布式文件系统	处理PB级别的超大文件,扩展性强; 流式地访问数据,针对一次写入多次读取的使用模式; 对硬件设备的要求比较低,使用低成本的硬件会导致集群中节点故障率升高; 数据在系统中的每个节点上均衡分布,充分利用磁盘空间。
数据仓库	支持SQL访问,只能处理结构化数据; 数据必须与数据仓库事先定义的模型吻合; 并行处理,对分析计算效率较高。
内存数据库	高并发、低延迟、高可用; 适合数据关联性强、数据模型复杂的场景,一定程度上比NoSQL数据库更能保证ACID事务。
对象存储	存储海量图片、音视频、日志等文件; 每个对象都有自己的属性,如创建时间、生命周期等。

2.3 数据质量管控

铁路数据服务平台的铁路数据来源广泛,各业务系统在数据采集密度、数据规范、需求场景、质量管理方面各不相同,对数据完整性、一致性、及时性、精确性有较大的影响^[5]。

通常来说,数据质量的提升要从管理和技术2个方面开展管控工作^[6];管理方面的工作主要包括建立数据质量管控组织、数据质量问题发现和分析、数据质量考核等;技术方面主要依靠数据质量度量规则管理和稽核工具。

在数据质量管控方面,铁路数据服务平台提供质量监督角色设置、度量规则定义以及生成数据质量报告等功能。此外,针对车站和线路字典数据,提供数据质量检查和稽核的工具,可将铁路主数据平台提供的主数据作为参考,对平台上来自具体业务系统的相关主数据进行质量检查,采用文本相似度分析和时间拉链等方法,查找并纠正车站和线路字典相关的数据问题,显著提高了数据质量检查效率,车站字典数据质量检查过程如图2所示。

2.4 数据分级管理

铁路数据分级是在数据分类的基础上,采用明确、规范的方法评估数据的重要性和敏感度差异,确定各类数据的级别。数据分级有助于业务部门根

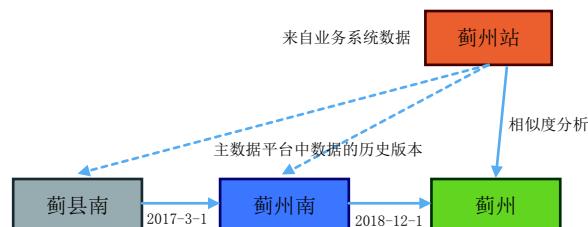


图2 车站字典数据质量检查过程示意

据数据不同级别,以及数据在其生命周期所处的阶段,确定需要采取的数据安全防护策略和管控措施,进而提高业务部门对数据的管理和安全防护水平,确保数据的完整性、保密性和可用性^[7]。依据铁路数据重要性不同,划分为公开、内部、机密和国家秘密4个等级,如图3所示。



图3 铁路数据分级示意

对于不同安全等级的数据,铁路数据服务平台采用对应的安全管理策略。如在数据共享环节,公开数据采取无条件公开方式,数据申请者可直接获取相关数据而无需审核。对于非公开数据,数据申请者提出的使用申请需要由对应安全级别的审批部门审核。在数据存储环节,内部数据采用按需加密,机密数据采用全量加密的策略。涉及国家秘密的数据,铁路数据服务平台暂不考虑汇聚。

2.5 数据共享管理

依据与原始数据的关系,铁路数据服务平台中的数据可分为未加工数据和加工数据,针对这2种数据提供数据目录和接口服务2种数据共享方式。

数据目录将未加工数据的摘要,包括数据说明、数据字段说明、数据来源、数据更新频率等,按照一定的分类方式整理形成数据目录并发布。对于加工后的数据(一般针对定制化加工需求),则发布在接口服务中。数据使用者对数据目录或者接口服务中的数据提出使用申请,申请被批准后,可通过

数据接口获得所需数据。

3 结束语

铁路数据服务平台已经在武清主数据中心部署，并针对京张高铁主题，盘点和汇聚多个业务系统的数据，支持智能京张高铁的大数据应用。今后，将继续完善铁路数据服务平台的分级管理体系，完善数据质量稽核算法，通过数据质量的提升来提高数据价值^[8]，为铁路各类业务数据融合和广泛共享奠定扎实的基础。

参考文献

- [1] 李雅雄, 倪 杉. 数据资产的会计确认与计量研究 [J]. 湖南财政经济学院学报, 2017, 33 (4): 82-90.

- [2] 闻克宇, 王 龙. 大数据环境下铁路运营数据资产价值分析与展望 [J]. 铁道经济研究, 2018 (3): 32-34.
- [3] 刘彦军, 李 平, 马小宁, 等. 铁路大数据资产管理平台的研究与设计[C]//第十二届中国智能交通年会大会论文集. 北京: 中国智能交通协会, 2017: 351-356.
- [4] 杨连报, 李 平, 马小宁, 等. 铁路主数据全生命周期管理研究[C]//第十二届中国智能交通年会大会论文集. 北京: 中国智能交通协会, 2017: 365-369.
- [5] 赵 冰, 李 平, 代明睿. 铁路大数据质量评估与优化方法研究 [J]. 中国铁路, 2018 (2): 63-67.
- [6] 邹 丹, 马小宁, 王 咸. 铁路大数据平台架构研究 [J]. 铁路计算机应用, 2019, 28 (8): 1-4.
- [7] 武 威, 马小宁, 刘彦军, 等. 铁路数据服务平台安全策略研究 [J]. 中国铁路, 2019 (8): 63-68.
- [8] 张艳君. 林业统计数据质量评估体系及其应用价值分析 [J]. 农村经济与科技, 2020, 31 (8): 37-38.

责任编辑 桑苑秋