

基于数据编织的铁路数据管理架构及典型场景应用

孙思齐¹, 邹丹¹, 张凯¹, 任爽², 刘敏¹

1. 中国铁道科学研究院集团有限公司电子计算技术研究所, 北京 100081;

2. 北京交通大学计算机科学与技术学院, 北京 100081

摘要

随着铁路行业数字化转型的深入, 多源异构数据的高效管理与融合应用成为提升运营效率的关键挑战。提出了基于数据编织技术的铁路数据管理新架构, 通过构建包含数据源层、元数据处理层、数据服务层等7层的协同体系, 实现了铁路车机工电辆供等业务系统数据的跨源整合与实时共享。聚焦主动元数据管理、知识图谱数据目录、跨源一体化查询等关键技术, 设计了工务设备故障处置等典型应用场景, 验证了数据编织技术在打破数据“孤岛”、提升故障处置效率及优化协同管理中的有效性。实践表明, 该架构可显著提升铁路数据治理能力, 为铁路数字化转型提供了可扩展的技术范式。

关键词

数据编织; 主动元数据; 知识图谱; 跨源查询; 故障协同处置

中图分类号: U29-39

文献标志码: A

doi:10.11959/j.issn.2096-0271.2026005

Railway data management architecture and typical scenario application schemes based on data weaving

Sun Siqi¹, Zou Dan¹, Zhang Kai¹, Ren Shuang², Liu Min¹

1. Institute of Computing Technologies, Chinese Academy of Railway Sciences, Beijing 100081, China

2. Beijing Jiaotong University, Beijing 100081, China

Abstract

With the deepening of digital transformation in the railway industry, efficient management and integrated utilization of multi-source heterogeneous data have become key challenges for improving operational efficiency. This paper proposes a novel railway data management architecture based on data weaving technology. By constructing a seven-layer collaborative system-comprising the data source layer, metadata processing layer, data service layer, and four additional supporting layers-it enables cross-source integration and real-time sharing of data from business systems such as train-ground communication, mechanical operations, electrical supply, and rolling stock management. The study focuses on critical technologies, including proactive metadata management, knowledge-graph-driven data cataloging, and unified cross-source querying. It also designs a typical application scenario for collaborative fault handling in track and equipment maintenance. Experimental results demonstrate that the proposed architecture effectively breaks down data silos, improves fault resolution efficiency, and optimizes collaborative management. This

architecture substantially enhances railway data governance capabilities and offers an extensible technical paradigm for the industry's digital transformation.

Key words

data weaving, proactive metadata management, knowledge graph, cross-source querying, collaborative fault handling

0 引言

在铁路智能化发展进程中，车机工电辆供等多专业系统产生的海量异构数据（如实时监测数据流、设备维修记录、巡检日志等）呈现爆炸式增长，传统数据管理模式面临数据“孤岛”林立、语义关联缺失、实时处理滞后等核心挑战。数据编织技术作为 Gartner 界定的战略技术趋势^[1]，通过元数据驱动的虚拟化与自动化手段，为跨源数据融合提供了创新性解决方案。国际上，IBM、Dell 等企业已在金融、医疗领域实现数据编织平台的规模化应用，而国内铁路行业在此领域的研究仍处于探索阶段。本文立足铁路业务特性，提出面向铁路场景的数据编织架构，旨在解决多源数据语义统一、实时关联与智能应用问题，为铁路数字化转型提供理论与技术支撑。为此，本文围绕主动元数据管理、基于知识图谱的数据目录、跨源一体化查询 3 项关键技术展开研究，并结合铁路工务设备故障处置的典型场景，充分验证数据编织架构在打破数据“孤岛”、提升故障处置效率及优化协同管理方面的有效性。

1 数据编织研究现状

数据编织作为一种先进的数据管理理念和技术体系，近年来在国外学术界和产业界引起了广泛关注。它旨在创建一个统

一的数据管理架构，能够跨各种数据源、环境和位置提供一致的数据访问、管理和治理能力。数据编织的核心理念是通过虚拟化和自动化手段，实现数据的无缝连接和统一管理，它的实现依赖于一系列关键技术，如数据虚拟化、数据目录、数据集成、元数据管理、数据治理和数据安全等。

（1）数据虚拟化

利用逻辑视图将物理上分散的数据集统一起来，使数据消费者可以像访问单一数据库一样访问所有数据。

（2）数据目录

构建一个全面的数据资产目录，包括数据的位置、类型、质量、所有权和敏感性等信息，便于数据的发现和使用。

（3）数据集成

利用抽取转换加载（extract transform load, ETL）、抽取加载转换（extract load transform, ELT）和应用程序编程接口（application programming interface, API）等工具，实现不同数据源之间的数据同步和转换。

（4）元数据管理

收集和维持关于数据的数据，包括数据血缘、数据质量指标和数据合规性信息，支撑数据的治理和分析。

（5）数据治理

确保数据的一致性、完整性和安全性，遵循行业标准和法律法规。

（6）数据安全

实施访问控制、数据加密和数据脱敏等措施，保护数据免受未经授权的访问和免于泄露。

1.1 国外研究现状

国外数据编织技术研究以 Gartner 提出的“元数据驱动-语义统一-动态适配”理论框架为核心，形成了从理论构建到行业落地的完整体系。在理论研究层面，齐艳平等^[2]指出数据编织通过逻辑化整合打破物理数据壁垒的核心优势，这与 LinT^[3]在基因组数据整合研究中验证的“语义层统一优先于物理集中”原则高度一致。

技术工具研发呈现平台化特征。IBM Cloud Pak for Data 通过集成主动元数据管理与数据虚拟化技术，实现跨云环境的数据逻辑整合，其内置的血缘分析模块可追溯数据全链路流转，这与《数据编织提升大数据使用效率》^[1]中强调的“动态数据治理”理念相吻合。Dell Boomi Atom Sphere 则聚焦动态编排技术，通过低代码界面实现 ETL 流程的智能调整。在零售行业案例中，该技术使跨渠道数据汇总时间从小时级缩短至分钟级。

行业应用已形成多领域渗透格局。在金融领域，Denodo 平台为美国大型银行构建分布式数据编织架构，实现 12 个州的客户数据实时访问，通过逻辑数据层替代物理集中，使合规审计效率大幅提升，印证了张连明等^[4]关于银行数据云构建中“虚拟化+元数据”技术组合的有效性。在医疗健康领域，ARPA-H 生物学数据项目整合全美临床与科研数据，采用差分隐私增强算法，在保护隐私的同时，其核心技术路径与 LinT 的基因组数据整合方法一脉相承。在交通运输领域，波兰铁路通过虚拟化数据中心改造，提升了调度系统计算能力，虽未明确采用完整数据编织架构，但其“动态资源适配”策略已体现数据编织的核心思想，为铁路场景应用提供了技术参考^[5]。

1.2 国内研究现状

相较于国际上数据编织技术的成熟应用，国内的数据编织技术研究和应用起步较晚，但在政策引导和市场需求的驱动下，发展势头强劲。

国内数据编织发展应用的主要特征是“技术应用优先于理论体系构建”。肖东君等^[6]的两级政务数据治理实践具有代表性，该研究通过增强数据目录与语义映射技术，整合省-市-县 3 级政务数据，实现跨部门服务办理时长减半，其采用的元数据注册中心技术与 IBM 研究院提出的 CRDT 算法原理一致，实现了分布式环境下的元数据冲突消解。在金融领域，张连明等^[4]详细阐述了某银行利用数据编织构建新一代数据云的实践，通过数据虚拟化整合核心系统与信贷数据，结合动态编排技术实现“实时审批-批量统计”的智能切换，使风险评估效能得到提升。该案例与国外 Denodo 银行方案形成技术呼应，但更强调本土合规要求下的权限管控模块优化。目前在铁路行业尚处于技术探索阶段，现有研究呈现“局部适配”特征：刘北胜等^[7]设计的铁路基础设施全生命周期数据模型，通过本体建模实现跨阶段数据语义对齐，其技术思路与数据编织的语义层构建原则高度契合；李新琴等^[8]构建的设备故障知识图谱，本质上是增强数据目录在铁路场景的专项应用，为数据编织的知识图谱模块提供了行业化验证；单馨漪等^[9]的铁路通信数据转型研究则暴露了现有技术在跨业务整合中的不足，其指出的标准不统一、接口不兼容问题，正是数据编织需要解决的核心痛点^[10]。

1.3 小结

综合来看，国内外研究均验证了数据

编织技术在打破数据“孤岛”中的有效性，但铁路领域仍存在明显缺口。不同行业或者厂商对数据编织的理解和实现路径不尽相同，对于构建数据编织方案的技术组成也不一致，抽离它们的共性为本文构建铁路专用数据编织架构提供了明确的方向。数据编织的主要共性技术如下。

- 主动元数据技术：支持版本控制、血缘追溯与影响分析。
- 增强数据目录：基于知识图谱提升资产可发现性。
- 数据语义层：统一定义数据实体间键值与依赖关系。
- 数据虚拟化：实现跨数据源、跨数据结构、跨部门、跨组织、跨云的数据统一访问和查询。
- 动态数据编排：智能调整基于语义的逻辑数据链路的即时计算和预计算的比例，还可以智能更新预计算数据，实现ETL自动化的效果。

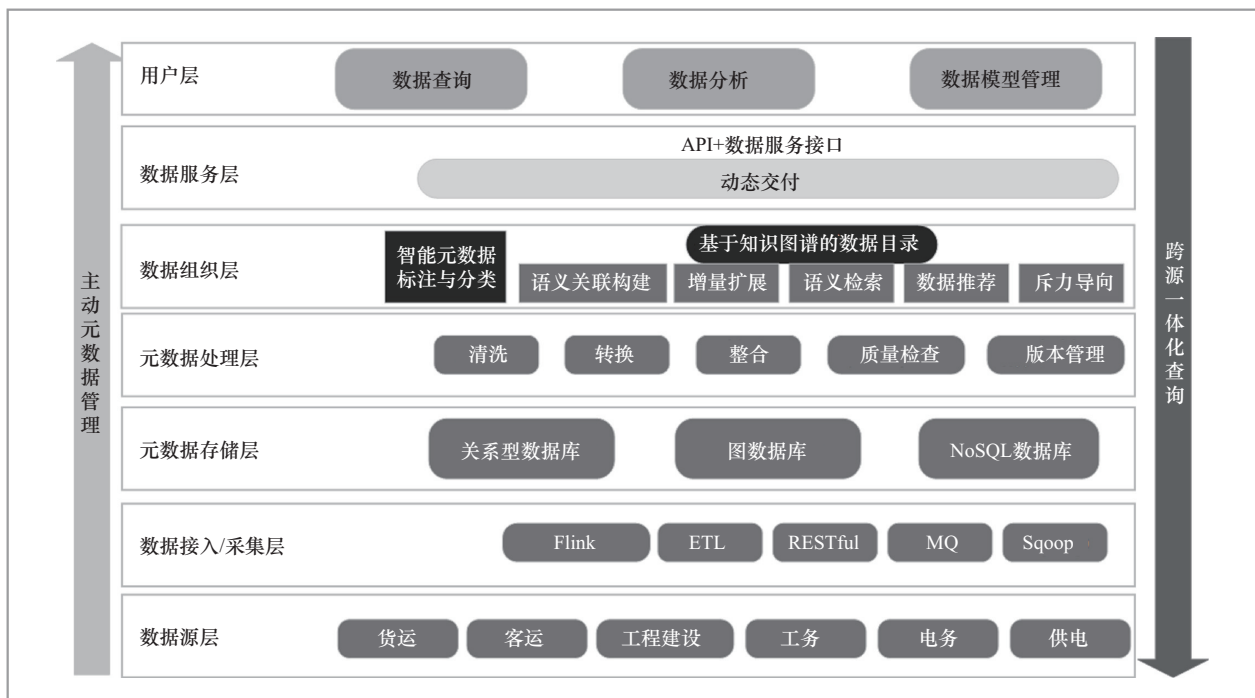
2 基于数据编织的铁路数据管理新架构

2.1 数据编织总体架构

基于数据编织的铁路数据管理新架构旨在打破铁路系统内多源异构数据之间的壁垒，实现数据的高效融合、实时共享与智能应用。该架构主要由数据源层、数据接入/采集层、元数据存储层、元数据处理层、数据组织层、数据服务层和用户层构成，各层之间紧密协作，为铁路数据管理提供强大的数据编织技术支撑。基于数据编织的铁路数据管理总体架构如图1所示。

(1) 数据源层

数据源层汇聚了铁路车机工电辆供等各类业务系统产生的多源异构数据，包括关系型数据库（如 Oracle、MySQL 等）、非关



系型数据库（如MongoDB、Redis、文件系统等）、实时数据流等。数据源层是整个架构的数据基础，为数据接入层提供原始数据。它与其他层的交互主要是被动的数据提供，根据数据接入层的请求将数据输出。

（2）数据接入/采集层

数据接入/采集层负责将数据源层的多源异构数据接入平台。它包含数据连接器、ETL工具、实时数据流处理框架等。数据接入/采集层接收数据源层的数据，经过处理后将其传递给元数据管理层。它是数据源层与上层架构之间的桥梁，负责数据的接入和初步处理。

（3）元数据存储层

元数据存储层是存储元数据的核心部分，采用合适的存储方式来保存元数据。常见的存储方式包括关系型数据库、图数据库和NoSQL数据库等，分别用于存储结构化元数据、复杂关联关系元数据以及海量半结构化/非结构化元数据。

（4）元数据处理层

该层对采集到的元数据进行处理和加工，以满足不同的应用需求。处理内容包括元数据清洗、转换、整合、质量检查以及元数据版本管理等。元数据处理层将处理后的高质量元数据返回存储层进行更新，同时为上层的数据组织提供支持。

（5）数据组织层

数据组织层基于元数据管理层的元数据构建铁路数据资产目录。它包括数据目录可视化界面、数据分类与标签、基于知识图谱的数据目录等。数据组织层为用户提供数据资产的导航和发现功能，用户通过数据目录发起数据请求。它与元数据处理层紧密相连，依赖元数据构建目录，同时将用户的查询需求反馈给数据服务层。

（6）数据服务层

数据服务层是数据编织架构的核心处

理层，负责接收和处理用户的数据请求。它包括数据请求解析器、跨源一体化查询引擎、数据缓存与优化模块等组件^[11]。数据服务层接收来自数据组织层的用户请求，根据数据处理层提供的元数据信息，从数据源层获取数据，并将处理结果返回给用户层。作为整个架构的数据处理中枢，数据服务层协调各层之间的数据交互，实现数据的高效利用。

（7）用户层

用户层面向多种类型的用户，包括数据科学家、分析师、工程师以及业务人员等。通过提供直观易用的界面和工具，用户层能够降低数据使用的门槛，提高数据分析和应用的效率，满足不同用户的需求。用户在使用过程中的反馈也可作用于数据组织层和元数据管理层，对数据目录和元数据进行优化，形成数据管理的闭环。

2.2 各层协同实现数据编织的流程

在基于数据编织的铁路数据管理新架构中，各层级紧密协同，共同实现数据编织的功能，打破数据壁垒，实现数据的高效融合、实时共享与智能应用。

各层协同实现数据编织的过程如图2所示，具体过程如下。

数据编织通过多个层级协同实现：首先，数据接入/采集层通过连接器获取数据源层（如列车时刻表）的各类原始数据，经ETL、实时流处理及数据虚拟化技术处理后，将标准化数据传递至元数据存储层；元数据存储层用多种数据库存储不同结构的元数据，并由元数据处理层进行清洗、整合、质量检查与版本管理，更新后存储并供给数据组织层；该层利用元数据构建铁路数据资产目录（支持分类、标签及知识图谱），响应用户检索需求并转交数据服务

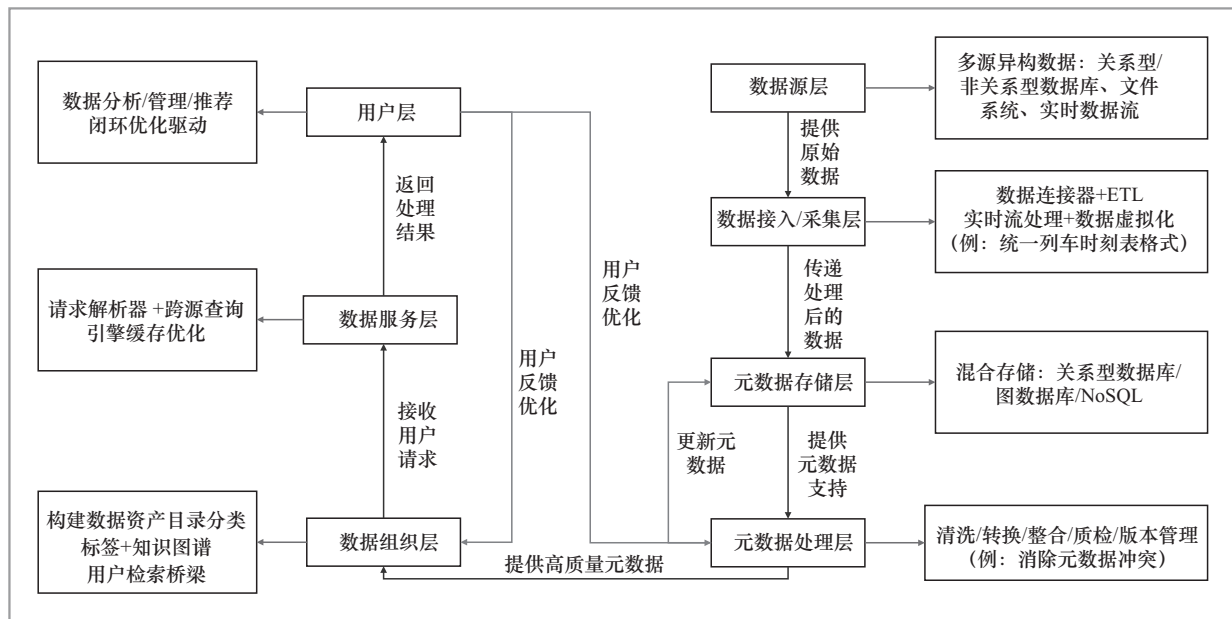


图2 各层协同实现数据编织的过程

层；数据服务层通过解析请求、跨源查询引擎与缓存优化处理用户查询，结果返回用户层；最终用户层的使用反馈闭环驱动数据目录与元数据层持续优化。7层架构的有机配合实现了对多源异构铁路数据从采集、处理到服务输出的全流程支撑，这种层次化设计确保数据在每一阶段都得到标准化和优化处理，各层协同机制也为数据编织场景应用提供了清晰的技术路径。

3 铁路数据编织关键技术及协同应用场景

数据编织技术通过整合多源异构数据，实现高效数据融合、共享与应用，其中在铁路行业，主动元数据管理技术、基于知识图谱的数据目录构建技术、数据跨源一体化查询技术在数据编织应用中发挥着关键作用。

3.1 主动元数据管理技术

主动元数据管理技术是对元数据进行自动采集与转换、智能标注与分类、动态监控与更新、自动化与智能化管理的技术，能够实时反映数据的变化及其历史演变。具体实现过程如下。在数据源端，各类传感器、数据库触发器和系统日志收集数据变化信息。元数据接入/采集层利用数据连接器获取这些信息，并通过ETL工具进行初步处理和格式转换。元数据存储层采用关系型数据库记录结构化元数据，图数据库存储元数据间的复杂关系，NoSQL数据库处理海量半结构化或非结构化元数据。元数据处理层中的清洗器、质量检查器和版本管理器协同工作，清洗无效或冗余数据，检查元数据质量问题，并管理元数据版本变更，确保元数据持续保持高质量和一致性^[12]。最终，高质量的元数据反馈用于更新元数据存储层，并为数据组织层提供支撑。主动元数据管理机

制可实现对元数据全生命周期的闭环维护，提高数据治理效率。

3.2 基于知识图谱的数据目录构建技术

基于知识图谱的数据目录构建技术是通过图数据库和知识图谱技术^[13]，构建数据之间的关联关系，通过图模型指导数据编织的数据组织技术。该技术主要包含两部分。

(1) 知识图谱构建

在铁路领域，知识图谱的构建首先需要人工设计出符合铁路业务逻辑的本体框架。该框架包含若干本体以及对应的属性和关系。本体是知识图谱的骨架，定义了铁路领域中的基本概念（例如，“列车”本体含{编号，最大载荷}等属性，“车站”本体与“列车”建立“停靠”关系）。完成基本框架设计后，根据结构化元数据中的实体和关系字段，将元数据项映射到相应的本体和关系。如果元数据中的实体或属性无法对应到已有本体/实体/属性，则需要对知识图谱进行动态更新，增补新的本体或关系^[14]。

(2) 知识图谱动态更新

知识图谱的动态更新指数据编织系统中，当底层数据源或其包含的具体数据发生变化时，知识图谱随之进行自动变更，包括本体、实体、关系、属性的增加、删除与更新等。这通常由数据接入/采集层检测到变化后触发更新机制。例如，当引入的新数据字段无法匹配现有图谱结构时，系统自动新增相应本体/属性；某设备实体属性值发生变化，触发更新图谱中对应节点的属性值。这种自动化的图谱更新机制可保证知识图谱与底层数据的一致性，保证数据目录的实时性和准确性。

3.3 数据跨源一体化查询技术

数据跨源一体化查询是指通过设计统一的数据查询接口，整合多个异构数据源的数据，使用户能够无缝访问不同数据库、数据平台或数据湖中的数据。用户通过数据服务层的统一接口提交查询请求，查询解析器根据元数据解析请求，确定数据来源和处理逻辑。跨源查询引擎将请求分解为子查询，并路由至相应数据源，利用分布式计算框架并行执行子查询。结果整合模块收集子查询结果，进行格式转换和合并，以统一格式返回给用户。跨源数据查询技术采用分布式查询架构，实现了元数据统一管理、多类型数据接入、统一语义管理、数据虚拟化、内存优化管理、高效的数据交换等功能。

- 统一的元数据管理：跨源数据查询引擎能够自动识别和管理不同数据源的元数据（如表结构、字段类型、索引、分区等），提供统一的元数据视图和接口，方便用户查询和管理。

- 多类型数据接入：支持关系型数据库、NoSQL数据库、文件系统、消息队列、流计算平台等多种类型数据源接入，实现对异构数据的一体化处理。

- 统一语义层：将不同数据源的逻辑模型抽象为统一的概念模型，支持用户使用标准的SQL语言进行跨源数据查询和分析，屏蔽底层数据源的差异和复杂性。

- 数据虚拟化：无须将不同数据源的数据提前抽取、转换和加载到中心化的仓库或湖中，而是在用户查询时动态地从各个数据源获取所需数据，实时整合计算，提高数据的时效性和准确性。

- 内存优化管理：有效地缓存和复用跨源数据查询过程中产生的中间结果，减少不必要的磁盘I/O和网络传输，提升查询性能和资源利用率。

以上关键技术为铁路数据编织提供了技术支持。

3.4 基于数据编织的铁路工务设备故障处置场景应用方案

铁路工务段管辖范围内的钢轨、道岔、桥梁等设备故障涉及多专业数据（如巡检记录、监测数据、维修工单等），而且数据分散在不同系统，关联关系复杂。例如，某次钢轨裂纹故障需关联巡检人员的巡检路线、设备历史维修记录、实时监测到的振动数据才能准确定位根源。采用数据编织技术构建统一的数据架构，可实现跨源数据的动态关联，打破原有数据“孤岛”，提高故障处置效率与人员协同能力。本节将工务设备故障处置作为典型场景，将主动元数据、知识图谱数据目录、跨源查询三大技术融合应用，验证其协同效果。

3.4.1 场景涉及的数据与实体

表1列出了该场景中的主要实体类型及其关键数据属性、数据来源系统和更新频率。

上述数据分散于多个业务系统：人员信息来自人力资源管理系统，设备状态和维修历史来自工务设备管理系统，故障记录来自发牌系统等。这些数据源在传统模式下各自独立，难以关联。本场景通过数据编织架构将相关数据实体连通，为故障

处置提供全面的信息支持。

3.4.2 数据编织关键技术协同机制

(1) 主动元数据管理：动态采集关联元数据与规则监控

在故障发生和处置过程中，主动元数据管理引擎实时捕获人员、设备、故障等数据资产的元数据变化，并进行智能标注和关联。例如，系统自动采集到巡检员李四的技能标签“钢轨探伤II级”、设备道岔D-03的属性“转辙机型号ZD6-E”、故障编号F-20240701的属性“紧急等级I级”等，将这些元数据标签赋予相应实体，方便后续关联分析。同时，主动元数据引擎通过自然语言处理（natural language processing, NLP）技术从维修工单文本中提取隐含关联，如解析出“巡检员王五在2024年6月10日巡检K15+000处时发现钢轨磨耗异常，关联后续维修工单WX-20240615”等信息，补充人员-设备-故障之间的隐性关联。主动元数据引擎还设有规则引擎进行动态监控。例如定义“巡检漏检预警”规则：如果某设备发生故障的位置在指定巡检路线范围内且该设备近30日内无巡检记录，则触发元数据告警标记“人员-设备-漏检风险”，并自动生成补检任务，推送至相应巡检员的移动终端。这些主动元数据管理措施，可使场景相关的人员、设备、故障数据得到及时关联和智能补全，为后续分析提供完整语境。

表1 实体类型与关键数据表

实体类型	关键数据	数据来源系统	数据更新频率
人员	巡检员工号、班组归属、技能等级、作业时间戳等	人员管理系统	实时更新
设备	钢轨编号、道岔型号、安装位置、服役年限、实时振动数据、历史维修记录等	工务设备管理系统等	分钟级/日级
故障	故障类型(如钢轨裂纹、扣件缺失)、发生时间、影响里程、处置状态、关联工单编号等	发牌系统等	实时

(2) 知识图谱数据目录：构建人员-设备-故障语义网络

利用知识图谱技术，将上述实体及其关系构造成可视的语义网络，为数据目录提供智能支撑^[15]（如图3所示）。该知识图谱以“人员-设备-故障”为核心节点，刻画出故障处置涉及的知识连接。例如，人员-设备关系方面，建立巡检人员与其负责线路区段内设备之间的关联（如“巡检员张三负责成渝线K10-K20区段设备巡检”）；设备-故障关系方面，将设备G-01与其发生的“焊缝裂纹”故障连接起来，并记录故障成因（如“由2023年焊接工艺缺陷引发”）；故障-人员关系方面，关联故障F-20240701与最初发现该故障的人员（如“巡检员李四通过轨检车监测数据首次发现该故障”）。上述语义关系的构建形成了直观的知识图谱网络，使运维人员可以通过数据目录语义检索相关信息。例如，只需查询“焊缝裂纹故障关联的人员”，系统即可通过图谱找到对应的巡检员和维修人员。该知识图谱数据目录不仅实现了数据的语义组织和展示^[16]，也为跨系统的数据关联查询提供了依据^[17]。另外，知识图谱引擎具备动态更新能力：当新增设备或人员时，相应节点和关系会自动加入图谱；故障处理完成后，图谱中故障节点状态随之更新。这样保证了知识图谱始终与最新元数据同步，维持数据目录的时效性和准确性。

3.4.3 跨源一体化查询:实时拉通多系统数据

跨源查询引擎为场景提供了统一的数据访问接口，使业务人员能够一键检索跨不同业务系统的数据。例如，在本场景中，查询引擎适配了多个异构接口类型：通过RESTful API获取工务设备管理系统中设

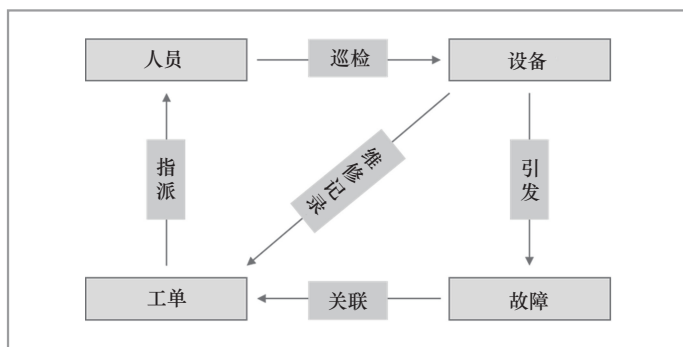


图3 实体关系建模示意图

备G-01的历史维修记录，通过消息队列（如Kafka）订阅轨道检测车监测平台的实时钢轨振动数据，通过直连数据库查询人力资源系统中巡检员王五的技能证书有效期，通过GraphQL接口获取工单管理系统中故障F-20240701的处置进度等。跨源查询引擎对上述不同接口类型进行了适配封装，屏蔽了底层技术差异，使用户无须关注数据来自何种系统、采用何种协议，只需要针对统一的虚拟数据视图编写查询。引擎接收到查询请求后，会自动拆分请求并分发到相应的数据源，实时汇聚结果返回用户。例如，调度人员在一条查询指令中即可同时获取指定故障的关联巡检记录、实时监测数据及工单处理状态等信息，极大提升了信息获取效率。这种知识驱动的跨源整合方法还可用于异构协议或数据格式的转换和互操作，实现不同系统间的数据渠道耦合^[18]。通过跨源一体化查询，该场景真正做到了各相关数据系统的无缝联通和信息即需即取，为及时准确地处置故障提供了保障。

3.4.4 技术协同工作流程与数据编织路径

图4所示为铁路故障处置场景中数据编织的协同流程。当监测到设备故障后，

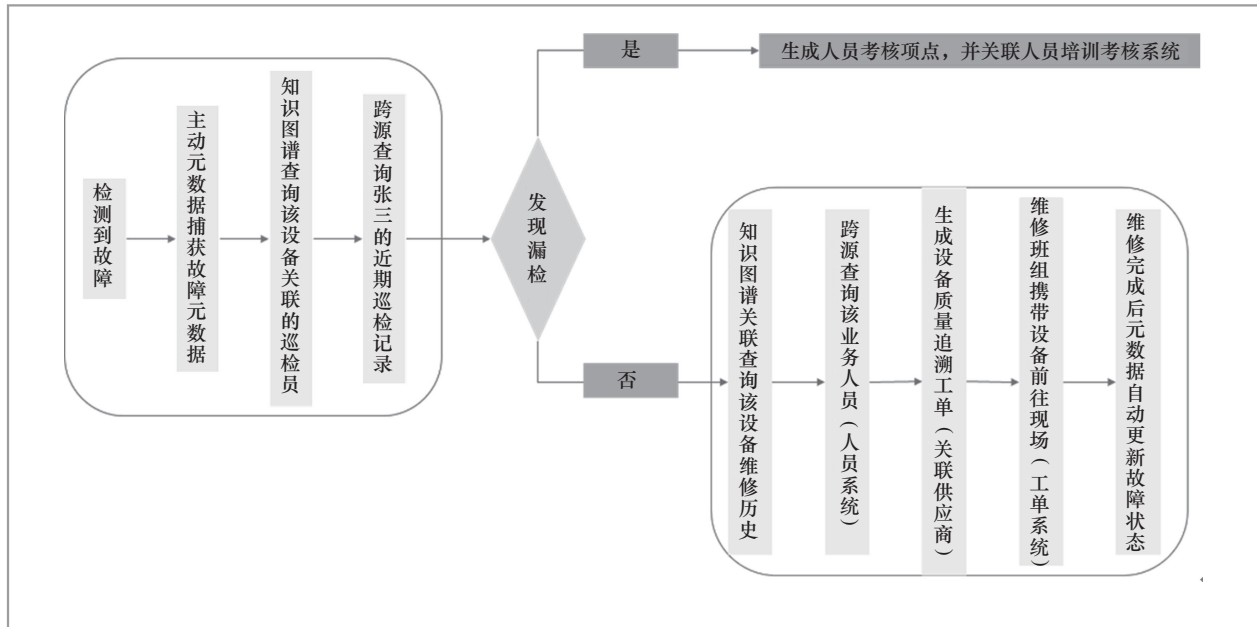


图4 技术协同工作流程与数据编织路径示意图

主动元数据引擎首先捕获记录该故障的元数据，并立即通过知识图谱引擎关联出负责该设备的巡检人员信息；随后，跨源查询引擎实时检索该巡检人员近期的巡检记录。如果查询结果发现存在“漏检”（即该设备区域内最近无巡检记录），则根据预定义规则触发元数据告警，同时自动生成一条补充巡检的工单任务指派给该人员，并通知其主管进行跟踪。如果未发现漏检情况，则系统通过知识图谱继续关联查询该设备的历史维修记录，结合跨源查询引擎调用人员系统获取相关业务人员（维修工程师）信息，生成一条设备质量追溯工单（关联可能的供应商质量问题），由维修班组通过工单系统领取，并前往现场处理。整个维修过程中，涉及的人员、设备状态、工单进展等信息不断通过主动元数据管理进行更新反馈：故障处理完成后，系统自动更新元数据仓库中该故障的状态为“已处置”，并触发知识图谱增量更新（故障节点与相关人员/设备的关系状态调整）。这

一系列自动化协同流程实现了“元数据刷新→图谱更新→一键查询→自动派单”的闭环处置。相较传统人工查询、人工派工的方式，该流程大大缩短了故障定位和响应时间，提高了处置效率和信息共享度。

3.4.5 可视化技术架构

图5所示为该场景下数据编织技术协同架构的示意图。底层的数据源层包括设备台账、实时监测数据、故障处置记录、人员信息等，源源不断产生原始数据并通过数据接入层汇聚到数据编织层。在数据编织层中，主动元数据引擎、知识图谱引擎和跨源查询引擎协同运作：其中元数据仓库直接连接主动元数据引擎，实现元数据的集中管理和驱动；语义查询解析模块对接知识图谱引擎，支撑语义级的智能检索；数据虚拟化组件支持跨源查询引擎，实现对异构数据的即时访问与融合^[19]。最终，处理后的数据和分析结果输送到用户

层，由业务人员获取和利用。

3项技术的深度协同实现了“人员-设备-故障”数据的动态关联与智能应用，解决了铁路传统业务中数据“孤岛”与协同效率低下的核心问题。与以往各系统割裂、人工协调的方式相比，本方案显著提升了故障根因分析和处置流程的效率。例如，在统一的数据编织平台支持下，运维人员可以在故障发生后第一时间获取关联的巡检记录和监测数据，快速判断故障原因并联动相关部门处置。综上，本应用场景验证了数据编织技术在铁路工务设备故障协同处置中的有效性，为后续在更大范围内推广数据编织提供了参考范式。

4 结束语

本文构建的铁路数据编织技术体系通过7层架构设计与三大关键技术协同应用，实现了从数据采集、语义建模到服务输出的全流程优化，为数据编织体系在铁路行业落地应用提供了可参考的技术范式。具体而言，7层架构清晰划分了数据接入、存储、处理、组织、服务到应用各环节的职能，使数据在每一层都得到针对性的整合与提升；从底层数据采集到顶层服务输出，各层有序衔接，显著简化了跨系统数据共享与业务协同流程。相比传统的“烟囱式”数据管理模式，本架构创新性地引入主动元数据驱动和知识图谱语义网，实现了数据语义统一和智能关联，消除了人工对接数据“孤岛”的烦琐步骤，尤其是通过知识图谱构建领域数据目录、主动元数据实时更新以及跨源查询按需取数等关键机制，打造了铁路数据管理从数据发现、治理到应用的闭环新模式，凸显了数据编织技术在铁路领域的创新应用价值。

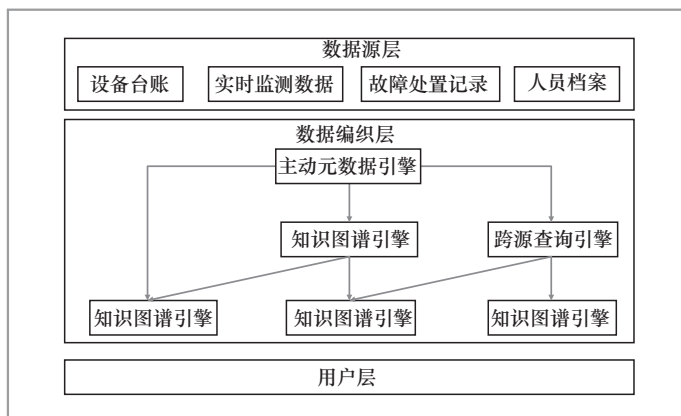


图5 可视化技术架构

面向未来，铁路数据编织技术可在3个方面深化发展。

- 技术融合：引入联邦学习与区块链技术，构建可信数据共享网络，解决跨单位数据协同的安全痛点^[20]。例如使用区块链确保跨部门数据交换的溯源和权限控制，通过联邦学习实现不同部门数据在本地训练下的协同智能。

- 场景拓展：向智慧货运领域延伸，基于货物轨迹数据编织实现运输路径动态优化、车载设备远程诊断等应用，提升铁路货运的智能化水平。

- 生态构建：与城市交通、物流等行业建立数据编织联盟，形成综合交通数据资产图谱，为国家智慧交通枢纽建设提供底层支撑，推动数据要素跨行业流通。

随着5G+物联网技术在铁路场景的全覆盖，数据编织技术将成为驱动“铁路大脑”建设的核心引擎，推动行业从“信息化”向“智慧化”的跨越式发展。

参考文献：

- [1] 数据编织提升大数据使用效率[J]. 电子质量, 2024(1): 57.

- Data weaving improves the efficiency of big data usage[J]. *Electronics Quality*, 2024(1): 57.
- [2] 齐艳平. 大数据时代数据编织方法及其对数据治理影响初探[J]. *中国国情国力*, 2022(1): 69-73.
Qi Y P. Data weaving method in big data era and its influence on data governance[J]. *China National Conditions and Strength*, 2022(1): 69-73.
- [3] Tang L. Genomics data integration[J]. *Nature Methods*, 2023, 20(1): 34.
- [4] 张连明, 王喆. 数据编织助力银行新一代数据云构建[J]. *中国金融电脑*, 2022(9): 66-69.
Zhang L M, Wang Z. Data fabric helps banks build a new generation of data cloud[J]. *Financial Computer of China*, 2022(9): 66-69.
- [5] Braga A J. Robust assessment of railway vehicle safety risks using data-driven approach[J]. *Reliability Engineering and System Safety*, 2024, 249: 110220.
- [6] 肖东君. 基于数据编织的两级政务数据治理实践[J]. *数字经济*, 2024(4): 54-58.
Xiao D J. Two-level government data governance practice based on data fabric[J]. *Digital Economy*, 2024(4): 54-58.
- [7] 刘北胜, 江若飞, 吕向茹, 等. 面向铁路基础设施全生命周期数据传递的通用数据模型研究[J]. *铁道标准设计*, 2025, 69(5): 28-34.
Liu B S, Jiang R F, Lyu X R, et al. Research on common data model for life-cycle data transmission in railway infrastructure[J]. *Railway Standard Design*, 2025, 69(5): 28-34.
- [8] 李新琴, 刘承亮, 代明睿, 等. 铁路运输设备故障知识图谱构建与应用研究[J]. *铁路计算机应用*, 2023, 32(1): 20-24.
Li X Q, Liu C L, Dai M R, et al. Construction and application of knowledge graph for railway transportation equipment fault[J]. *Railway Computer Application*, 2023, 32(1): 20-24.
- [9] 单馨漪, 郭梓焯, 官科, 等. 铁路通信数字化转型发展策略研究[J]. *铁道通信信号*, 2024, 60(9): 1-10.
Shan X Y, Guo Z Y, Guan K, et al. Research on digital transformation and development strategy in railway communications[J]. *Railway Signalling & Communication*, 2024, 60(9): 1-10.
- [10] 高珊, 袁宛竹, 卢卫, 等. 面向列语义识别的共现属性交互模型构建与优化[J]. *软件学报*, 2023, 34(3): 1010-1026.
Gao S, Yuan W Z, Lu W, et al. Construction and optimization of co-occurrence-attribute-interaction model for column semantic recognition[J]. *Journal of Software*, 2023, 34(3): 1010-1026.
- [11] 马友忠, 孟小峰. 云数据管理索引技术研究[J]. *软件学报*, 2015, 26(1): 145-166.
Ma Y Z, Meng X F. Research on indexing for cloud data management[J]. *Journal of Software*, 2015, 26(1): 145-166.
- [12] 蔡珉官, 王朋. 数据湖技术研究综述[J]. *计算机应用研究*, 2023, 40(12): 3529-3538.
Cai M G, Wang P. Survey of data lake technology research[J]. *Application Research of Computers*, 2023, 40(12): 3529-3538.
- [13] 马忠贵, 倪润宇, 余开航. 知识图谱的最新进展、关键技术和挑战[J]. *工程科学学报*, 2020, 42(10): 1254-1266.
Ma Z G, Ni R Y, Yu K H. Recent advances, key techniques and future challenges of knowledge graph[J]. *Chinese Journal of Engineering*, 2020, 42(10): 1254-1266.
- [14] 社会芳, 王昊奋, 史英慧, 等. 知识图谱多跳问答推理研究进展、挑战与展望[J]. *大数据*, 2021, 7(3): 60-79.
Du H F, Wang H F, Shi Y H, et al. Progress, challenges and research trends of reasoning in multi-hop knowledge graph-based question answering[J]. *Big Data Research*, 2021, 7(3): 60-79.
- [15] 江琳, 王煜, 李仕旺, 等. 基于知识图谱的12306智能运维应用研究[J]. *铁路计算机应用*, 2022, 31(12): 26-29.
Jiang L, Wang Y, Li S W, et al. Application of 12306 intelligent operation and

- maintenance based on knowledge graph[J]. Railway Computer Application, 2022, 31(12): 26-29.
- [16] 林映利, 李红辉, 张春, 等. 基于知识图谱的协议转换方法研究[J]. 集成技术, 2023, 12(1): 26-41.
Lin Y L, Li H H, Zhang C, et al. Research of protocol conversion based on knowledge graph[J]. Journal of Integration Technology, 2023, 12(1): 26-41.
- [17] 张丁荣, 王恪铭, 冯心妍. 基于知识图谱和故障树的高速铁路事故致因分析[J]. 铁路计算机应用, 2023(7): 14-18.
Zhang D R, Wang K M, Feng X Y. Cause analysis of high-speed railway accidents based on knowledge graph and fault tree [J]. Railway Computer Application, 2023(7): 14-18.
- [18] 韩铖山, 李红辉, 闫佳和, 等. 基于知识图谱的多源异构信息通道耦合技术研究[J]. 集成技术, 2023, 12(3): 48-60.
Han C S, Li H H, Yan J H, et al. Research on coupling technology of multi-source heterogeneous information channels based on knowledge graph[J]. Journal of Integration Technology, 2023, 12(3): 48-60.
- [19] Xu X J, Tatge L, Xu X L, et al. Blockchain applications in the supply chain management in German automotive industry[J]. Production Planning & Control, 2024, 35(9): 917-931.
- [20] 屈靖淇, 李红辉, 崔家昇, 等. 基于区块链的可溯源政务大数据共享方法研究[J]. 集成技术, 2023, 12(3): 19-33.
Qu J Q, Li H H, Cui J S, et al. Research on blockchain-based traceable government big data sharing method[J]. Journal of Integration Technology, 2023, 12(3): 19-33.

作者简介



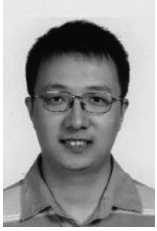
孙思齐 (1993-), 女, 中国铁道科学研究院助理研究员, 主要研究方向为铁路大数据治理。



邹丹 (1981-), 女, 中国铁道科学研究院高级工程师, 主要研究方向为铁路大数据治理。



张凯 (1993-), 男, 中国铁道科学研究院助理研究员, 主要研究方向为铁路大数据治理。



任爽（1979-），男，博士，北京交通大学计算机科学与技术学院副教授，主要研究方向为计算机科学与技术、区块链工程。



刘敏（1991-），女，中国铁道科学研究院高级工程师，主要研究方向为铁路大数据。

收稿日期: 2025-06-24

通信作者: 邹丹, 15350805159@163.com

基金项目: 中国铁道科学研究院集团有限公司科研专项项目(No. 2024YJ262)

Foundation Item: Special Scientific Research Project of China Academy of Railway Sciences Group Co., Ltd. (No. 2024YJ262)