

# 基于元数据与知识图谱的省级气象数据管理方法及应用

张平

山东省气象数据中心, 山东 济南

收稿日期: 2026年3月9日; 录用日期: 2026年4月10日; 发布日期: 2026年4月20日

## 摘要

针对省级气象业务中数据来源多样、系统分散、语义关联不足以及数据服务智能化水平不高等问题, 提出了一种基于元数据与知识图谱融合的省级气象数据知识化管理与智能应用方法。该方法以气象领域本体为统一语义模型, 通过对多源异构气象元数据的语义映射与融合, 构建省级气象知识图谱, 实现气象数据从“以存储管理为主”向“以知识组织与关联服务为主”的转变。在此基础上, 设计并实现了语义检索、关联发现和智能问答等应用场景。验证结果表明, 该方法能够有效打破省级气象业务系统间的数据壁垒, 显著提升复杂业务查询与信息获取效率, 可为省级气象数据资源的精细化管理和智能服务提供了可行技术路径参考。

## 关键词

气象数据管理, 元数据, 知识图谱, 语义关联, 数据服务

# Leveraging Metadata and Knowledge Graphs for Provincial Meteorological Data Management: Methods and Applications

Ping Zhang

Shandong Meteorological Data Center, Jinan Shandong

Received: March 9, 2026; accepted: April 10, 2026; published: April 20, 2026

## Abstract

Addressing issues in provincial-level meteorological operations such as diverse data sources,

fragmented systems, insufficient semantic associations, and a lack of intelligent data services, this paper proposes a method for knowledge-based management and intelligent application of provincial meteorological data, integrating metadata with knowledge graphs. Using a meteorological domain ontology as a unified semantic model, the method semantically maps and fuses multi-source heterogeneous meteorological metadata to construct a provincial meteorological knowledge graph. This shifts the focus of meteorological data from “storage-centric management” to “knowledge organization and association services.” On this basis, application scenarios including semantic search, association discovery, and intelligent Q&A are designed and implemented. Application verification results demonstrate that the method can effectively break down data silos between provincial meteorological business systems, significantly improve the efficiency of complex operational queries and information retrieval, and provide a feasible technical pathway for the refined management and intelligent services of provincial meteorological data resources.

## Keywords

Meteorological Data Management, Metadata, Knowledge Graph, Semantic Association, Data Service

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

### 1.1. 研究背景与意义

气象数据是支撑防灾减灾、农业生产、交通运行和气候变化研究的重要基础性信息资源。随着观测手段的不断丰富和业务体系的持续拓展，省级气象部门在长期业务运行过程中积累了来源多样、类型复杂、时效性强的海量气象数据。然而，受限于以业务系统或数据文件为中心的传统管理模式，数据往往分散存储于不同系统之中，缺乏统一的语义描述与关联机制，难以形成可计算、可推理的整体认知结构，制约了气象数据在综合分析 with 智能服务中的深度应用。

在当前智慧气象建设背景下，省级气象业务对数据管理提出了更高要求，不仅需要实现数据的统一汇聚与共享，更需要支持跨业务、跨类型的数据关联分析以及面向业务决策的智能服务。如何将分散的数据资源转化为可组织、可关联、可推理的知识资源，已成为省级气象数据治理亟需解决的关键问题。

### 1.2. 国内外研究现状

在气象数据管理方面，国际上世界气象组织(WMO)推动的 WIS 2.0 框架强调基于核心元数据标准实现气象数据的发现与交换[1]，美国、欧洲等已构建较为成熟的元数据目录服务体系。国内依托国家级气象大数据云平台“天擎”，初步实现了气象数据的统一汇集与编目管理，并形成了较为完整的气象数据元数据标准体系。然而，现有研究与实践多聚焦于数据汇聚与目录检索，元数据主要用于数据发现，尚未充分发挥其在数据语义关联和知识化管理中的作用。

近年来，知识图谱技术在气象领域逐步开展应用研究[2]，主要集中在气象灾害知识库构建、气象概念关系建模以及预报解释等方面。这类研究为气象知识表达与关联提供了有益探索，但多以局部知识或静态知识建模为主，缺乏与省级实际业务数据体系的深度融合，尚难支撑复杂业务场景下的智能查询与决策推理需求。

### 1.3. 现有问题与研究思路

现有研究多面向国家级平台或单一业务领域，缺乏面向省级多业务系统、以元数据为统一入口的知识化管理方法，难以在实际业务运行环境中形成可推广、可复用的整体解决方案。

综合分析现有研究和省级气象业务实践，当前数据管理主要面临以下问题：一是数据分散在不同业务系统中，缺乏统一的语义关联机制，形成事实上的“数据孤岛”；二是数据管理以存储和检索为主，数据背后的业务知识难以系统化组织和复用；三是数据服务方式仍以人工检索为主，难以满足复杂业务场景下对关联分析和智能决策支持的需求。

针对上述问题，本文引入元数据与知识图谱相结合的技术思路，构建面向省级气象业务的数据知识化管理与智能应用方法。通过将气象元数据由“目录描述”提升为“知识组织入口”，并利用知识图谱实现多源数据的语义关联和推理服务，以期能为提升省级气象数据管理效率和智能服务能力，探索可行路径。

## 2. 研究方法与技术路线

### 2.1. 方法总体思路

针对省级气象数据来源多样、业务关联复杂以及语义割裂严重等特点，本文提出一种基于元数据与知识图谱融合的气象数据知识化管理方法[3]。该方法以气象领域本体作为统一语义模型，将分散在不同业务系统中的多源异构气象元数据进行语义映射与融合，构建省级气象知识图谱，并在此基础上支撑语义检索、关联发现和智能问答等应用。

方法总体思路如图 1 所示，概括为“统一语义模型 - 多源元数据知识化 - 知识图谱构建 - 智能应用验证”四个环节，核心在于将传统元数据从“数据目录描述”提升为“知识组织与关联入口”，实现气象数据由静态存储向动态知识服务的转变。



Figure 1. Technical roadmap for provincial meteorological data knowledge management based on metadata and knowledge graphs  
图 1. 基于元数据与知识图谱的省级气象数据知识化管理技术路线

### 2.2. 气象领域本体模型构建

气象领域本体[4]是实现多源气象数据语义统一和知识关联的基础。结合省级气象业务特点，采用自顶向下与自底向上相结合的方法[5]构建气象领域本体模型。首先需求分析，梳理形成顶层概念框架；然后结合现有气象元数据标准和业务系统字段，对概念层级、属性和关系进行细化与补充；最后通过本体推理规则对模型进行一致性校验与优化。

拟构建的气象领域本体主要包括四个核心模块：(1) 时空实体模块：定义时间、空间位置及其约束关系，为观测数据、预报产品和服务事件提供统一的时空语义锚点。(2) 观测与数据模块：描述观测站、观测平台、气象要素及其质量属性，支持多类型观测数据的统一表达。(3) 预报与预警模块：定义预报产品、预警信号及其发布、演变关系，体现气象业务流程特征。(4) 服务与影响模块：刻画气象服务产品、服务对象及灾害影响关系，为业务应用和决策分析提供支撑。通过上述构建，基本实现了对省级气象数据关键业务语义的统一刻画，为后续多源数据知识化整合奠定基础。

### 2.3. 多源气象元数据的语义整合方法

在气象领域本体的约束下，对省级气象业务中多源异构元数据进行语义整合，将其转化为结构化、

可关联的知识单元。

### 2.3.1. 结构化元数据的知识化映射

对于来源于业务数据库和数据管理系统的结构化元数据,采用基于规则的映射方式实现知识化转换。具体包括:将业务数据表映射为本体中的概念类,将记录映射为实体实例,将字段映射为实体属性,并依据业务逻辑关系生成实体之间的语义关联。例如,通过站点编码与行政区划编码的关联,自动构建观测站与行政区划之间的“位于”关系。

采用 R2RML (RDB to RDF Mapping Language)作为结构化数据到知识图谱的标准映射语言,并利用 D2RQ 平台提供的映射引擎批量执行转换。针对省级气象业务中涉及的十余个关系型数据库(如观测数据库、预报产品库、预警信息库等),编写映射文件。映射过程通过 Python 脚本调度,每日增量同步更新至知识图谱。

### 2.3.2. 非结构化信息的知识抽取

针对气象服务材料、决策专报等非结构化文本数据,采用自然语言处理技术进行信息抽取,识别文本中的气象实体、时间实体和空间实体,并抽取实体间的关键关系与事件要素。

算法模型:采用基于 BERT-BiLSTM-CRF 的联合模型进行命名实体识别(NER),其中 BERT 提供上下文相关的字向量表示,BiLSTM 捕捉序列依赖,CRF 层保证标签的全局最优。关系抽取则使用 PCNN (Piecewise Convolutional Neural Networks)模型,结合注意力机制,从句子中抽取预定义的关系类型(如“发生于”、“造成”、“发布”等)。

训练数据:从省级气象历史服务材料中人工标注了 8000 条句子,涵盖 12 类实体(包括气象灾害、气象要素、时间表达式、地理位置、机构名称等)和 9 类关系。标注数据按照 8:1:1 划分为训练集、验证集和测试集。

效果:在测试集上,NER 模型的精确率、召回率和 F1 值分别达到 92.3%、90.7%和 91.5%;关系抽取的 F1 值为 85.6%。抽取结果通过与本体概念对齐,实现非结构化信息的结构化表达和知识化入库。

### 2.3.3. 知识融合与一致性控制

来自不同数据源的知识单元可能存在实体重复或信息冲突问题。通过基于标识符匹配与属性相似度计算的方法实现实体对齐,并结合数据来源权威性和时间优先级等策略进行冲突消解,保证知识图谱的一致性和可靠性。对于观测站点、行政区划等具有权威编码的实体,直接采用标识符匹配;对于气象灾害事件等缺乏统一编码的实体,计算属性(名称、时间、地点等)的编辑距离与 Jaccard 相似度,设定阈值判定是否为同一实体。以权威业务系统(如国家级下发数据)为优先,同一系统内以最新时间为准。根据不同业务系统特点进行配置与迭代调整,以适应业务中多源数据长期演化和差异化管理需求。

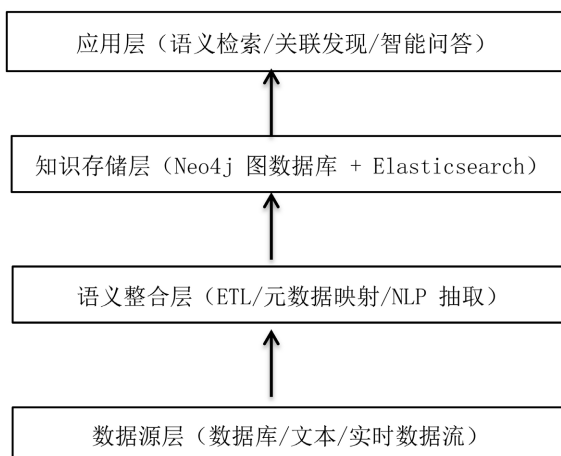
## 2.4. 省级气象知识图谱构建与更新机制

设计完成本体模型和知识抽取规则后,构建省级气象知识图谱,实现气象数据的集中存储与关联计算。知识图谱构建采用“本体先行、数据驱动、持续更新”[6]的方式:先导入气象领域本体,建立统一语义框架;再将历史气象数据和元数据批量转换为知识实体和关系,形成初始知识图谱;通过增量更新机制将新产生的数据持续融入图谱,实现知识的动态演化。

图数据库采用 Neo4j 作为知识图谱的存储引擎,主要基于其原生图存储、支持 ACID 事务、提供高效的图遍历查询语言 Cypher,并且具备良好的扩展性和可视化能力。针对实体及其关系,Neo4j 以节点和边的形式存储,并建立相应的索引以加速查询。为支持复杂关联查询和推理分析,结合图查询与规则推理机制,使用 Neo4j 的 APOC 库实现自定义路径遍历和规则推理,实现对多跳关联关系和业务逻辑规

则的高效处理。

系统技术架构整体分为四层：数据源层、语义整合层、知识存储层和应用层。数据源层：包含结构化数据库、非结构化文本文件及实时数据流，作为原始数据输入；语义整合层：通过 ETL 工具、NLP 抽取模块和映射引擎，将多源数据转换为结构化知识。知识存储层：以 Neo4j 图数据库为核心存储知识图谱，辅以 Elasticsearch 提供全文检索能力。应用层：基于 Spring Boot 构建 RESTful API，对外提供语义检索、关联发现和智能问答等服务。数据流向自下而上：数据源经语义整合后存入知识存储，最终由应用层接口服务调用。如图 2 所示。



**Figure 2.** Technical architecture of the provincial meteorological knowledge graph system

**图 2.** 省级气象知识图谱系统技术架构图

## 2.5. 技术路线总结

本文提出的方法以气象领域本体为核心[7]，通过多源气象元数据的语义映射与融合，构建省级气象知识图谱，并将其作为智能应用的统一数据与知识支撑。该方法既保持了现有气象元数据体系的规范性，又增强了数据之间的语义关联能力，为省级气象数据管理和智能服务提供了一种可复用、可扩展的技术路径。

## 3. 应用验证与分析

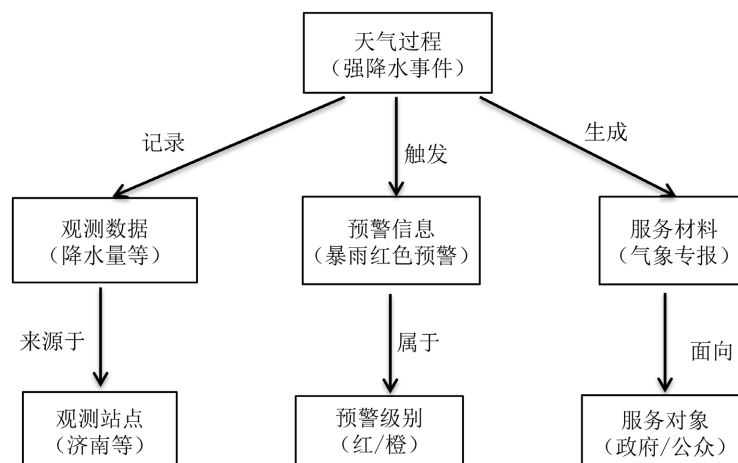
为验证所提出的基于元数据与知识图谱[8]的省级气象数据知识化管理方法的有效性，结合省级气象业务实际需求，设计并实现了多个典型应用场景，并通过对比分析验证其在数据关联、信息获取效率和智能服务能力等方面的改进效果[9]。

### 3.1. 语义关联检索应用验证

以一次具体的天气过程为中心，通过线条向四周辐射，连接起原本分散在不同业务系统中的观测数据、预报预警、服务材料等模块。如图 3 所示，该方法能解释知识图谱如何通过语义关系(比如“导致”、“发生在”、“记录了”)，把孤岛数据串联起来。

#### 3.1.1. 验证场景设计

在传统省级气象业务中，观测数据、预报产品和气象服务材料通常分散存储于不同业务系统，业务人员需分别检索并人工关联相关信息，效率较低且易遗漏。为验证知识图谱在语义关联检索方面的优势，选取典型天气过程作为验证对象，设计跨系统、多类型信息的一体化检索场景[10]。



**Figure 3.** Semantic association retrieval based on the provincial meteorological knowledge graph

**图 3.** 基于省级气象知识图谱的语义关联检索应用示意

### 3.1.2. 实现方式

基于构建的省级气象知识图谱，将与同一天气过程或同一空间区域相关的观测数据、预报产品、预警信息和服务材料通过语义关系进行统一关联。用户仅需输入天气过程名称或时间、区域等条件，系统即可返回相关信息的关联结果，实现“一次检索、全景获取”。

### 3.1.3. 验证结果分析

以一次区域性强降水过程为例进行验证。传统方式下，业务人员需分别在多个系统中以时间和区域为条件进行查询，并人工整理结果，整体耗时约 10~15 min。基于知识图谱的语义关联检索方式下，系统能够在数秒内返回该过程相关的观测实况、预警信号和服务产品等信息。为量化评估知识图谱的覆盖度，统计当前图谱中包含 2000 多个观测站点、180 种气象要素、40 类预报产品、12 类预警信号以及 3 万份历史服务材料，共形成 3.8 万条语义关系。对于该强降水过程，知识图谱通过“发生于”关系关联了 45 个受影响站点，通过“引发”关系关联了 3 条预警信号，通过“生成”关系关联了 8 份服务专报，实现了信息的全景关联。

对比结果表明，所提出的方法在复杂业务场景下显著提升了信息获取效率，并减少了人工操作负担，验证了其在打破数据孤岛和提升数据关联能力方面的有效性。

## 3.2. 关联发现与知识化管理能力验证

### 3.2.1. 验证场景设计

为验证系统在知识化管理和关联发现方面的能力，选取历史天气过程与实时观测数据相结合的业务场景，重点考察系统对相似案例发现和异常信息关联分析的支持能力。

### 3.2.2. 实现方式

在知识图谱中对天气过程、观测站点和气象要素等实体进行统一建模，并通过属性相似度和关联关系实现历史相似过程的自动匹配。当新的天气过程或异常观测值出现时，系统可自动关联历史数据和相关业务信息，为业务分析提供辅助支持。

### 3.2.3. 验证结果分析

验证结果表明，当新的强天气过程被纳入系统后，知识图谱能够基于路径、强度和影响区域等属性，

自动关联历史相似过程，并快速定位相关观测数据和服务材料。与传统依赖人工经验和手工查阅的方式相比，该方法显著缩短了历史案例分析时间(从平均 30 分钟缩短至 2 分钟)，提高了业务分析的系统性和完整性。

### 3.3. 智能问答与业务辅助应用验证

#### 3.3.1. 验证场景设计

针对气象业务中常见的复杂查询需求，设计基于自然语言输入的智能问答场景，验证系统对多源数据关联查询和简单业务推理的支持能力。

#### 3.3.2. 实现方式

通过对用户自然语言问题进行实体识别和意图解析，将问题映射为知识图谱查询条件，并在必要时结合预定义业务规则进行推理分析，生成查询结果或辅助决策建议。

#### 3.3.3. 验证结果分析

选取 100 条典型业务问题构建测试集，涵盖单跳查询(如“某站某时的温度”)、多跳查询(如“某次暴雨过程影响了哪些站点，发布了哪些预警?”)以及带约束的查询(如“过去一周降水量最大的三个站点”)。系统对问题意图和实体的识别准确率为 91%，查询结果正确率为 88%。与传统基于关键词检索的方式相比，该方法在多跳查询场景下具有更高的准确性和更好的用户体验，初步验证了其在提升气象数据服务智能化水平方面的应用价值。

此外，针对多跳关联查询的耗时对比实验表明：在传统关系型数据库(MySQL)中，完成一次三跳关联查询(如查询某站点某次过程的观测数据、关联的预警信息以及该预警的服务材料)平均耗时约 4.8 秒(需多次 join 操作)；而基于 Neo4j 的图查询使用 Cypher 语句仅需 0.3 秒，查询效率提升 16 倍。

## 4. 结论与讨论

### 4.1. 结论

针对省级气象数据管理中存在的数据分散、语义关联不足以及数据服务智能化水平不高等问题，本文提出了一种基于元数据与知识图谱融合的省级气象数据知识化管理与智能应用方法，并结合省级气象业务场景开展了应用设计和验证。主要结论如下：

(1) 构建了以气象领域本体为核心的省级气象数据知识化管理方法。

通过引入统一的气象领域本体模型，对多源异构气象元数据进行语义映射与融合，将传统以目录描述为主的元数据体系扩展为支撑知识组织与语义关联的统一语义基础，实现了省级气象数据的知识化表达。

(2) 实现了多源气象数据的语义关联与知识化管理。

基于构建的省级气象知识图谱，有效整合了观测数据、预报产品、预警信息和气象服务材料等多类型数据资源，打破了业务系统之间的数据壁垒，显著提升了复杂业务场景下的数据关联发现与综合分析能力。

(3) 验证了该方法在提升数据服务效率和智能化水平方面的应用效果。

通过语义关联检索、关联发现和智能问答等应用场景的验证，结果表明，该方法能够在省级气象业务中有效缩短信息获取时间，降低人工操作负担，为复杂业务查询和辅助决策提供了更高效的数据服务方式。

综合来看，本文提出的方法为气象数据从“以存储管理为主”向“以知识组织与智能服务为主”的

转变提供了可行的技术路径，对提升气象数据资源利用效率和业务支撑能力具有一定参考价值。

## 4.2. 讨论

本文方法仍存在一定局限性，有待在后续研究中进一步完善。首先，气象领域本体模型的覆盖范围仍需持续扩展。本文构建的本体主要聚焦于观测、预报、预警和服务等核心业务领域，对于农业气象、能源气象和气候服务等专业领域的细化概念与关系仍有待补充，未来需在更广泛的业务实践中不断迭代完善。其次，知识图谱对实时数据流的处理能力仍有提升空间。当前方法主要面向分钟级及以上时间尺度的数据更新，对于高频实时观测数据的流式知识化处理与即时推理能力尚需进一步研究。而且，智能问答和业务推理能力在复杂场景下仍依赖于规则完备性与语义解析精度。后续可结合更先进技术，进一步提升系统对复杂、模糊查询的理解与应对能力。

## 参考文献

- [1] World Meteorological Organization (2020) WMO Information System (WIS) 2.0 Implementation Plan. WMO.
- [2] Noy, N.F. and McGuinness, D.L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory.
- [3] 王琪, 王艺, 陈凯华, 等. 基于 Neo4j 的气象数据治理知识图谱构建与应用[J]. 微型电脑应用, 2025, 41(2): 78-83.
- [4] 王平, 谢雨霏, 端文卓, 等. 基于本体的气象记录档案知识图谱构建与应用研究[J]. 档案学研究, 2024(3): 94-103.
- [5] 张欢, 易善楨, 冷创, 等. 基于本体和图数据库的流域干旱灾害风险知识图谱构建及应用[J]. 水电能源科学, 2024, 42(5): 24-28.
- [6] Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G.D., Gutierrez, C., et al. (2021) Knowledge Graphs. *ACM Computing Surveys*, **54**, 1-37. <https://doi.org/10.1145/3447772>
- [7] Ehrlinger, L. and Wöß, W. (2016) Towards a Definition of Knowledge Graphs. *Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems (SEMANTiCS 2016)*, Leipzig, 13-14 September 2016, 13-16.
- [8] 邱明慧, 谢能付, 姜丽华, 等. 农业气象灾害知识图谱构建研究进展[J]. 中国农业气象, 2024, 45(10): 1216-1235.
- [9] 王益鹏, 张雪英, 党玉龙, 等. 顾及时空过程的台风灾害事件知识图谱表示方法[J]. 地球信息科学学报, 2023, 25(6): 1228-1239.
- [10] 向历霓, 李刚, 李海江. 基于知识图谱和 GPT 模型的可靠性代码自动生成方法[J]. 计算力学学报, 2024, 41(2): 217-225.