

Article

Interoperable Knowledge Graphs for Localized Supply Chains: Leveraging Graph Databases and RDF Standards

Vishnu Kumar 

Department of Industrial and Systems Engineering, Morgan State University, Baltimore, MD 21251, USA;
vishnu.kumar@morgan.edu

Abstract

Background: Ongoing challenges such as geopolitical conflicts, trade disruptions, economic sanctions, and political instability have underscored the urgent need for large manufacturing enterprises to improve resilience and reduce dependence on global supply chains. Integrating regional and local Small- and Medium-Sized Enterprises (SMEs) has been proposed as a strategic approach to enhance supply chain localization, yet barriers such as limited visibility, qualification hurdles, and integration difficulties persist. **Methods:** This study proposes a comprehensive knowledge graph driven framework for representing and discovering SMEs, implemented as a proof-of-concept in the U.S. BioPharma sector. The framework constructs a curated knowledge graph in Neo4j, converts it to Resource Description Framework (RDF) format, and aligns it with the Schema.org vocabulary to enable semantic interoperability and enhance the discoverability of SMEs. **Results:** The developed knowledge graph, consisting of 488 nodes and 11,520 edges, enabled accurate multi-hop SME discovery with query response times under 10 milliseconds. RDF serialization produced 16,086 triples, validated across platforms to confirm interoperability and semantic consistency. **Conclusions:** The proposed framework provides a scalable, adaptable, and generalizable solution for SME discovery and supply chain localization, offering a practical pathway to strengthen resilience in diverse manufacturing industries.

Keywords: supply chain resilience; knowledge graphs; graph database; RDF; schema.org



Academic Editor: Rob Handfield

Received: 2 September 2025

Revised: 9 October 2025

Accepted: 10 October 2025

Published: 13 October 2025

Citation: Kumar, V. Interoperable Knowledge Graphs for Localized Supply Chains: Leveraging Graph Databases and RDF Standards.

Logistics **2025**, *9*, 144.

<https://doi.org/10.3390/logistics9040144>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent geopolitical conflicts, trade disruptions, and economic volatility have exposed hidden vulnerabilities and intensified longstanding challenges within global supply chains. In response, many U.S.-based manufacturers have increasingly adopted reshoring strategies, relocating production and services to the domestic market to reduce exposure to external risks and enhance supply chain resilience [1]. However, this strategic shift has introduced new operational challenges: large enterprises must now identify and collaborate with regional partners in a timely manner to construct agile and reliable supply networks without compromising quality, responsiveness, or cost-efficiency [2]. In this context, Small- and Medium-Sized Enterprises (SMEs), which are regarded as the backbone of the American economy, are increasingly recognized as essential contributors to the reshoring effort [3]. Their agility, niche expertise, and geographic proximity make them ideal contributors to regionalized supply chains. Despite this, two persistent barriers have limited their integration: (i) Large enterprises often lack structured, searchable data on the capabilities and availability of local SMEs [2], and (ii) Many SMEs lack the visibility and digital infrastructure needed to participate in complex supply networks [4]. These challenges have

hindered effective supplier discovery and slowed the localization of manufacturing and supply chain ecosystems.

Although recent research has advanced understanding of supply chain resilience and digitalization, most efforts have focused on large firms, global optimization, or simulation-based models [5]. Few studies have developed practical, semantically enriched digital frameworks that make SMEs discoverable and interoperable within regional supply networks. Current approaches are typically limited to static supplier databases or survey-driven assessments, which lack the capacity to capture complex inter-enterprise relationships or to enable cross-platform integration. This leaves an urgent gap for scalable, machine-interpretable solutions that directly support SME integration into reshoring initiatives.

To bridge this gap, this work proposes a knowledge graph-based framework for representing SMEs and their inter-enterprise relationships within regional supply chains. In this framework, supply chain entities are modeled using a graph database and semantically enriched to form a knowledge graph, enabling machine-interpretable representation, relationship mapping, and contextual search functionality. The semantic layer facilitates structured querying, dynamic information retrieval, and improved interoperability across digital platforms. A proof-of-concept implementation was then conducted using enterprise data from the BioPharma sector in Pennsylvania (PA), a state in the U.S. with a high density of manufacturing SMEs [6]. A dynamic search mechanism was developed, enabling multi-conditional, human-readable queries that support efficient exploration of complex supply chain relationships. These capabilities were shown to outperform conventional keyword-based search tools in both speed and precision. To ensure broader applicability, the knowledge graph was converted into the Resource Description Framework (RDF), a World Wide Web Consortium (W3C) standard, and further aligned with Schema.org vocabulary to enhance semantic consistency, interoperability and enable integration with web-based systems [7].

In essence, this work addresses the following research gaps: (i) the absence of structured, semantically enriched representations of regional SMEs; (ii) limited mechanisms for querying complex inter-enterprise relationships; and (iii) a lack of interoperable frameworks to support data sharing across systems.

Accordingly, this study is guided by two research questions (RQ):

RQ1: How can SMEs and their relationships within a regional supply chain be effectively modeled using knowledge graphs to enhance dynamic information retrieval and discoverability?

RQ2: In what ways can semantic web technologies (such as RDF and Schema.org) improve the interoperability of SME data?

To investigate RQ1, a graph-based query mechanism was designed and implemented to support flexible, context-aware search for regional SME capabilities and connections. To address RQ2, semantic interoperability was validated through RDF serialization, Schema.org alignment, and cross-platform knowledge sharing scenarios.

The remainder of this paper is structured as follows: Section 2 reviews related work on supply chain resilience, graph modeling, and knowledge representation. Section 3 presents the methodology for knowledge graph construction, including data modeling, querying, validation, and implementation using Pennsylvania BioPharma data. Section 4 extends this methodological framework by addressing semantic interoperability: covering RDF serialization, Schema.org alignment, and cross-platform validation. Section 5 discusses key findings, implications, advantages and limitations. Section 6 concludes with the main contributions and directions for future research.

2. Related Work and Motivation

Globalization and lean manufacturing have transformed supply chains into efficient yet vulnerable systems. To minimize costs and enhance responsiveness, firms increasingly rely on globally dispersed suppliers and maintain minimal inventories, making them highly

susceptible to disruptions from natural disasters, geopolitical events, and pandemics [8]. Such disruptions often propagate across interconnected networks, amplifying systemic risks [9–11]. Limited visibility into upstream and downstream components further hampers timely response and recovery. Consequently, supply chain resilience, defined as the ability to resist, absorb, and recover from disruptions while maintaining essential functions, has emerged as a critical strategic capability [9–11]. While strategies such as supplier diversification, localization, redundancy, and enhanced visibility have been proposed [12–14], much of the literature remains theoretical or descriptive, with few scalable or empirically validated implementations [5].

To capture the systemic nature of supply chains, researchers increasingly model them as complex networks, where entities are connected via material, information, or financial flows [15]. Graph-based models allow representation of inter-dependencies, simulation of disruptions, and quantification of resilience. Approaches such as centrality analysis [16,17], agent-based modeling [18], and shock propagation [19] reveal vulnerabilities but remain largely conceptual, focusing on large enterprises while excluding small- and medium-sized enterprises (SMEs). This exclusion is problematic given SMEs' growing role in resilient, regionalized manufacturing ecosystems.

SMEs are critical contributors to supply chain resilience, especially under the shift toward modular and platform-based manufacturing [3,12,20–22]. However, SMEs often remain invisible in global sourcing platforms due to limited discoverability, outdated data, and lack of standardized representation. Platforms such as Wikidata [23], Thomasnet [24], Mergent Intellect [25], and Dun & Bradstreet [26] offer centralized repositories but suffer from incomplete data, poor update mechanisms, and limited query expressiveness. Wikidata requires SPARQL expertise, while commercial directories such as Thomasnet rely on keyword search, often yielding imprecise results. These barriers highlight the need for scalable, semantically rich frameworks that enhance SME visibility, enable dynamic querying, and support cross-platform integration.

Knowledge graphs offer a promising solution by enabling semantic modeling of enterprise capabilities, relationships, and context. They support interoperability with standards such as RDF and Schema.org and allow multi-conditional querying that can surface hidden connections. Recent research has begun to explore knowledge graphs in supply chains, but most contributions remain partial or domain generic. For example, Zhou et al. [27] proposed a supply chain knowledge graph for market convergence prediction, employing representation learning to infer macro-level trends. While methodologically sound, their work remains focused on high-level markets and does not address regional implementation, SME discoverability, or semantic web alignment. Other studies have leveraged zero-shot frameworks combining knowledge graphs with large language models (LLMs) to map complex, multi-tiered dependencies in electric vehicle supply chains [28]. While these systems map multi-tiered dependencies and support intelligent querying using language models, they do not incorporate RDF or Schema.org, and the design assumes access to extensive enterprise data, limiting their utility for SMEs.

Düggelin and Laurenzi [29] developed an ontology-driven decision support system for risk analysis using knowledge graphs, validated through domain expert review. While demonstrating strong knowledge modeling practices, their approach focuses on organizational-level risk rather than SME integration or semantic interoperability. Tu et al. [30] applied graph neural networks (GNNs) to a large-scale automotive supply chain knowledge graph for supplier recommendation. Their system showed superior predictive accuracy compared to non-graph methods, but it does not address dynamic querying, RDF compliance, or regional contexts, and SMEs remain underrepresented. Karam et al. [31] proposed a hybrid framework combining knowledge graphs and Bayesian networks for

bottleneck detection. While novel in combining structural and probabilistic models, it lacks a schema for supplier discovery or semantic querying.

Kosasih et al. [32] advanced the field by applying GNNs for risk propagation and reasoning over supply chain knowledge graphs. They encode complex inter-firm relationships and use GNN-based link prediction to surface hidden vulnerabilities. However, their system does not address how such knowledge graphs are constructed, semantically modeled, or deployed in web-based regional applications. Similarly, the distant supervision method proposed by Huang et al. [33] focuses on automated knowledge extraction for supply chain knowledge graphs, using natural language processing (NLP) and relation extraction to build graphs from text. While critical for scalability, the method does not extend to RDF or Schema.org alignment, nor is it embedded in an end-user system for SME discovery or regional graph querying.

Several recent works leave gaps relevant to SME-focused supply chain resilience. Rolli et al. [34] introduced an axiomatic RDF-based knowledge graph for Industry 4.0, but their work remains conceptual and lacks SME-focused implementation and supplier discovery features. Saidi et al. [35] developed reconfigurable supply chain models leveraging semantic technologies, but these remain largely theoretical and do not address SME integration. Similarly, Di Pierro [36] proposed ontology-enriched graph databases without validation in real supply chain applications. McEachen and Lewis [37] advanced spatial knowledge interoperability, and Thakker [38] investigated RDF–property graph translation, but both contributions are outside the scope of SME supply chains.

Earlier foundational works such as Shadkam [39], Petersen et al. [40], and Grangel-González [41] established semantic integration approaches in supplier selection and Industry 4.0 contexts. Complementing these, Petersen et al. [42] demonstrated an RDF-based information model for manufacturing companies, showing how enterprise and asset data can be exposed via RDF vocabularies and queried using SPARQL. Ameri et al. [43] applied an ontological RDF in agri-food supply chains to support traceability, validating semantic modeling and query execution in real-world contexts. Similarly, Petersen et al. [44] developed a federated, semantics-based supply chain analytics framework using SCORVoc RDF vocabularies and distributed SPARQL queries. While these studies provide concrete industrial applications of RDF in supply chain contexts, they remain disconnected from dynamic web deployment, Schema.org compliance, and SME-focused supplier discovery, motivating the distinct contributions of this work.

To systematically compare these contributions, Table 1 summarizes their core features, focus areas, and limitations relative to the proposed framework. The format of Table 1 is adapted from [45,46] to clearly present a comparative analysis of prior works and highlight the novelty of this study. This analysis reveals persistent gaps, notably in SME integration, dynamic querying capabilities, semantic modeling, and real-world application. Most prior works remain conceptual, lack SME focus, and do not adopt dynamic querying or semantic modeling standards such as RDF or Schema.org. Furthermore, they seldom demonstrate real-world deployment or validation, limiting their practical impact. The proposed research addresses these gaps by integrating semantic structuring, dynamic querying, SME-focused design, and real-world application within a domain-specific context.

Table 1. Comparative analysis of related works.

Study	Knowledge Graph Use	SME Focus	Querying Capability	Semantic Modeling	Real-World Application
Rolli et al. (2025) [34]	Yes	No	No	Yes	No
Saidi et al. (2025) [35]	Yes	No	No	Yes	No
Di Pierro (2025) [36]	Yes	No	No	Yes	No

Table 1. *Cont.*

Study	Knowledge Graph Use	SME Focus	Querying Capability	Semantic Modeling	Real-World Application
Zhou et al. (2024) [27]	Yes	No	Limited	Partial	No
AlMahri et al. (2024) [28]	Yes	No	Yes	No	No
Düggelin & Laurenzi (2024) [29]	Yes	No	No	Yes	No
Tu et al. (2024) [30]	Yes	No	Yes	No	Yes
Kosasih et al. (2024) [32]	Yes	No	Limited	No	No
Huang et al. (2024) [33]	Yes	No	Limited	No	No
Karam et al. (2023) [31]	Yes	No	Limited	Partial	No
McEachen et al. (2023) [37]	Yes	No	No	Yes	No
Ameri et al. (2022) [43]	Yes	No	No	Yes	Yes
Thakker (2021) [38]	Yes	No	No	Yes	No
Shadkam (2021) [39]	Yes	No	No	No	No
Grangel-González (2019) [41]	Yes	No	No	Yes	No
Petersen et al. (2017) [42]	Yes	No	Yes	Yes	Yes
Petersen et al. (2016) [44]	Yes	No	Yes	Yes	Yes
This Work	Yes	Yes	Yes	Yes	Yes

3. Knowledge Graphs for Supply Chain Modeling

Modern manufacturing supply chains are complex, multi-tiered ecosystems involving various types of entities, including suppliers, contract manufacturers, logistics providers, and testing centers. These entities are connected by inter-dependent material, information, and financial flows that form dense, dynamic networks. Traditional supply chain models based on linear, tabular representations (e.g., relational databases) are ill-suited to capture such complexity, especially when modeling relationships that are multi-directional, indirect, or conditional [47].

In contrast, graph-based representations provide a more natural and flexible alternative [15]. In this approach, entities such as companies (e.g., Company A, Company B) are represented as nodes, and their relationships (e.g., “Company A supplies Company B” or “Company B works with Company A”) are represented as edges. The graph-based structures can be stored and managed using a specific type of database called a graph database. Unlike traditional relational databases, graph databases natively store relationships and connections between entities alongside data, enabling seamless querying and manipulation [47]. To maximize the utility of graph structures for decision-making, they can be enriched with domain knowledge or semantics, thereby creating a knowledge graph [47]. For example, Figure 1a illustrates a simple graph with two connected nodes, while Figure 1b shows the corresponding knowledge graph where the nodes (labeled as Company A and Company B) are enriched with additional attributes such as company size, name, location, and relationship details like the quantity of materials supplied from A to B. This enriched representation not only captures the existence of relationships but also their meaning and context. A formal mathematical definition and example of the knowledge graph structure used in this work are provided in Appendix A.

3.1. Knowledge Graph Implementation for SME Supply Chains

This section presents a step-by-step methodology for constructing a domain-specific knowledge graph designed to enhance visibility and partner discovery within regional SME supply chains, following the approach of [48]. The process comprises three main stages:

- (1) **Data Collection and Preprocessing:** Gathering enterprise-level data relevant to the target sector, followed by cleaning, structuring, and formatting for knowledge graph construction.

- (2) Knowledge Graph Schema Design: Importing processed data into Neo4j graph database and defining nodes, edges and categories that structure the knowledge graph.
- (3) Search Mechanism Design: Developing a dynamic, human-interpretable search mechanism to enable enterprise scouting and discovery within the knowledge graph.

The proof-of-concept implementation focuses on Pennsylvania's BioPharma sector but is generalizable to other industries. Figure 2 presents a pictorial flowchart illustrating this methodology. The following sections describe each stage in detail.

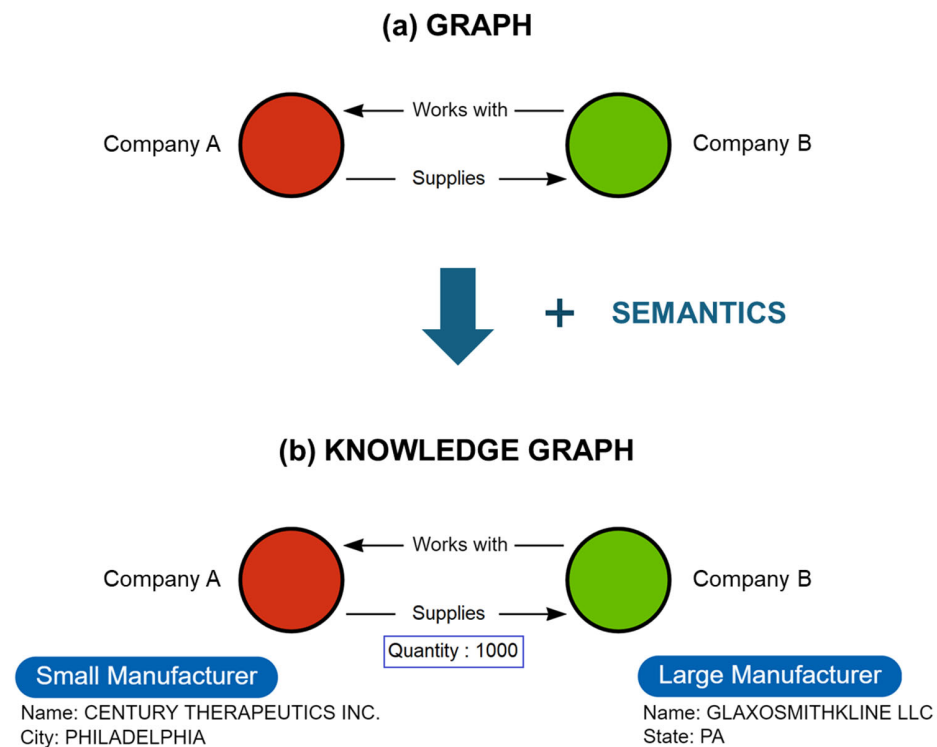


Figure 1. (a) Basic graph representation with nodes and edges showing entity connections; (b) Graph enriched with semantic information such as entity attributes (Name, City) and relationship properties (Quantity), resulting in a knowledge graph.

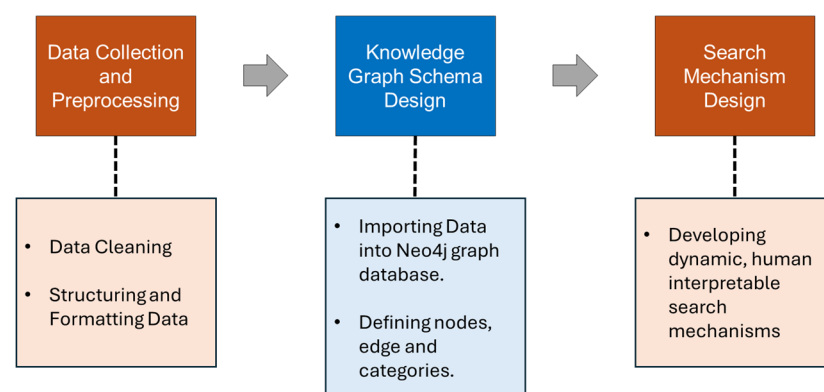


Figure 2. Methodological framework for constructing SME supply chain knowledge graph: data collection and preprocessing, knowledge graph schema design, and search mechanism for enterprise scouting.

3.1.1. Data Collection and Preprocessing

Enterprise-level data were collected from Mergent Intellect [25], a commercial data platform that aggregates structured business information from Dun & Bradstreet. The dataset included 488 enterprises located in Pennsylvania and categorized under North

American Industry Classification System (NAICS) Code 3254 (Biopharmaceuticals). This NAICS code classifies companies based on their primary business activities related to biopharmaceutical manufacturing, ensuring relevance to the sector. For each enterprise, attributes such as company name, location, employee count, sales volume, and business function were extracted. The data were cleansed by removing duplicates and incomplete records, then formatted into a tabular structure for knowledge graph construction.

3.1.2. Knowledge Graph Schema Design

The cleaned dataset was imported into the Neo4j graph database platform to develop the knowledge graph. Once the data were imported, enterprises were modeled as nodes, while their relationships were modeled as edges using Neo4j graph database. The companies are categorized into 8 groups based on the products made/service rendered, namely, Large Scale Manufacturer ($n = 45$), and SMEs including: Small Manufacturer ($n = 165$), Contract Manufacturer ($n = 34$), Material Supplier ($n = 52$), R&D Company ($n = 96$), Equipment Supplier ($n = 10$), Logistics and Supply ($n = 64$) and Testing Centers ($n = 22$). After node creation, edges labeled “Works With” were established between companies located within the same or neighboring regions, identified via their zip codes and sector relevance. This iterative process resulted in a complete knowledge graph representing Pennsylvania’s BioPharma manufacturing and supply chain ecosystem. The final graph consists of 488 nodes and 11,520 edges. Figure 3 illustrates a partial view of this knowledge graph.

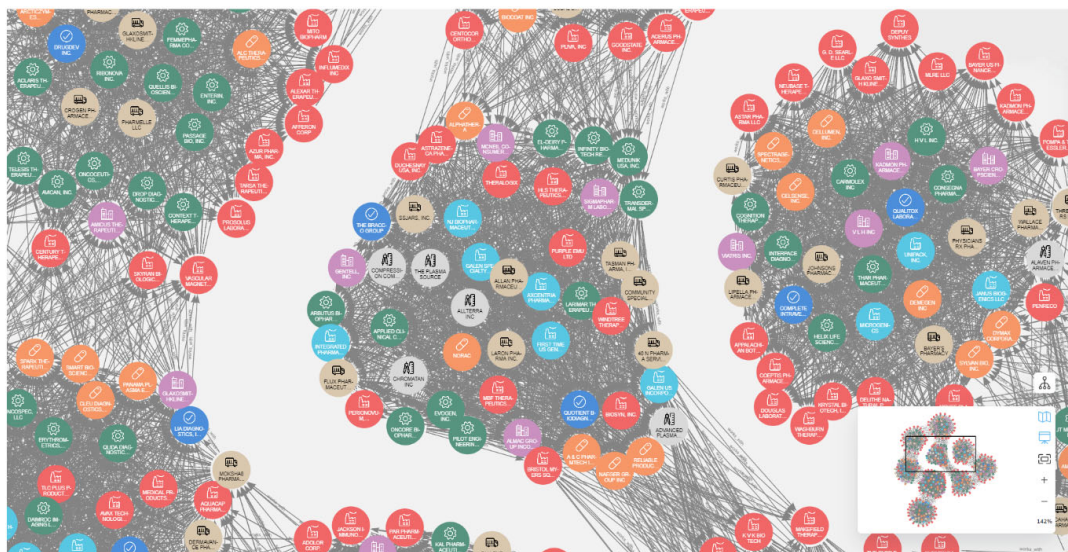


Figure 3. Partial view of the Knowledge Graph for PA enterprises in BioPharma Sector.

Unlike existing repositories that focus on isolated enterprise records, the approach presented in this work models SMEs within their broader relational and geographic ecosystem, enabling both direct and multi-hop discovery use cases essential for supply chain resilience.

3.1.3. Search Mechanism Design and Use Case Scenarios

Effective enterprise scouting, particularly for SMEs, is critical to supply chain diversification, risk mitigation, and regional economic development. However, traditional search tools often fail to uncover indirect partnerships or alternative suppliers, limiting their value for strategic sourcing and resilience planning. By integrating entity attributes with inter-firm relationships in a unified structure, the constructed knowledge graph enables dynamic, relationship-driven search capabilities. To assess its practical utility, this work examined both simple lookups and complex, multi-conditional queries. The simple queries reflect fundamental supplier discovery tasks, such as identifying firms by location or organizational

type, which are widely used for initial screening. In contrast, the multi-conditional queries capture complex decision-making situations where firms must evaluate multiple criteria simultaneously, for example, balancing specialization, geographic proximity, and collaborative ties. These types of tasks are emphasized in the supplier selection literature [49] and in studies on supply chain resilience and adaptation under disruption [50], confirming that the chosen scenarios are representative of real-world supply chain challenges.

(a) Simple Relationship-Based Search

The knowledge graph built using the Neo4j graph database allows intuitive exploration of companies and their direct connections using human language style search prompts. For example, a user searching for all Logistics and Supply enterprises can retrieve matching companies and their relationships by selecting category filters, as shown in Figure 4. Figure 4a shows search options to find and filter enterprises by predefined categories such as Contract Manufacturing, Logistics and Supply, R&D, etc. These filters are based on tagged metadata stored in the knowledge graph and appear in the interface as dropdown menus or searchable text fields. Selecting a category displays all enterprises matching that classification. Figure 4b shows relationship-based prompts to locate, for instance, Logistics and Supply enterprises and explore their connections to other enterprise categories. When activated, this option displays Logistics and Supply enterprises along with their first-degree (direct) connections to other companies.

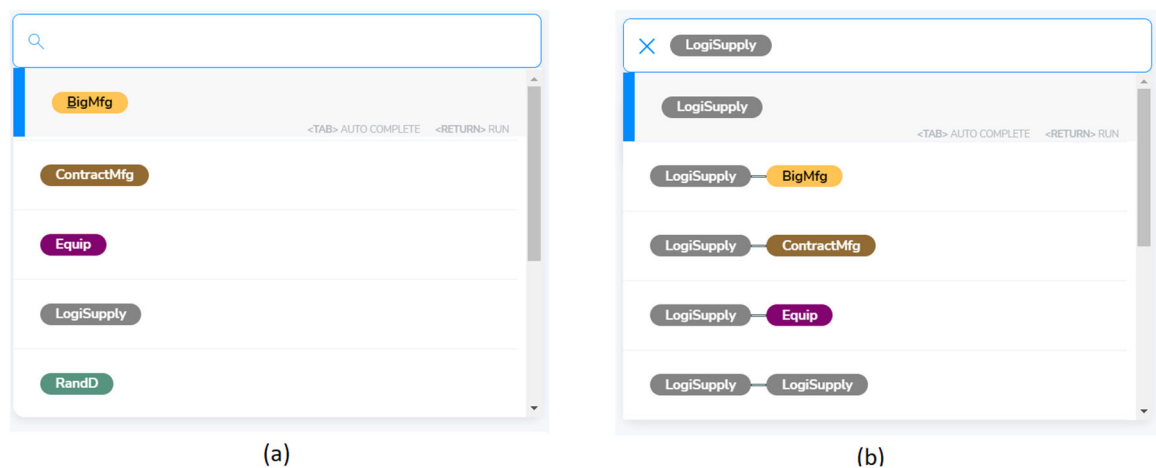


Figure 4. Simple search using built-in search prompts in Neo4j graph database: (a) Interface for searching enterprises by category (e.g., BigMfg, ContractMfg); (b) Interface for locating Logistics and Supply enterprises and exploring their relationships with other categories such as BigMfg and ContractMfg.

Users can further refine results by selecting individual enterprises to reveal direct relationships and neighboring entities within the network (Figure 5). In Figure 5, the selected enterprise node (ENDO PHARMACEUTICAL, INC.) expands to show all directly connected entities, such as raw material suppliers, contract manufacturers, or other related companies. The visual layout helps users quickly understand the local network context of a given enterprise, offering insights into the company's position within the regional supply chain. For example, a Logistics and Supply enterprise connected to many other entities may indicate a central role in distribution or supply coordination, while a manufacturer with numerous connections could represent a key production hub within the regional network. While this supports quick exploratory analysis, more detailed or conditional discovery requires advanced queries, discussed in the following section.

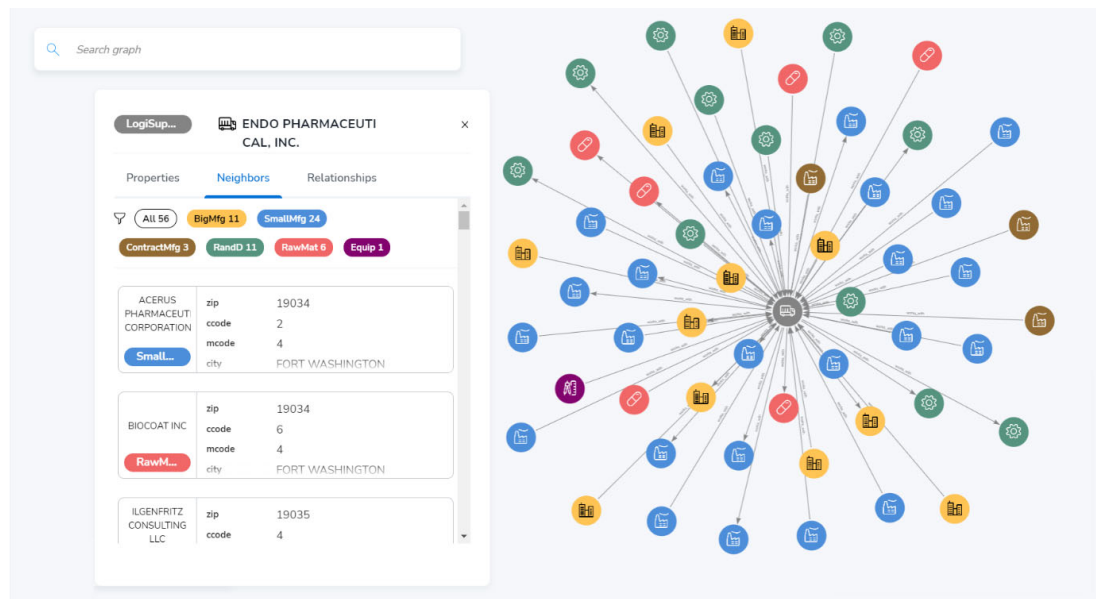


Figure 5. Simple search results showing the enterprise ENDO PHARMACEUTICAL, INC. (gray node at the center) and its immediate neighboring connections. Colored nodes represent different enterprise categories, such as BigMfg and ContractMfg.

(b) Enterprise Scouting Using Complex Multi-Conditional Queries

Beyond simple searches, the knowledge graph built on the Neo4j graph database supports advanced multi-conditional queries that traverse multiple relationships to discover highly specific enterprise connections. These complex queries are essential in realistic supply chain scenarios where users seek entities based on combinations of attributes and indirect relationships. Such queries, expressed in human-readable templates, can be dynamically adjusted by the user to explore alternative conditions or partners.

Scenario 1: Identifying a Supplier's Supplier

Consider a large enterprise seeking to collaborate with a Pennsylvania-based contract manufacturer who meets both of the following conditions: (i) the contract manufacturer sources raw materials from GNOSIS USA INC., and (ii) the contract manufacturer works with another large manufacturer, STI PHARMA LLC. Such a scenario reflects a realistic due diligence or strategic sourcing situation where enterprises seek partners that are already embedded in a reliable supply network. Traditional supply chain platforms rarely support such multi-hop queries across indirect relationships. In contrast, the knowledge graph efficiently executes this multi-conditional query, instantly returning the set of qualifying enterprises. Figure 6 shows the results of the query: "Contract Manufacturers that get raw materials from supplier GNOSIS USA INC. and work with STI PHARMA LLC". This type of query supports transparency, reveals non-obvious connections, and enhances strategic supplier scouting.

Scenario 2: Locating Regional Suppliers with Specific Collaborations

In a second scenario, an enterprise might seek all raw material suppliers located in a specific county (e.g., Philadelphia) that also supply a particular large company, such as MEDICAL PRODUCTS LABORATORIES, INC. This query combines geographic filtering with relationship-based criteria, functionality rarely supported by traditional supplier databases. Using the knowledge graphs built on graph databases, this complex search was executed with ease, again allowing dynamic input of county names or target companies. Figure 7 displays the query results for: "All raw material suppliers in PHILADELPHIA who work with MEDICAL PRODUCTS LABORATORIES, INC".

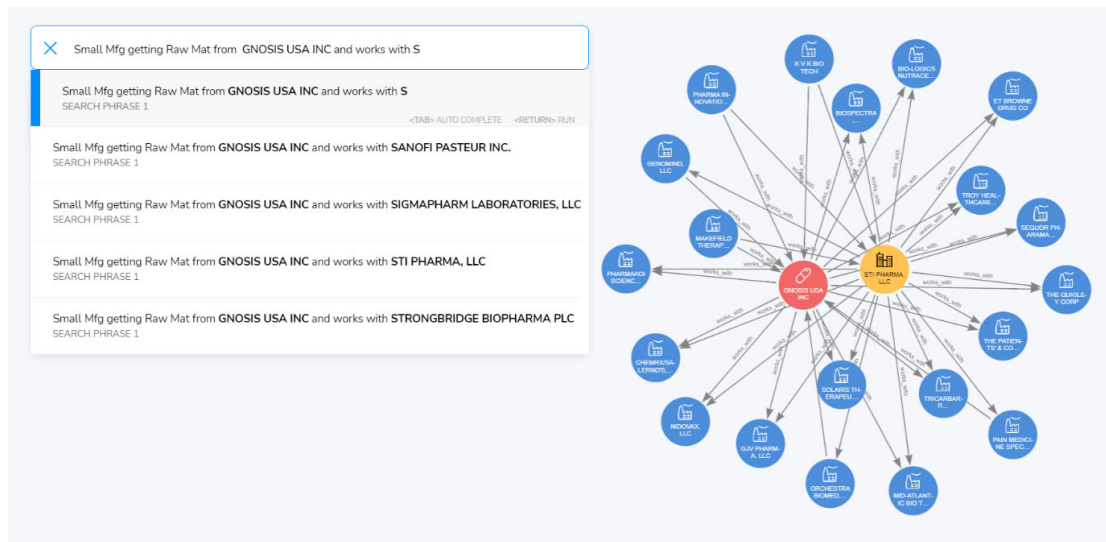


Figure 6. Multi-conditional query results for Scenario 1, showing enterprises (blue nodes) that (i) source raw materials from GNOSIS USA INC. and (ii) work with STI PHARMA LLC.

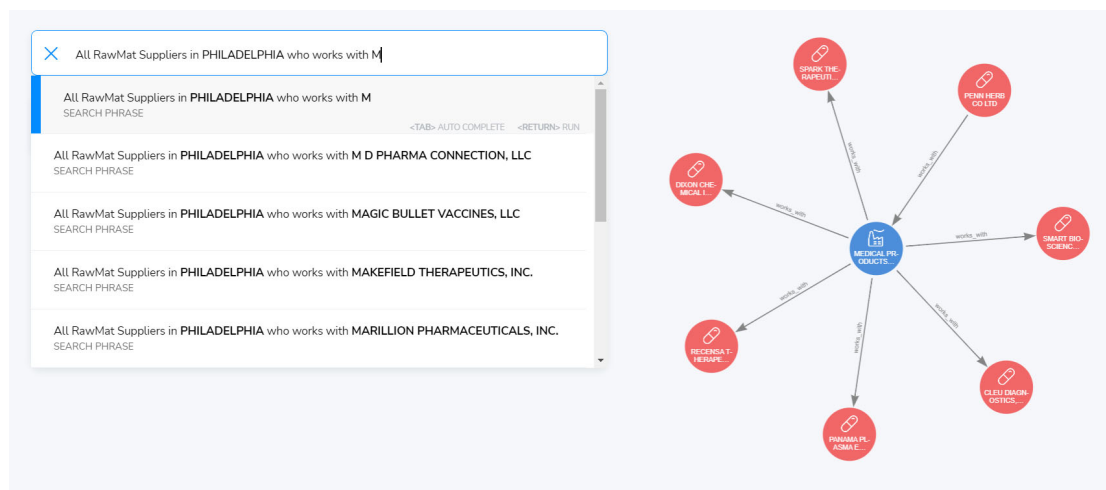


Figure 7. Multi-conditional query results for Scenario 2, showing suppliers in Philadelphia (red nodes) that work with MEDICAL PRODUCTS LABORATORIES, INC.

It is also worthwhile noting that the queries are dynamic, meaning the knowledge graph updates the list of companies available for selection in the dropdown menu in real time as queries are typed. This capability enhances localized supplier discovery, improves supply chain transparency, and supports regional economic development by enabling enterprises and policymakers to identify and strengthen trusted local supply networks.

3.2. Performance Evaluation and Comparison

To assess the performance of the proposed knowledge graph model, a series of test queries discussed in Section 3.1.3 were executed on the constructed knowledge graph. The evaluation focused on the following key metrics: (i) query execution time (Q) in milliseconds (ms) and (ii) the number of nodes and edges returned. The test cases included both (i) simple linear queries with single-hop lookups within the network, and (ii) complex multi-conditional queries requiring traversal of multiple hops and filtering based on combined conditions. Table 2 summarizes the tested queries (referenced in Section 3.1.3) and their outcomes.

Table 2. Performance Summary of Knowledge Graph Queries.

Query Type	Scenario	# Nodes	# Edges	Q (ms)
Simple Linear Search	3.1.3 (a)	57	56	<1
Multi-Conditional Query	3.1.3 (b) Scenario 1	22	41	5
Multi-Conditional Query	3.1.3 (b) Scenario 2	8	7	8

The knowledge graph-based model delivered near-instantaneous results for simple queries (<1 ms) and maintained execution times under 10 milliseconds even for complex multi-hop queries. This contrasts sharply with existing platforms like Wikidata [23], which rely on manual or semi-automated updates and offer incomplete industry coverage. For example, a SPARQL query on Wikidata for BioPharma enterprises (Q507443) in the United States (Q30) returned only 217 results, far below actual industry counts and took approximately 802 milliseconds for a simple query. Moreover, platforms such as Wikidata lacks support for complex, multi-conditional relationship queries like those demonstrated here. Therefore, the proposed knowledge graph-based approach not only supports richer, relationship-driven queries but does so with significantly faster response times

3.3. Validation and Accuracy Check

To assess the accuracy of the constructed knowledge graph and its querying mechanism, a targeted manual validation was performed. A stratified random sample of eight enterprises, one from each of the eight key categories (e.g., manufacturers, suppliers, logistics providers), was selected. For each enterprise, direct relationships (first-degree connections) identified by the automated query mechanism were manually cross-checked against known partnerships recorded in the graph database. Additionally, three complex multi-hop queries, similar to those illustrated in Section 3.1.3 (b), were executed and verified.

4. Interoperability and Standardization

Knowledge graphs and graph databases have seen increasing adoption across industries due to their growing relevance and applications in diverse domains [51]. The knowledge graph developed in this work for the Pennsylvania (PA) BioPharma sector to enhance supply chain resilience represents one such application. However, to extend the utility and applicability of this knowledge graph across a broader ecosystem of SMEs, it must support key capabilities such as data exchange and integration, semantic reuse, creation of new knowledge, and shared understanding across different systems. These capabilities are collectively encompassed by the concept of interoperability, defined as the ability of multiple systems to seamlessly exchange, interpret, and utilize information [52]. In the context of graph databases and knowledge graphs, interoperability implies that data must be represented, stored, and shared in a manner that is both machine-readable and intelligible across different platforms and domains [53]. One of the most effective ways to achieve interoperability is through the adoption of standardized frameworks and protocols.

4.1. Semantic Web Standards

Standardization is driven by consensus among stakeholders including industry, academia, user groups, and regulatory bodies, and ensures common protocols and data formats [54]. The World Wide Web Consortium (W3C) plays a central role in defining such standards for web-based information exchange. The W3C's Semantic Web initiative (also known as Web 3.0) introduces a framework for encoding data with well-defined semantics, enabling it to be shared and reused across applications, communities, and enterprises. According to W3C, the Semantic Web provides foundational technologies such as [55]:

- (i) RDF (Resource Description Framework): The standard model for data interchange.
- (ii) SPARQL: The standard query language for RDF-based data.

These technologies facilitate the creation of interoperable datasets and ontologies, supporting flexible integration and reasoning across domains. To make the PA BioPharma knowledge graph interoperable, it was mapped into the RDF format, the W3C standard for structured data exchange on the web. This translation allows seamless integration into other applications and domains, enabling a scalable, interoperable knowledge ecosystem.

4.2. RDF Data Model

The Resource Description Framework (RDF) is a W3C-standardized data model designed for structured information exchange on the Web [6]. It provides a way to represent and link data so that users can understand the meaning of entities and their relationships across different domains [56]. At its core, RDF represents data as triples composed of Subject, Predicate, and Object. The subject and object represent nodes or resources, while the predicate defines the relationship between them. For example, the statement “Entity 1 works with Entity 2” can be modeled using the following RDF triples:

- Subject (Node): Entity 1
- Predicate (Edge): works with
- Object (Node): Entity 2

Each RDF triple uniquely identifies these elements, making it possible to clearly model entities and their relationships. These relationships are often described using shared vocabularies such as FIBO [57], FOAF [58], or Schema.org [59] to support semantic interoperability across systems.

RDF data can be serialized in various formats such as Turtle, RDF/XML, N-Triples, or JSON-LD. For instance, the statement “Entity 1 is located in Philadelphia and works with Entity 2” can be represented in Turtle format as shown below:

```
@prefix sch: <http://www.example.org/schema/>
<entity:1> sch:Works_With <entity:2>
<entity:1> sch:City "Philadelphia"
```

In this snippet, “entity:1” and “entity:2” represent entities (e.g., companies or organizations), and “sch:Works_With” is a property linking them. The predicate “sch:City” associates Entity 1 with the city “Philadelphia.” The prefix “sch:” serves as a shorthand for the full URI (Uniform Resource Identifier): <http://www.example.org/schema/> (accessed on 13 May 2025). The usage of the prefixes such as “sch:” helps improve readability and reduce repetition in RDF syntax.

Figure 8 illustrates a simplified RDF graph based on this example. Since RDF decomposes each data statement, every entity (e.g., Entity 1 and Entity 2) and its properties (e.g., City) are represented as distinct components, often as separate nodes or edges, in the RDF graph. Formal definitions of RDF triple structure, data types, and Turtle serialization syntax are provided in Appendix B for reference.

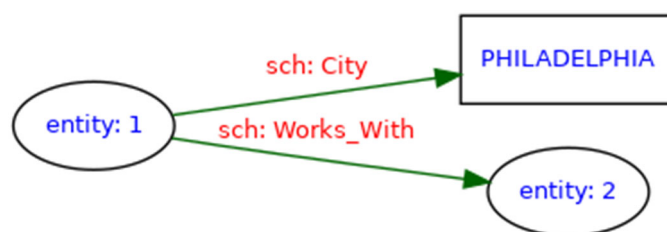


Figure 8. RDF graph representation showing two triples in the order [Subject, Predicate, Object]: (i) [entity:1, sch:City, “Philadelphia”]; (ii) [entity:1, sch:Works_With, entity:2].

4.3. Database Mapping

In simple terms, database mapping is the process of translating a database from a source database format to a target database format. To enhance interoperability, the knowledge graph for PA BioPharma, which is a Labeled Property Graph (LPG), needs to be mapped to the RDF format and back in a lossless manner (i.e., preserving all the information). This mapping is performed based on the following procedures:

- (i) Props \rightarrow W (Property names from LPG mapped to URIs in RDF).
- (ii) Values \rightarrow L (Property values from LPG mapped to Literals in RDF).
- (iii) Labels \rightarrow L \cup W (Labels from LPG mapped to Literals or URIs in RDF).
- (iv) V \cup E \rightarrow W \cup B (Nodes and Edges from LPG mapped to URIs or BNodes in RDF).

In other words, every element in the LPG, whether node, edge, label, or property, is transformed into a corresponding RDF triple. This enables the preservation of both data structure and semantics.

Figure 9 illustrates this mapping process, showing how the nodes and edge in the LPG are converted into seven distinct RDF triples. In the RDF representation, each entity, relationship, and property is decomposed into triples consisting of a Subject, Predicate, and Object. This granular decomposition allows all information from the LPG to be captured explicitly, with each property and value represented as a separate node or edge in the RDF graph.

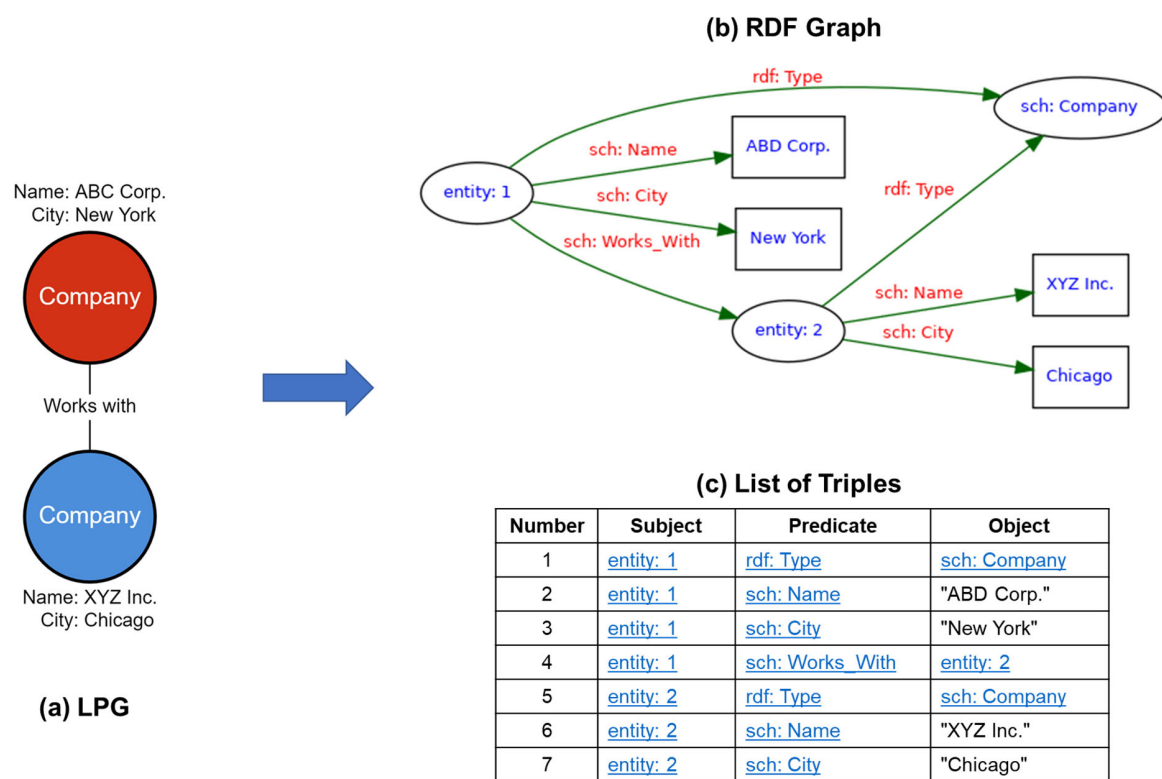


Figure 9. Mapping from LPG to RDF. (a) LPG Representation, (b) RDF Representation, (c) List of Triples obtained from the RDF.

As shown in Figure 9c, some triples use prefixes such as “sch:” or “rdf:”. These act as shorthands for full URIs (e.g., rdf: corresponds to <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>> (accessed on 13 May 2025), and sch: to <<http://www.example.org/schema/>> (accessed on 13 May 2025). While such custom or default vocabularies work for general use, they lack semantic alignment with widely adopted standards. To enhance broader web-based interoperability, these prefixes could be mapped to standardized vocabularies such as Schema.org [59].

4.4. Interoperability Workflow for PA BioPharma Knowledge Graph

To address the interoperability needs of the knowledge graph developed for the PA BioPharma sector, this work follows a three-step methodological framework:

- (1) Knowledge Graph Creation: Building the domain-specific knowledge graph as described in Section 3, using the Neo4j graph database.
- (2) RDF Mapping: Mapping the knowledge graph to the RDF standard.
- (3) Schema.org Extension: Extending the RDF graph with Schema.org vocabulary to achieve semantic interoperability.

Figure 10 illustrates this overall interoperability workflow, providing a conceptual roadmap for the detailed steps discussed in the sections that follow.

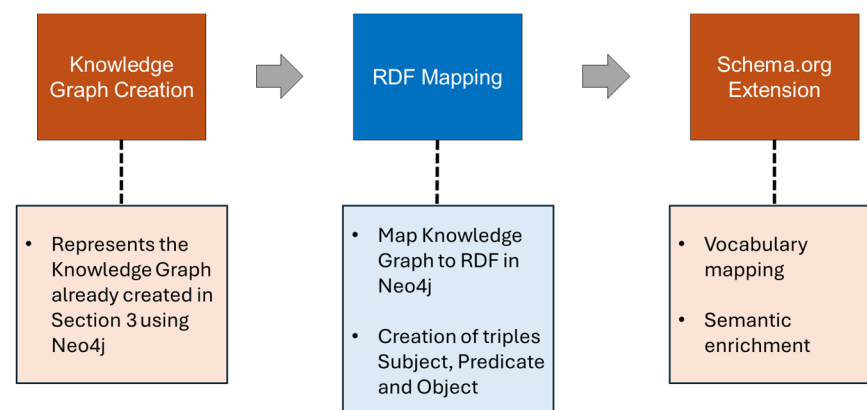


Figure 10. Interoperability workflow showing three stages: knowledge graph creation, mapping the knowledge graph to RDF and extending it with Schema.org for semantic integration across SME supply chains.

4.4.1. RDF Mapping

The knowledge graph constructed in Section 3 serves as the input for the RDF mapping process. This existing graph was implemented in Neo4j and subsequently mapped to the RDF using the Neosemantics (n10s) extension. This extension enables seamless interoperability by allowing data stored in Neo4j's Labeled Property Graph (LPG) model to be transformed into standard RDF serializations. Data insertion is performed using Cypher query language, and the output is, by default, serialized in Turtle format, though other RDF formats such as RDF/XML can be produced through content negotiation. The original LPG-based knowledge graph comprised 488 nodes and 11,520 edges. After conversion to RDF, the graph expanded into a total of 16,086 RDF triples, reflecting RDF's more granular, triple-based representation. Figure 11 presents a partial view of the RDF serialization in RDF/XML format.

As seen in Figure 11, the RDF output uses a default vocabulary automatically generated based on the Neo4j schema. This vocabulary assigns generic URI prefixes (e.g., neo4j://graph.xxx) to represent node labels, properties, and relationships. Although this schema-specific vocabulary is sufficient for serialization and internal querying, it lacks interoperability for broader web-based data integration. To address this, it is necessary to map the terms in the RDF graph to publicly recognized vocabularies. Specifically, the generic Neo4j vocabulary is aligned with equivalent terms in Schema.org, a widely adopted RDF schema that enhances semantic consistency and facilitates data sharing and integration across heterogeneous systems. This vocabulary alignment ensures that the RDF representation of the PA BioPharma knowledge graph is both machine-readable and semantically interoperable with external data sources and ontologies.

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:n4sch="neo4j://graph.schema#"
  xmlns:n4ind="neo4j://graph.individuals#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="neo4j://graph.individuals#73">
    <rdf:type rdf:resource="neo4j://graph.schema#LogiSupply" />
  </rdf:Description>

  <rdf:Description rdf:about="neo4j://graph.individuals#72">
    <rdf:type rdf:resource="neo4j://graph.schema#Company" />
    <n4sch:name>COMPLETE INTRAVENOUS ACCESS SERVICES INC</n4sch:name>
  </rdf:Description>

  <rdf:Description rdf:about="neo4j://graph.individuals#73">
    <rdf:type rdf:resource="neo4j://graph.schema#Company" />
    <n4sch:name>PHYSICIANS RX PHARMACY OF CLINTON, INC</n4sch:name>
  </rdf:Description>

  <rdf:Description rdf:about="neo4j://graph.individuals#72">
    <rdf:type rdf:resource="neo4j://graph.schema#Testing" />
  </rdf:Description>

```

Figure 11. Partial view of the RDF representation (in RDF/XML format) of the PA BioPharma Knowledge Graph.

4.4.2. Schema.org Extension

Schema.org is a widely adopted public vocabulary comprising an extensive collection of schemas for structured data. It facilitates machine readability and enhances semantic understanding of content on the internet, significantly improving data discoverability and search engine performance [60–62]. Developed collaboratively by major search engines including Google, Microsoft, Yahoo, and Yandex, Schema.org provides definitions for commonly used entities such as organizations, places, products, services, and their associated properties [63]. Currently, Schema.org vocabularies are used by over 10 million websites and applications to make data more accessible and interpretable by intelligent systems. Figure 12 illustrates the rich, structured definitions available through Schema.org, specifically for the tag “Organization” (<https://schema.org/Organization>, accessed on 13 May 2025) and its associated property “address” (<https://schema.org/address>, accessed on 13 May 2025).

Organization
A Schema.org Type
Thing > Organization

An organization such as a school, NGO, corporation, club, etc. [more...]

Property	Expected Type	Description
Properties from Organization		
actionableFeedbackPolicy	CreativeWork or URL	For a NewsMediaOrganization or other news-related Organization, a statement about public engagement activities (for news media, the newsroom's), including involving the public - digitally or otherwise -- in coverage decisions, reporting and activities after publication.
address	PostalAddress or Text	Physical address of the item.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
alumni	Person	Alumni of an organization. Inverse property: alumniOf
areaServed	AdministrativeArea or GeoShape or Place or Text	The geographic area where a service or offered item is provided. Supersedes serviceArea.
award	Text	An award won by or for this item. Supersedes awards.
brand	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.
contactPoint	ContactPoint	A contact point for a person or organization. Supersedes contactPoints.
correctionsPolicy	CreativeWork or URL	For an Organization (e.g. NewsMediaOrganization), a statement describing (in news media, the newsroom's) disclosure and correction policy for errors.
department	Organization	A relationship between an organization and a department of that organization, also described as an organization (allowing different uris, logos, opening hours). For example: a store with a pharmacy, or a bakery with a cafe.
dissolutionDate	Date	The date that this organization was dissolved.

address
A Schema.org Property
Thing > Property :: address
Physical address of the item.

Values expected to be one of these types
PostalAddress
Text

Used on these types
GeoCoordinates
GeoShape
Organization
Person
Place

Figure 12. Rich definition of the tags “Organization” and “address” in Schema.org.

To ensure semantic interoperability of the RDF representation of the PA BioPharma knowledge graph, the generic Neo4j schema vocabulary is mapped to Schema.org vocabulary. This process involves identifying a corresponding term in Schema.org for each node label, property, and relationship in the original Neo4j schema. As a result, the Neo4j-specific URI prefixes (e.g., neo4j://graph.xxx) are systematically replaced with Schema.org URIs (e.g., <https://schema.org/xxx>), producing a semantically enriched RDF graph that aligns with publicly accepted web standards. Table 3 presents the mapping between Neo4j schema elements and their equivalent Schema.org elements. A partial graphical view of the RDF graph updated with Schema.org vocabulary is shown in Figure 13.

Table 3. Neo4j schema elements and their equivalent Schema.org schema elements.

Neo4j Schema Element	Element Type	Schema.org Schema Elements (Accessed on 13 May 2025)
Company	label	https://schema.org/Organization
Name	property	https://schema.org/legalName
Type	property	https://schema.org/affiliation
NAICS Code	property	https://schema.org/naics
City	property	https://schema.org/location
Zip	property	https://schema.org/postalCode
Works_With	relationship	https://schema.org/seller

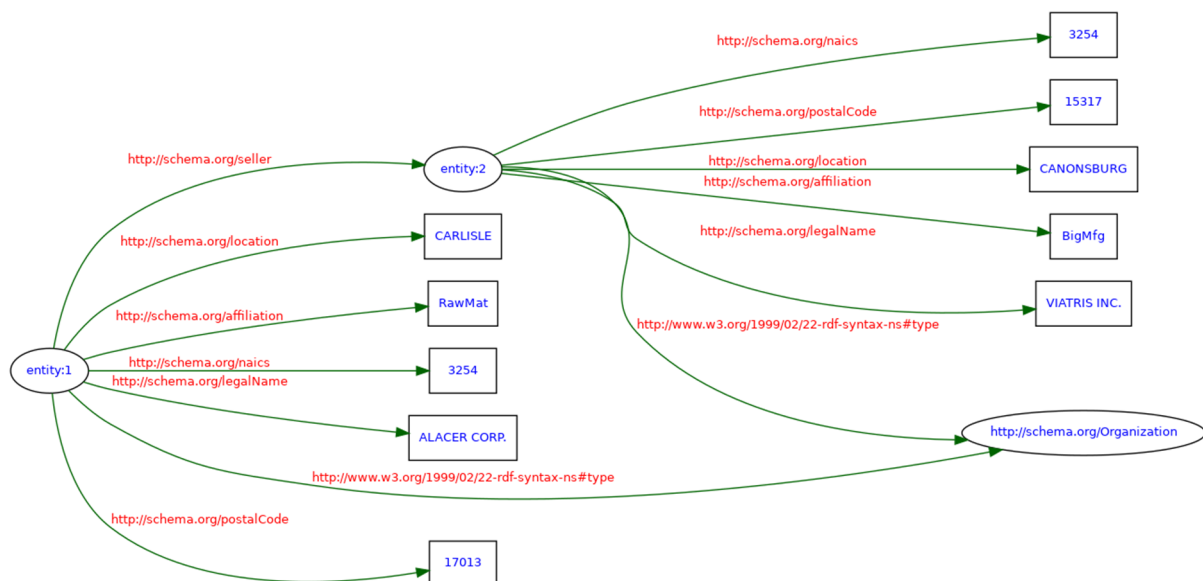


Figure 13. Graphical representation of RDF triples mapped using Schema.org vocabulary.

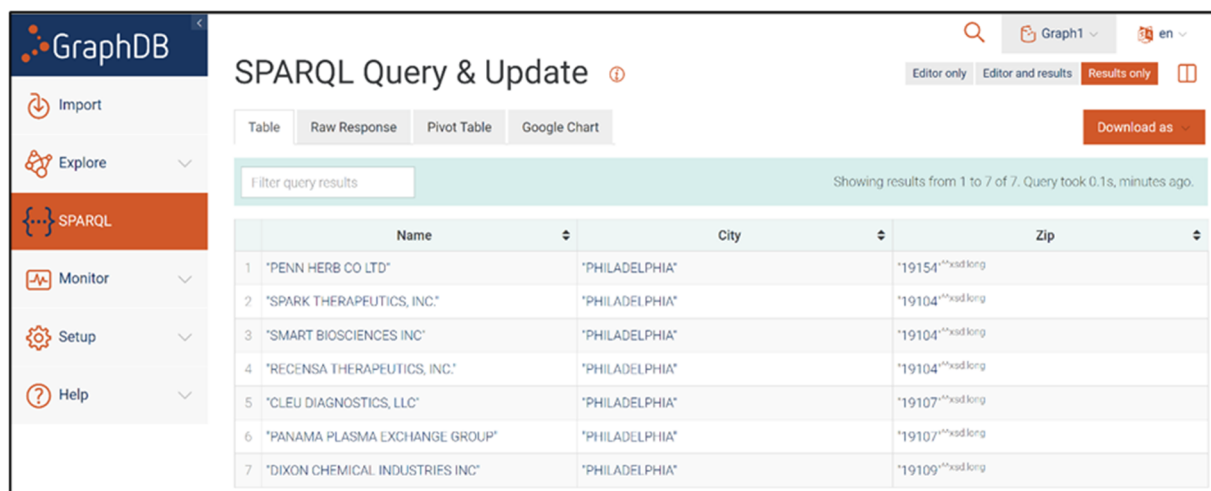
It is important to note that both Neo4j's LPG model and RDF are based on underlying graph structures, which greatly simplifies the mapping process compared to data models based on tabular (relational) or document-oriented schemas [63]. This structural similarity ensures that the transformation between LPG and RDF, especially when enhanced with Schema.org, is not only seamless but also preserves semantic context and data fidelity.

The transition of the PA BioPharma knowledge graph to RDF format extends beyond technical representation. RDF serialization aligned with Schema.org enables semantic interoperability, allowing the graph to integrate across platforms and applications. This alignment facilitates incorporation into Linked Open Data ecosystems, supplier discovery portals, and other web-based platforms, thereby extending the utility of the knowledge graph beyond its initial implementation. As a result, RDF serialization supports practical

supply chain use cases such as enhanced partner discovery, cross-domain querying, and dynamic integration with emerging datasets.

4.5. Interoperability Validation

To evaluate the interoperability of the RDF-based graph data model generated from the Neo4j knowledge graph, the RDF serialization (comprising 16,086 triples) was imported into GraphDB [64], a leading RDF triple store known for its native SPARQL support. The RDF file was loaded via the GraphDB browser interface, with the loading process completing in approximately 100 milliseconds. To assess interoperability, the same query scenarios discussed in Section 3.1.3 were re-executed on the RDF graph within GraphDB. Each Cypher query used in Neo4j was translated into a semantically equivalent SPARQL query and implemented in GraphDB. The results obtained from GraphDB matched exactly with those from the Neo4j implementation, both in terms of returned entities and relationships. This confirms that the RDF representation preserved the structural and semantic integrity of the knowledge graph, demonstrating cross-platform operability between Neo4j and GraphDB. Figure 14 illustrates the query results for Scenario 2 (Section 3.1.3 (b)) as executed within GraphDB.



	Name	City	Zip
1	"PENN HERB CO LTD"	"PHILADELPHIA"	"19154"^^xsd:long
2	"SPARK THERAPEUTICS, INC."	"PHILADELPHIA"	"19104"^^xsd:long
3	"SMART BIOSCIENCES INC"	"PHILADELPHIA"	"19104"^^xsd:long
4	"RECENSA THERAPEUTICS, INC."	"PHILADELPHIA"	"19104"^^xsd:long
5	"CLEU DIAGNOSTICS, LLC"	"PHILADELPHIA"	"19107"^^xsd:long
6	"PANAMA PLASMA EXCHANGE GROUP"	"PHILADELPHIA"	"19107"^^xsd:long
7	"DIXON CHEMICAL INDUSTRIES INC"	"PHILADELPHIA"	"19109"^^xsd:long

Figure 14. RDF Query results for Scenario 2, Section 3.1.3 (b) implemented using SPARQL in GraphDB.

5. Discussion and Implications

Global supply chain disruptions have accelerated the shift toward localized manufacturing models, highlighting the strategic importance of regional small- and medium-sized enterprises (SMEs) in strengthening supply chain resilience. Despite this, large enterprises often face significant challenges in identifying and evaluating local SME partners due to fragmented data sources, lack of visibility, and limited access to reliable information. Conversely, SMEs struggle to gain exposure and integrate into larger supply chains, often hindered by resource constraints and limited networking opportunities.

This study presents a knowledge graph driven approach to model and analyze supply chain networks, using the PA biopharma sector as a representative case. Implementation of knowledge graphs using the Neo4j graph database effectively mapped SMEs and their interrelationships, demonstrating how such a framework can enhance supply chain scouting and diversification efforts. This aligns with prior research emphasizing the value of network-based models for representing inter-dependencies and enabling resilience analysis [15,16,19]. A core strength of the knowledge graph-based approach is its support for complex, multi-hop querying, enabling users to uncover indirect relationships and hidden collaboration opportunities that may not be apparent through traditional search tools. This

directly addresses the gaps noted in existing SME discovery approaches [27,29,33]. The flexibility of the graph database also enables real-time updates and schema evolution, allowing the system to adapt as the supply chain network grows or changes, consistent with supply chain theory on dynamic capabilities for resilience [9,10]. This demonstrates a clear answer to RQ1: knowledge graphs can effectively represent SMEs and their relationships within a regional supply chain, enabling dynamic information retrieval and discoverability through multi-hop querying.

Moreover, by converting the knowledge graph into RDF format and aligning it with the Schema.org vocabulary, the knowledge graph becomes interoperable with external systems, facilitating broader integration and reuse across platforms. This feature addresses RQ2, showing that semantic web technologies improve the interoperability of SME data, enabling seamless exchange and integration across systems, and supporting cross-domain queries and collaboration. Validation was performed through the inspection of representative query results and relationship structures, confirming alignment with known supplier networks based on publicly available company data.

To contextualize the contributions of this work, a comparison was drawn with publicly accessible platforms such as Wikidata and Thomasnet. Although both platforms offer useful SME-related information, several limitations were observed. Wikidata, despite supporting complex SPARQL queries, suffers from partial and incomplete data coverage in domain-specific contexts such as the biopharma sector and exhibits slower response times compared to the proposed knowledge graph. Furthermore, Wikidata relies on manual or semi-automated updates, resulting in delayed or outdated information [27,31]. It also lacks support for complex, multi-conditional relationship queries and semantic enrichment necessary for advanced supply chain analysis. Thomasnet, on the other hand, primarily functions as a keyword-based directory, offering simple text search capabilities without support for graph-based queries or multi-hop relationship discovery. Its database structure does not facilitate exploration of indirect supplier connections or complex inter-enterprise relationships. Crucially, neither Wikidata nor Thomasnet support semantic interoperability or alignment with widely adopted vocabularies such as Schema.org, limiting their integration potential within broader web-based or linked data ecosystems.

In contrast, the proposed knowledge graph framework delivers curated, domain-specific data with flexible multi-hop querying, enabling discovery of complex inter-enterprise relationships. Its near-instantaneous query response times (about 10 milliseconds even for multi-hop queries) demonstrate a significant performance advantage. Additionally, its real-time update capability and semantic interoperability via RDF serialization and Schema.org alignment provide superior functionality and integration potential compared to existing platforms. These findings build upon prior work advocating for domain-specific, semantically interoperable supply chain knowledge graphs as enablers of SME integration and resilience [27,29,34,42]. A structured comparison of these approaches is presented in Table 4.

Table 4. Comparison with Existing Platforms.

Feature	Wikidata	Thomasnet	Knowledge Graph (This Work)
Data Coverage	Partial, general-purpose	Partial, industry listings	Focused, domain-specific
Query Complexity	High (SPARQL)	Low (Keyword)	Low to High (Cypher Queries)
Multi-Hop Querying	Limited	Not Supported	Supported
Real-Time Updates	No	No	Yes (for user-supplied data)

The RDF serialization and Schema.org alignment of the PA BioPharma knowledge graph have significant implications for inter-enterprise data sharing and system integration. By adopting widely recognized semantic web standards, the framework enables seamless

exchange of structured, machine-readable data across organizations [34,36,42,44]. This supports integration with external systems and open data repositories, enhancing partner discovery, enabling cross-domain querying, and fostering dynamic collaboration across supply chain networks. Such interoperability positions the knowledge graph as a scalable foundation for shared supply chain intelligence ecosystems, aligning with artificial intelligence (AI) and machine learning (ML) initiatives, platform-based manufacturing [65], and enabling the development of digital supply chain twins.

While implementing a knowledge graph entails costs for software, infrastructure, and maintenance, several factors make this approach increasingly accessible for SMEs. Open-source graph database platforms like Neo4j Community Edition provide a cost-effective starting point, and cloud-based graph services offer scalable, pay-as-you-go options that lower upfront investment. Automation tools and prebuilt integration templates can further reduce technical complexity and cost. Additionally, shared industry platforms or consortium-based models present opportunities for SMEs to collectively access knowledge graph capabilities at a fraction of individual cost. With growing availability of affordable solutions and support models, the proposed framework is well-positioned to be a viable, practical tool for SMEs aiming to enhance visibility, collaboration, and resilience within regional supply chains.

Limitations

While the proposed knowledge graph framework demonstrates significant potential, several limitations must be acknowledged to contextualize its current scope and applicability. First, the approach relies heavily on the availability, quality, and diversity of input datasets. This study uses a single primary data source (Mergent Intellect) chosen for its reliability and domain-specific coverage of Pennsylvania's biopharma sector. While this ensures quality for the proof-of-concept, reliance on one source limits coverage, generalizability, and the completeness of the knowledge graph. Missing or incomplete data may impact relationship discovery and reduce robustness. Furthermore, automating data ingestion remains a challenge that could impact scalability if not addressed through dedicated pipelines, as noted in prior work [33,38]. Expanding data sources and developing automated ingestion pipelines would strengthen coverage, resilience, and applicability.

Second, the targeted manual query validation was limited in scope, involving a stratified random subset of enterprises. This approach served as a proof-of-concept to confirm the feasibility and correctness of the knowledge graph's querying methodology. However, larger-scale and more exhaustive validation would improve the robustness and generalizability of the conclusions. This could be addressed by developing automated validation tools, which are not currently available.

While the current implementation of the knowledge graph framework demonstrates robust performance on curated datasets, its scalability to larger networks warrants careful consideration. For instance, a comprehensive evaluation of graph database systems across three computing architectures and four scaling factors ranging from 1 GB to 1 TB revealed that Neo4j, while efficient for smaller datasets, exhibits performance bottlenecks for complex queries as graph size approaches 1 TB [66]. Although the study reported occasional query timeouts in certain scenarios, Neo4j was generally able to execute queries within approximately 1000 s at the largest tested scale [66]. These findings suggest that knowledge graph frameworks can be extended to handle larger datasets; however, increasing the scaling factor further, particularly beyond several million nodes and edges, may introduce challenges related to query execution time, timeouts, and performance degradation. Addressing these challenges may require alternative implementations, hardware optimizations, or architectural adjustments to ensure consistent performance as dataset sizes grow.

Additionally, extending this framework to other sectors or regions would require addressing variations in data availability, domain-specific ontologies, and sector-specific regulatory requirements, which could impact both scalability and applicability.

Additional limitations relate to data confidentiality and legal considerations, which pose significant challenges for sharing RDF triples between platforms. Confidential business information embedded in RDF graphs raises privacy concerns, necessitating measures such as data anonymization, secure sharing protocols, and role-based access control. Furthermore, legal and regulatory restrictions such as intellectual property rights and data protection laws may limit RDF data exchange across organizations and jurisdictions. Addressing these issues is essential to fully realize the potential of RDF-based knowledge graphs for cross-platform supply chain integration.

6. Conclusions

This study developed a knowledge graph-driven framework for modeling and analyzing supply chain networks, using the Pennsylvania biopharma sector as a case study. The approach demonstrated that a curated, domain-specific knowledge graph implemented in Neo4j, converted to RDF, and aligned with Schema.org can significantly enhance visibility, interoperability, and discovery of complex inter-enterprise relationships. Key strengths of this work include support for multi-hop queries, real-time updates, and semantic interoperability, which together enable faster and deeper supplier discovery compared to existing platforms such as Wikidata and Thomasnet.

Stakeholders stand to benefit in several ways: large enterprises gain improved access to reliable, structured SME information for partner scouting and risk mitigation, while SMEs gain enhanced visibility and integration opportunities in larger supply chains. This addresses critical gaps in supply chain resilience and fosters dynamic collaboration, particularly in sectors where rapid adaptation to disruptions is essential.

While the findings provide strong proof-of-concept evidence, limitations in data scope, scalability, and automated validation highlight opportunities for future work. Future research should move beyond the manual proof-of-concept validation by developing automated methods to improve scalability and accuracy. Extending the RDF-serialized knowledge graph for integration with external systems such as Linked Open Data or supplier discovery portals could further enhance interoperability and broaden its applicability across platforms. Since Neo4j performance may degrade at extremely large datasets, addressing scalability through optimization, alternative graph platforms, or distributed architectures will be essential for large-scale adoption. Additional opportunities include expanding data sources to improve generalizability across regions and industries, automating ingestion pipelines to reduce manual effort, developing user-friendly query interfaces such as natural language processing (NLP) to increase accessibility for non-technical users, and integrating predictive analytics and AI-driven insights to enable proactive supply chain risk management.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study can be obtained from Mergent Intellect, <https://www.mergentintellect.com> (accessed on 13 May 2025).

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A

Mathematical Definition and Example of Knowledge Graph Structure

A knowledge graph is a labeled property graph (LPG) comprising nodes (entities), edges (relationships), and associated labels and properties that semantically enrich the data for advanced querying and analysis. Mathematically, a knowledge graph can be expressed as a tuple: $G(V, E, s, p, r)$, where

1. V represents a finite set of nodes or vertices.
2. E represents the finite set of edges, where $V \cap E = \emptyset$.
3. $s: E \rightarrow (V \times V)$ represents a total function.
4. $p: (V \cup E) \rightarrow \text{Labels}$ represents a total function.
5. $r: (V \cup E) \times \text{Properties} \rightarrow \text{Values}$ represents a partial function

Figure A1 illustrates a knowledge graph comprising two nodes (labeled as Company A and Company B), each enriched with attributes such as company size, name, and location. The relationship between the nodes includes details such as the quantity of materials supplied from A to B.

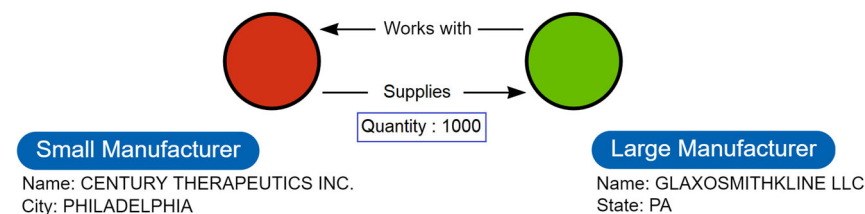


Figure A1. Knowledge graph comprising two nodes and their attributes.

The Knowledge graph shown in Figure A1 consists of the following elements:

$V = \{v_1, v_2\}$	$p(v_1) = \text{"Small Manufacturer"}$	$p(v_2) = \text{"Large Manufacturer"}$
$E = \{e_1, e_2\}$	$p(e_1) = \text{"Works With"}$	$p(e_2) = \text{"Supplies"}$
$s(e_1) = (v_2, v_1)$	$s(e_2) = (v_1, v_2)$	
$r(v_1, \text{"Name"}) = \text{"CENTURY THERAPEUTICS INC."}$		
$r(v_1, \text{"City"}) = \text{"PHILADELPHIA"}$	$r(v_2, \text{"Name"}) = \text{"GLAXOSMITHKLINE LLC"}$	
$r(v_2, \text{"State"}) = \text{"PA"}$	$r(e_2, \text{"Quantity"}) = 1000$	

Appendix B

Formal RDF Definitions and Serialization Examples

Assume three disjoint infinite sets L, B, W corresponding to the set of all Literals, Blank Nodes (BNodes) and URIs (Uniform Resource Identifiers), respectively. A URI is a unique link to a web resource, Blank Node is any anonymous resource that has no URI label, and Literal is anything with a value associated with it such as string, number, currency, date, etc. An RDF triple can be represented using a tuple $t = (s, p, o)$ where $s \in W \cup B$ is the subject, $p \in W$ is the predicate, $o \in L \cup B \cup W$ is the object. Therefore, a subject can be a Blank Node or a URI, predicate can be a URI, and object can be a Blank Node, URI or a Literal. An example of an RDF representation (or serialization) in Turtle data format is shown below:

```

@prefix sch: <http://www.example.org/schema/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix xmls: <http://www.w3.org/2001/XMLSchema#>.
<entity:1>
    a sch:Company;
    sch:Works_With <entity:2>;

```

```

sch:City "CARLISLE";
sch:Zip "17013" ^^xmls:long.

<entity:2>
  a sch:Company;
  sch:City "CANONSBURG";
  sch:Zip "15317" ^^xmls:long.

```

In this RDF code snippet, the first three lines denote the prefix definitions and start with @prefix. A prefix definition (e.g., sch) is used as a placeholder for lengthy URIs (e.g., <http://www.example.org/schema/>, (accessed on 13 May 2025)). Therefore, a full URI like <http://www.example.org/schema/City> (accessed on 13 May 2025) can be abbreviated as sch:City using the prefix definition. The BNodes are denoted by _: followed by a number, e.g., entity:1. The BNodes do not have an URI but have a local node index (NodeID) for reference purposes and this index is denoted by the number assigned to the Blank Node. Finally, the Literals are classified into two types: (i) simple literal, which is a string (e.g., "CANONSBURG") and (ii) a typed literal which comprises a datatype URI along with the string (e.g., "15317" ^^xsd:long). There can be many possible combinations of RDF triples (s, p, o); in the above code snippet depicts three kinds of valid RDF triples: (i) (Blank Node, URI, Literal), e.g., (entity:1, sch:City, "CARLISLE"), (ii) (Blank Node, URI, Blank Node), e.g., (entity:1, sch:Works_With, entity:2), and (iii) (Blank Node, URI, URI), e.g., (entity:1, rdf:type, sch:Company).

The RDF triples can be represented as a graph to visualize the connection between the subject, predicate and object effectively as shown in Figure A2. Here, the nodes represent the subject and/or the object and the edges represent the predicate.

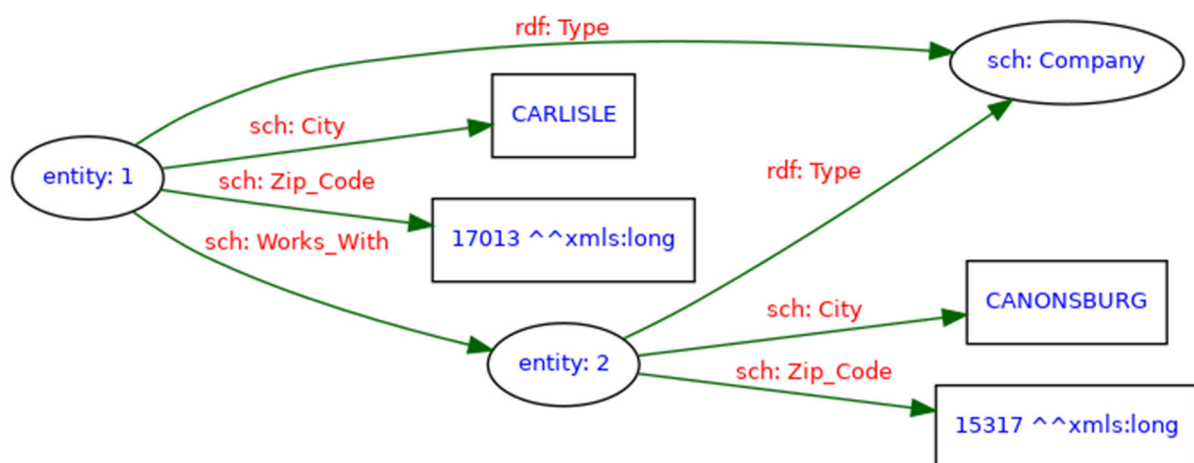


Figure A2. Graphical representation of RDF triples.

In Figure A2, the URI nodes and BNodes are represented by elliptical nodes. This can be either "subject" or "object" elements of the RDF triple. The Literal nodes, which are invariably the "object" elements, are represented by rectangular nodes. The Literal values can either be a text (such as the City name: "CANONSBURG") or a number (such as the Zip: "15317"). The "predicate" elements are represented by the arrows or connections and denote the property of the nodes (such as Name, City, Zip represented as a URI) or relationships between nodes (such as Works_With relationship represented again as a URI). This means that in an RDF representation, the data are broken down to the maximum possible extent, separating out the entities, their relationships and properties, and the literal values.

References

1. Harvard Business Review. Global Supply Chains in the Post Pandemic World. 2020. Available online: <https://hbr.org/2020/09/global-supply-chains-in-a-post-pandemic-world> (accessed on 13 May 2025).
2. Brookings Institution. Report on Reshoring Advanced Manufacturing and Supply Chains. Available online: <https://www.brookings.edu/articles/reshoring-advanced-manufacturing-supply-chains-to-generate-good-jobs/> (accessed on 13 May 2025).
3. McKinsey; Company. Report on SMEs and Their Growth. Available online: <https://www.mckinsey.com/industries/public-and-social-sector/our-insights/unlocking-growth-in-small-and-medium-size-enterprises> (accessed on 13 May 2025).
4. Kumar, V. A Novel IoT Based Solution for Cold Chain Monitoring in the Pharmaceutical Supply Chain. In *Lecture Notes in Networks and Systems, Proceedings of Tenth International Congress on Information and Communication Technology, London, United Kingdom, 18–21 February 2025*; Yang, X.S., Sherratt, S., Dey, N., Joshi, A., Eds.; Springer: Singapore, 2025; Volume 1413, pp. 393–403. [CrossRef]
5. El Baz, J.; Ruel, S.; Fozouni Ardekani, Z. Predicting the Effects of Supply Chain Resilience and Robustness on COVID-19 Impacts and Performance: Empirical Investigation through Resources Orchestration Perspective. *J. Bus. Res.* **2023**, *164*, 114025. [CrossRef]
6. U.S. Small Business Administration. 2022 Small Business Profile: Pennsylvania. Office of Advocacy. Available online: <https://advocacy.sba.gov/wp-content/uploads/2023/11/2023-Small-Business-Economic-Profile-PA.pdf> (accessed on 13 May 2025).
7. World Wide Web Consortium (W3C). Resource Description Framework (RDF). Available online: <https://www.w3.org/RDF> (accessed on 13 May 2025).
8. Tan, W.J.; Zhang, A.N.; Cai, W. A Graph-Based Model to Measure Structural Redundancy for Supply Chain Resilience. *Int. J. Prod. Res.* **2019**, *57*, 6385–6404. [CrossRef]
9. Shishodia, A.; Sharma, R.; Rajesh, R.; Munim, Z.H. Supply Chain Resilience: A Review, Conceptual Framework and Future Research. *Int. J. Logist. Manag.* **2023**, *34*, 879–908. [CrossRef]
10. Hashmi, A. Factors Affecting the Supply Chain Resilience and Supply Chain Performance: Supply Chain Resilience and Supply Chain Performance. *South Asian J. Oper. Logist.* **2022**, *1*, 65–85.
11. Adobor, H. Supply Chain Resilience: An Adaptive Cycle Approach. *Int. J. Logist. Manag.* **2020**, *31*, 443–463. [CrossRef]
12. Zamani, E.D.; Smyth, C.; Gupta, S.; Dennehy, D. Artificial Intelligence and Big Data Analytics for Supply Chain Resilience: A Systematic Literature Review. *Ann. Oper. Res.* **2023**, *327*, 605–632. [CrossRef]
13. The White House. Report on 100 Day Review of US Supply Chain. 2022. Available online: <https://bidenwhitehouse.archives.gov/wp-content/uploads/2021/06/100-day-supply-chain-review-report.pdf> (accessed on 13 May 2025).
14. Capgemini Research Institute. Fast Forward: Rethinking Supply Chain Resilience for a Post-COVID-19 World. Available online: https://www.capgemini.com/wp-content/uploads/2020/11/Fast-forward_Report.pdf (accessed on 13 May 2025).
15. Aldrichetti, R.; Battini, D.; Ivanov, D.; Zennaro, I. Costs of Resilience and Disruptions in Supply Chain Network Design Models: A Review and Future Research Directions. *Int. J. Prod. Econ.* **2021**, *235*, 108103. [CrossRef]
16. Lavassani, K.M.; Movahedi, B. Firm-Level Analysis of Global Supply Chain Network: Role of Centrality on Firm's Performance. *Int. J. Glob. Bus. Compet.* **2021**, *16*, 86–103. [CrossRef]
17. Li, Y.; Zobel, C.W.; Seref, O.; Chatfield, D. Network Characteristics and Supply Chain Resilience Under Conditions of Risk Propagation. *Int. J. Prod. Econ.* **2020**, *223*, 107529. [CrossRef]
18. Axtell, R.L.; Farmer, J.D. Agent-based modeling in economics and finance: Past, present, and future. *J. Econ. Lit.* **2025**, *63*, 197–287. [CrossRef]
19. Gomez, M.; Garcia, S.; Rajtmajer, S.; Grady, C.; Mejia, A. Fragility of a Multilayer Network of Intranational Supply Chains. *Appl. Netw. Sci.* **2020**, *5*, 71. [CrossRef] [PubMed]
20. Ernst & Young (EY). Report on Impact of COVID-19 on Supply Chains. 2021. Available online: https://www.ey.com/en_us/insights/supply-chain/how-covid-19-impacted-supply-chains-and-what-comes-next (accessed on 13 May 2025).
21. Kumar, V.; Srinivasan, V.; Kumara, S. Towards Smart Vaccine Manufacturing: A Preliminary Study During COVID-19. In Proceedings of the ASME 2021 International Mechanical Engineering Congress and Exposition, Virtual Online, 1–5 November 2021. [CrossRef]
22. Kumar, V.; Srinivasan, V.; Kumara, S. Smart Vaccine Manufacturing Using Novel Biotechnology Platforms: A Study During COVID-19. *J. Comput. Inf. Sci. Eng.* **2022**, *22*, 040903. [CrossRef]
23. Vrandečić, D.; Pintscher, L.; Krötzsch, M. Wikidata: The making of. In Proceedings of the ACM Web Conference, Austin, TX, USA, 30 April–4 May 2023; pp. 615–624.
24. Thomasnet Product Sourcing and Supplier Discovery Platform. Available online: <https://www.thomasnet.com> (accessed on 13 May 2025).
25. Mergent Intellect by FTSE Russell. Available online: <https://www.mergentintellect.com> (accessed on 13 May 2025).
26. Dun & Bradstreet: Intelligent Data for Business Performance. Available online: <https://www.dnb.com/en-us> (accessed on 13 May 2025).
27. Zhou, S.; Liu, Y.; Liu, Y. A Market Convergence Prediction Framework Based on a Supply Chain Knowledge Graph. *Sustainability* **2024**, *16*, 1696. [CrossRef]

28. AlMahri, S.; Xu, L.; Brintrup, A. Enhancing Supply Chain Visibility with Knowledge Graphs and Large Language Models. *arXiv* **2024**, arXiv:2408.07705.
29. Düggelin, W.; Laurenzi, E. A Knowledge Graph-Based Decision Support System for Resilient Supply Chain Networks. In Proceedings of the International Conference on Research Challenges in Information Science, Guimaraes, Portugal, 14–17 May 2024; Springer Nature: Cham, Switzerland, 2024; pp. 66–81.
30. Tu, Y.; Li, W.; Song, X.; Gong, K.; Liu, L.; Qin, Y.; Liu, S.; Liu, M. Using Graph Neural Network to Conduct Supplier Recommendation Based on Large-Scale Supply Chain. *Int. J. Prod. Res.* **2024**, *62*, 8595–8608. [[CrossRef](#)]
31. Karam, N.; Matini, S.; Laas, R.; Hoppe, T. A Hybrid Knowledge Graph and Bayesian Network Approach for Analyzing Supply Chain Resilience. In Proceedings of the European Semantic Web Conference, Crete, Greece, 28 May–1 June 2023; Springer Nature: Cham, Switzerland, 2023; pp. 27–31.
32. Kosasih, E.E.; Margaroli, F.; Gelli, S.; Aziz, A.; Wildgoose, N.; Brintrup, A. Towards Knowledge Graph Reasoning for Supply Chain Risk Management Using Graph Neural Networks. *Int. J. Prod. Res.* **2024**, *62*, 5596–5612. [[CrossRef](#)]
33. Huang, F.; Cheng, L. Distant Supervision Knowledge Extraction and Knowledge Graph Construction Method for Supply Chain Management Domain. *Auton. Intell. Syst.* **2024**, *4*, 7. [[CrossRef](#)]
34. Rolli, F.; Parretti, C.; Barbieri, R.; Polidoro, A.; Bindi, B. RDF Knowledge Graphs Designed with Axiomatic Methodology to Enhance Industry 4.0. *Machines* **2025**, *13*, 58. [[CrossRef](#)]
35. Saidi, C.; Hamani, N.; Benaissa, M.; Rolf, B.; Reggeline, T.; Lang, S. Modeling Reconfigurable Supply Chains Using Knowledge Graphs: Towards Supply Chain 5.0. *Prod. Eng.* **2025**, *19*, 795–818. [[CrossRef](#)]
36. Di Pierro, D. Ontology-enriched Graph Databases: An Interoperable Framework for Knowledge Integration and Management. Ph.D. Thesis, The University of Bari Aldo Moro, Bari, Italy, 2025.
37. McEachen, N.; Lewis, J. Enabling Knowledge Sharing by Managing Dependencies and Interoperability between Interlinked Spatial Knowledge Graphs. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 117–124. [[CrossRef](#)]
38. Thakker, H.V. On Supporting Interoperability Between RDF and Property Graph Databases. Ph.D. Dissertation, Universitäts- und Landesbibliothek Bonn, Bonn, Germany, 2021.
39. Shadkam, E. Multi-objective Supplier Selection with the New Hybrid COAW Method. *Int. J. Supply Chain. Oper. Resil.* **2021**, *5*, 60–78. [[CrossRef](#)]
40. Petersen, N. Towards Semantic Integration of Supply Chain and Production Data. Ph.D. Dissertation, Universitäts- und Landesbibliothek Bonn, Bonn, Germany, 2020.
41. Grangel-González, I. A Knowledge Graph Based Integration Approach for Industry 4.0. Ph.D. Dissertation, Universitäts- und Landesbibliothek Bonn, Bonn, Germany, 2019.
42. Petersen, N.; Halilaj, L.; Grangel-González, I.; Lohmann, S.; Lange, C.; Auer, S. Realizing an RDF-based Information Model for a Manufacturing Company—A Case Study. In Proceedings of the International Semantic Web Conference, Vienna, Austria, 21 October 2017; Springer International Publishing: Cham, Switzerland, 2017; pp. 350–366.
43. Ameri, F.; Wallace, E.; Yoder, R.; Riddick, F. Enabling Traceability in Agri-Food Supply Chains Using an Ontological Approach. *J. Comput. Inf. Sci. Eng.* **2022**, *22*, 051002. [[CrossRef](#)]
44. Petersen, N.; Lange, C.; Auer, S.; Frommhold, M.; Tramp, S. Towards Federated, Semantics-Based Supply. In *Lecture Notes in Business Information Processing, Proceedings of the Business Information Systems: 19th International Conference, Leipzig, Germany, 6–8 July 2016*; Springer: Cham, Switzerland, 2016; Volume 255, p. 436.
45. Tayyab, M.; Jemai, J.; Lim, H.; Sarkar, B. A sustainable development framework for a cleaner multi-item multi-stage textile production system with a process improvement initiative. *J. Clean. Prod.* **2020**, *246*, 119055. [[CrossRef](#)]
46. Sarkar, B.; Takeyeva, D.; Guchhait, R.; Sarkar, M. Optimized radio-frequency identification system for different warehouse shapes. *Knowl.-Based Syst.* **2022**, *258*, 109811. [[CrossRef](#)]
47. Robinson, I.; Webber, J.; Eifrem, E. *Graph Databases: New Opportunities for Connected Data*, 2nd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015.
48. Samanta, S.; Dubey, V.K.; Sarkar, B. Measure of influences in social networks. *Appl. Soft Comput.* **2021**, *99*, 106858. [[CrossRef](#)]
49. Ivanov, D. Supply chain viability and the COVID-19 pandemic: A conceptual and formal generalisation of four major adaptation strategies. *Int. J. Prod. Res.* **2021**, *59*, 3535–3552. [[CrossRef](#)]
50. Sarkis, J.; Dhavale, D. Supplier selection for sustainable operations: A triple-bottom-line approach using a Bayesian framework. *Int. J. Prod. Econ.* **2015**, *166*, 177–191. [[CrossRef](#)]
51. Angles, R.; Gutierrez, C. An Introduction to Graph Data Management. In *Graph Data Management*; Springer: Cham, Switzerland, 2018; pp. 1–28.
52. Park, A.; Wilson, M.; Robson, K.; Demetis, D.; Kietzmann, J. Interoperability: Our exciting and terrifying Web3 future. *Bus. Horiz.* **2023**, *66*, 529–541. [[CrossRef](#)]
53. Otto, B.; Ten Hompel, M.; Wrobel, S. *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*; Springer Nature: Cham, Switzerland, 2022; p. 580.

54. Iliadis, A.; Acker, A.; Stevens, W.; Kavakli, S.B. One schema to rule them all: How Schema.org models the world of search. *J. Assoc. Inf. Sci. Technol.* **2025**, *76*, 460–523. [CrossRef]
55. W3C Semantic Web. Available online: <https://www.w3.org/standards/semanticweb> (accessed on 13 May 2025).
56. Ali, W.; Saleem, M.; Yao, B.; Hogan, A.; Ngomo, A.C.N. A survey of RDF stores & SPARQL engines for querying knowledge graphs. *VLDB J.* **2022**, *31*, 1–26.
57. EDM Council. About FIBO. Available online: <https://edmcouncil.org/frameworks/industry-models/fibo> (accessed on 13 May 2025).
58. The FOAF Project. Friend of a Friend (FOAF) Vocabulary. Available online: <http://xmlns.com/foaf/spec> (accessed on 13 May 2025).
59. Schema.org. Schema.org Home Page. Available online: <https://schema.org> (accessed on 13 May 2025).
60. Issemantic. Schema Markup Validator. Available online: <https://issemantic.net/schema-markup-validator> (accessed on 13 May 2025).
61. Brinkmann, A.; Primpeli, A.; Bizer, C. The Web Data Commons Schema.org data set series. In Proceedings of the ACM Web Conference 2023, Austin, TX, USA, 30 April–4 May 2023; pp. 136–139.
62. Guha, R.V.; Brickley, D.; Macbeth, S. Schema.org: Evolution of Structured Data on the Web. *Commun. ACM* **2016**, *59*, 44–51. [CrossRef]
63. Barrasa, J. Creating a Schema.org Linked Data Endpoint on Neo4j. Available online: <https://jbarrasa.com/2018/10/18/quickgraph7-creating-a-schema-org-linked-data-endpoint-on-neo4j-in/> (accessed on 13 May 2025).
64. Ontotext GraphDB. Available online: <https://www.ontotext.com/products/graphdb> (accessed on 13 May 2025).
65. Kumar, V.; Srinivasan, V. Role of Platform Technologies and AI/ML in Advancing Pharmaceutical Manufacturing. In Proceedings of the ASME 2025 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Anaheim, CA, USA, 17–20 August 2025.
66. Rusu, F.; Huang, Z. In-Depth Benchmarking of Graph Database Systems With the Linked Data Benchmark Council (LDBC) Social Network Benchmark (SNB). Available online: <https://arxiv.org/abs/1907.07405> (accessed on 13 May 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.