HARMONIZING HETEROGENEOUS DATA: A MATERIALIZED KNOWLEDGE GRAPH APPROACH FOR FEDERATED INFORMATION SYSTEMS

Abid Ali Fareedi*, Ahmad Ghazawneh$^b$

*The Academy for Information Technology (ITE), Halmstad University, Halmstad, Sweden.

$^b$Kristiania University College, Oslo, Norway.

Abstract. Healthcare is one of the major industries where sharing information with a common understanding is essential. This research presents a federated knowledge graph framework with its materialized knowledge graph (MKG) approach. It also explores the transformative potential of federated knowledge graphs (FKGs) as a solution to address the persistent challenges of data integration and interoperability in today's complex information landscape, especially within federated information system development (ISD). The authors tailored Design Science Research Methodology (DSRM) to develop design artifacts embedded with a core domain-oriented ontological metadata model focused on cardiovascular diseases. By adopting FKGs, organizations can streamline data integration efforts, improve cross-system data harmonization, and foster seamless data exchange with metadata standardization among diverse information sources. This research highlights the advantages of FKGs in enhancing data connectivity with semantic alignment, seamless data transformation and interlinking, facilitating on-demand data retrieval to support optimization of query performance, scalability, and more effective decision-making processes within healthcare settings.

Keywords: Materialized knowledge graphs, Federated information systems, Design science research method, Ontology, Data Integration

1. Introduction

The healthcare industry has witnessed remarkable growth over the past few decades, primarily attributed to the integration of digital technologies [1]. These advancements have paved the way for cutting-edge approaches in managing, designing, and shaping the evolution of digital healthcare interoperability [2]. Many healthcare institutions are progressively adopting diverse health information systems (HIS), including electronic health record (EHR) systems, to enhance healthcare quality and patient care [3]. However, the digital transformation is not without challenges [4], and in fact, several HISSs are non-interoperable [5] and designated with no conformity to interoperability guidelines [6]. With the digital transformation of the healthcare industry and the digital data explosion [8], organizations have more data available and storage than ever to be consumed for countless business processes and operations [7]. However, the dominance of commutable data and its storage with a shared understanding of semantic meaning over analogue data is a recent phenomenon, and Lopez [8] estimated that the amount of digital data created by mankind surpassed the amount of analogue data during 2002-2003, rose to 94% in 2007 and is currently close to 100% of storables created by mankind. The annual growth rate of digital data was 58% during the years 1986-2007; consequently, the annual growth increased by 60% and the 100% share of digital data from all data created sources by mankind in 2010 [8]. During 2018, mankind created every third month as much storables digital data as we created from 10,000 BC until the end of 2013 in any format. In 2028, the same phenomenon will happen in every 20th hour; in 2038, it will occur in every 11th minute; in 2048, it will pop up in every 6th second [8].

Data integration and semantic interoperability are crucial in the information systems development (ISD)
discipline and play a pivotal role in developing information systems (ISs) for various domains because of its data versatility, such as structured or semi-structured forms that stem from diverse sources are often unsuccessful [9], especially in the healthcare sector. In healthcare organizations, the challenge of integrating, exchanging, and achieving semantic data interoperability across distributed health-related data sources is a serious and pressing issue [10] for the development of consolidation of distributed patient data and building a sustainable e-health system [11]. Semantic interoperability is primarily concerned with discovering innovative ways to assert equivalence relationships between data points from disparate sources. It is critical in various applications and use cases that need to query across autonomous and heterogeneous data sources [12]. Multiple stakeholders have dedicated their efforts to establishing a common standard that facilitates the seamless dissemination of information from diverse sources. This endeavour necessitates the creation of mappings that bridge the gap between disparate information sources and a unified taxonomy or shared upper-level ontology [13]. Since the inception of the semantic web, there has been a continuous incremental demand for developing systems that enable semantic interoperability using various methods, such as meta-databases or ontologies encompassing algorithms for identifying mappings [14, 13].

The literature on semantic interoperability can be divided into a two-part framework, as outlined by Liu et al. in 2020 [13]: 1) Data model development and 2) Semi-automatic or fully automatic semantic data integration within the domain of data model development. The focus has primarily been establishing and defining data standards across diverse fields. These fields include e-health [15], pharmaceutical drug discovery [16], and web service interoperability [17]. During their literature review, the authors encountered a range of challenges, with a notable one being the translation of information needs into standardized database queries. Collecting, integrating, reconciling, and efficiently extracting information from heterogeneous and autonomous data sources is widely acknowledged as a significant challenge in the broader corporate landscape [18].

This research targets the following RQ: How does the federated knowledge graph framework improve data integration and interoperability issues using a materialized knowledge graph approach in federated information systems (IS)?

The research is inspired by the ontology-based data access (ODBA) [19] paradigm and aims to highlight the importance of the virtual knowledge graph (VKG) approaches to data access or integration and also emphasizes the design and development of federated knowledge graphs (FKGs) framework using ontological paradigms to incorporate global or integrated schema from disparate data sources or data models and map the data sources to the global schema that is referred federation of data [20]. Achieving this goal necessitates adopting highly flexible data models for the global schema within federated knowledge graphs (FKGs), wherein vocabulary is articulated through an ontology.

The suggested framework serves as an intermediary layer in contrast to the conventional data integration approach, which relies on a relational global schema and extraction-transform-load (ETL) processes, as seen in materialized federated knowledge graphs (MFKGs). The MFKGs approach is the reciprocal of the virtual knowledge graphs (VKGs) approach. The MFKG approach is a methodology for managing and accessing data in a distributed and federated information system development and is different from VKGs because of its salient features: it does not require committing early on to a specific structure; it provides better accommodate heterogeneity phenomenon; better deal with missing or incomplete information from different sources; does not require a complex restructuring operation to accommodate new information or inclusion of new data sources. It also enhances query performance and reduces the need for constant data synchronization from different sources. This dynamic and volatile behaviour helps the corporation own it and create value within organizations as a severe contribution in today’s complex information landscape.

This research is structured with the following sections: Section 2 provides a brief overview of desktop research, a taxonomy of heterogeneity in health data sources and repositories, health data standards for federated IS, and Integrating heterogeneous IS to federated IS using applied ontologies. Section 3 presents the methodology: data collection, data analysis, problem identification and motivation for a proposed solution and relevance; Anatomy of materialized federated knowledge graph (MFKGs) - Architectural artifact. Section 4 presents the experimental results. Section 5 presents evaluation and testing procedures using SPARQL and DL query structures in the Ontotext graph database for data visualization. Section 6 describes the discussion section, and lastly, Section 7 offers a conclusion and future possibilities and directions.

2. Theoretical Background

2.1. A Taxonomy of Heterogeneity in Health Data Sources and Repositories

The taxonomy of heterogeneity in healthcare can be categorized into three main classes: 1) Data-level heterogeneity, 2) Ontological heterogeneity, and 3) Temporal heterogeneity [21]. Data-level heterogeneity refers
to variations, differences, or disparities in data at the granular level within a data set. It indicates that data points or observations within a data set exhibit diversity, dissimilarity, or inconsistency regarding their attributes, formats, values, or structures. Data-level heterogeneity arises from various sources and poses challenges for data analysis, integration, interlinking, and interpretation when the same entity is expressed and represented differently in different contexts, as data-level heterogeneity [21]. Ontological heterogeneity refers to differences in how data is structured, categorized, and represented within a data set, particularly regarding the underlying concepts or ontologies used to describe and organize the data. It is a type of heterogeneity that occurs when data sources or systems use different applied ontologies or vocabularies to represent similar or related information. Ontological heterogeneity can make it challenging to integrate, query, and interpret data from diverse sources, as the differences in underlying semantics can lead to misunderstandings and errors [21]. Temporal heterogeneity refers to variations or differences in data over time. It indicates that data points or observations within a data set exhibit changes, fluctuations, or patterns concerning time-related attributes. Temporal heterogeneity is particularly relevant when analyzing data in which the temporal aspect is essential, such as time series or longitudinal data [21]. In this paper, the proposed framework for an ontological virtual knowledge graph can also effectively tackle structural and syntactic ontological heterogeneities.

2.2. Healthcare Data Standards for Federated Information Systems

Developing a better-functioning healthcare system requires comprehensive patient-related data, i.e. electronic health record (EHR) available at the time of diagnosis and accessible within context according to the demand of healthcare professionals (HPs) during ward-round [22] point of care [10]. In the traditional healthcare landscape, data sharing has been carried out through various methods. These methods encompass faxing reports, dispatching paper copies, the manual or electronic submission of claims, offering data downloads, and electronically transmitting data. This data sharing is especially prevalent when patients are part of the same healthcare-integrated delivery network within a healthcare organization that adheres to a health information exchange (HIE) framework. An HIE involves exchanging health information for patient care, bridging the gap across conventional business boundaries in the healthcare sector, as highlighted by Hersh in 2009 [23]. It also facilitates the development of federated information systems (FIS), as described by Busse et al. [24].

In constructing a comprehensive federated information repository, data must be built upon a standard layer, data elements and terminology, structure, and organization in order to share data or information from various sources. These requirements are called interoperability. The term interoperability is classified into functional and semantic interoperability. Functional interoperability refers to the group participants supporting standard functions and procedures. Similarly, Semantic interoperability explains that the communication language must be understood by a computer at the receiving end of the communication [10]. Interoperability requires the use of golden standards. Wager et al. [25] defined standards and categorized them into three further main categories: 1) Classification, vocabulary, and terminology standards; 2) Data interchange standards; and 3) Health record content standards. Several recognized types of standards development organizations (SDOs), such as general standards, data components, data interchange, knowledge representation, EHR, and application-level support, are working to achieve interoperable standards internationally [26]. The authors used SDO standards in this research (illustrated in Table 1).

<table>
<thead>
<tr>
<th>Types of Standard</th>
<th>Example Standards</th>
<th>SDO Creating the Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>General standards</td>
<td>XML, TCP/IP, 802.11, Web services, security,</td>
<td><em>W3C</em>, IETF, IEEE, OMG, HL7</td>
</tr>
<tr>
<td>Data components</td>
<td>Reference Information Model (RIM), data elements, data types, terminology, templates, clinical statements, clinical document architecture</td>
<td>HL7, CEN, ISO, openEHR, IHTSDO, LOINC, RxNorm, UMLS, WHO</td>
</tr>
<tr>
<td>Data interchange</td>
<td>Structured and free-form documents, images</td>
<td>HL7, ASTM, DICOM, IEEE 1073, NCPDP, X12N, CEN, ISO</td>
</tr>
<tr>
<td>Knowledge representation</td>
<td>Guidelines and protocols, decision support algorithms</td>
<td>HL7, ASTM, <em>Ontologies, Knowledge graphs (KGs)</em></td>
</tr>
<tr>
<td>Electronic health record (EHR)</td>
<td>Functional requirements, EHR models, Continuity of Care Record (CCR), patient summary record, personal health record</td>
<td>HL7, ASTM, openEHR, CEN</td>
</tr>
<tr>
<td>Application level support</td>
<td>Identifiers, resource registries, disease registries, tool sets, conformance requirements, implementation manuals</td>
<td>HIPAA, HL7, ASTM, ISO, CEN</td>
</tr>
</tbody>
</table>

*Table 1. Conceptual Framework for Standards Development Organizations [27]*
2.3. Integrating Heterogeneous Information Systems to Federated Information System Using Applied Ontologies

The creation of a federated information system (FIS) is a critical endeavour that aims to tackle inherent heterogeneities present at two distinct levels: 1) The system level and 2) The semantics level. This is imperative because one of the primary challenges is integrating a diverse array of distributed, autonomous, and heterogeneous IS [28]. System-level heterogeneities arise from variations among individual information systems concerning their platforms and data models. [29] (e.g., pre-relational and object-oriented schemas, XML files, Excel files, CSV files, ADL files, Unknown-formats, and OWL models) [30], data types and sizes, and database management systems (DBMS). Semantic-level heterogeneities arise from variations in perspectives and the terminologies employed by individual information systems [29]. Diverse methods are employed to address these challenges. One approach involves the integration of local schemas into a unified global or federated schema [31]. Additionally, the Wrapper/mediator approach [32] is used to tackle and resolve system-level heterogeneities to build federated IS or federated database management systems (DBMS) by utilizing proprietary bridges, adapters, wrappers or common intermediaries such as open database connectivity (ODBC) [29].

Semantic web technologies and ontological approaches are employed and entertained to facilitate information integration of very different systems because of the capability to get the precise meaning for the exchanged information. This phenomenon can be achieved through ontologies or web ontology language (OWL) structure (ontology is considered the cornerstone item of the semantic web; it is a modelling language that has the ability to build models) [33, 34]. The OWL is the de facto standard for ontology implementation and has established itself as the prevailing standard for ontology implementation. It empowers the precise articulation of data semantics. Within OWL, there exists a subset rooted in Description Logics (DL) called OWL-DL. This particular subset harnesses the capabilities of DL reasoning. In the realm of clinical models and clinical data, OWL-DL enables the execution of inference tasks, facilitating advanced reasoning and analysis [35]. In this research, authors explored semantic web technologies and concentrated on ontological models for developing federated IS using the materialized knowledge graphs (MKGs) approach.

3. Methodology

This research employs a tailored and holistic approach (illustrated in Fig. 1.), which involves iterative processes to elaborate on different phases of the design science research method (DSRM) [36]. The objective is to design and evaluate novel digital artifacts inspired by the theory of knowledge design [37]. The authors used a tailored DSR methodology and a customized federated materialized approach to design and build enterprise knowledge graphs (EKGs). This strategy helps to understand contextual knowledge within the healthcare domain (e.g., cardiovascular ontological metadata model) and other data models and their integration and knowledge management to facilitate end-users with shared understanding and unification of the data to resolve the data interoperability and data exchange issues. This DSR method provides a systematic way to understand the problem and motivation, presentation, evaluation and demonstration of the digital artifacts on different stages that showcase the data interoperability resolution through semantic web techniques such as ontologies and knowledge graphs (KGs). These steps are listed in the following section:

- **Problem Centered Initiation:** The first stage of our study is dedicated to identifying the knowledge gap surrounding the critical significance of data integration and resolving interoperability issues in developing data-driven applications. The authors delve into potential solutions, techniques, or approaches to achieve data standardization from the Information Systems (IS) perspective, focusing on the healthcare domain. The authors did desktop research and supplemented their findings by discussing them with colleagues and domain experts. Through this collaborative effort, we aim to underscore the significance of data standardization to overcome data integration and interoperability challenges within contemporary data-driven applications across diverse domains.

- **Objective Centered Initiation:** The second stage explores various standards, including general standards established by W3C, data standards such as Unified Medical Language System (UMLS), and data representation standards like ontologies and knowledge graphs. These standards are pivotal in presenting standardized data within federated health information systems (HIS).

- **Design and Development Initiation:** The third stage is constructing federated information systems (IS). For this purpose, authors devised a framework for federated knowledge graphs (FKGs), leveraging semantic web technologies paradigms to generate FKGs using a materialized KGs approach in the health care context.
- **Context Initiation**: The fourth stage is designated for developing various digital artifacts to demonstrate the systematic process of addressing data integration and interoperability issues and outlining the procedure for generating standardized data to support the development of various healthcare-related services.

- **Evaluation**: The fifth stage outlines the evaluation of various digital artifacts, particularly federated knowledge graphs (KGs), which serve as the foundation for developing Health Information Systems (HISs). The evaluation process may involve different methods such as the System Usability Scale (SUS)\(^1\), artificial evaluation strategies in controlled environments, and assessment through Competency Questions (CQs) followed by Description Logics (DL) and SPARQL queries as test runs. The authors adopted the FEDs approach and utilized DL and SPARQL queries to validate the CQs.

- **Communication**: The authors endeavoured to disseminate the concept of materialized knowledge graphs (MKGs) through various information sessions with domain experts or technical specialists to gather valuable feedback for further refinement. This initiative aimed to set the stage for systematically addressing data integration and interoperability challenges, particularly within the healthcare context.

---

**Fig. 1.** Customized Design Science Research (DSR) Approach

**3.1. Activity 1 & 2: Problem Identification and Motivation for Heterogeneous Data Collection - Construct Artifact**

This research is based on the healthcare context of cardiovascular disease and aims to develop an ontology-driven clinical recommended system. In the data collection phase, authors tried to incorporate the domain knowledge related to the context of cardiovascular diseases and embed production rules in the knowledge represented in the ontological metadata model OWL files for generating reasoning ability to facilitate healthcare professions (HPs) and health users. For this purpose, one challenging task is integrating various data models from different sources and formats in a decentralized environment. Resolving the data integration and interlinking, the data interoperability issues, the authors followed the steps to make this discussion more realistic:

- First, conduct a desktop research process, analyze them and create an analogy to understand the problems and their importance. In this process, authors found some exciting approaches to address data integration and interoperability issues, especially materialized knowledge graphs (MKGs), which help construct federated knowledge graphs (FKGs).
- Second, the authors also had informal discussions with domain experts and industry colleagues having promising experience in building data-driven applications and their data-driven pipelines.
- Third, The author used the sample of published data related to heart-failure clinical records data set from Kaggle\(^2\).

---


\(^2\) [https://www.kaggle.com/datasets/durgeshrao9993/heart-failure-dataset](https://www.kaggle.com/datasets/durgeshrao9993/heart-failure-dataset)
3.2. Activity 3: Defining the Objective for a Proposed Solution and Relevance

3.2.1. Data Analysis

The data analysis phase is delineated as a comprehensive five-phase process, comprising knowledge elicitation, modelling of tacit knowledge, design of the architectural artifact for federated materialized knowledge graphs (FMKGs), conceptual modelling to formulate the ontological metadata model artifact, and evaluation and testing employing Description Logics (DL) and SPARQL queries based on Competency Questions (CQs) (refer to Table 2 for further details).

<table>
<thead>
<tr>
<th>Phases</th>
<th>Tasks</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| **Knowledge Elicitation** (Construct-level Artifact) | • Articulate and define objective and research questions using a systematic approach.  
• Organize the extracted data in a structured format, such as a spreadsheet.  
• Construct competence questions (CQs).  
• Analyze ontological frameworks and semantic web techniques to solve data interoperability issues.  
• Analyze mapping techniques for data integration and resolve data interoperability issues.  
• Analyze mapping plug-ins such as “DataMaster”, “Cellfie”, “SPARQL”, “DL”, and “Ontop”, etc. | • Research data repositories.  
• Identify numerous critical concepts related to data integration and data interoperability issues.  
• Mentioned different conceptual modelling tools (e.g., MS Visio, CMap, Protege, etc.).  
• Mentioned ontological frameworks (e.g., FKGs).  
• Knowledge graphs (KGs) expressed in RDF.  
• Ontology (O) expressed in OWL2QL format.  
• Mapping (M) expressed in R2ML.  
• Query expressed in DL and SPARQL languages. |
| **Tacit Knowledge Modelling** (Neuronal-level Artifact) | • As a knowledge engineer, draw conceptual models in suitable tools and share them with knowledge mentors.  
• Create data models in diverse formats (e.g., .xml, .xlsx, .csv, .db) using various tools like Protege, MySQL, and more. | • Process view of the domain model based on cardiovascular disease context.  
• Information sets, processes |
| **Architectural Artifact of Materialized Knowledge Graphs (MKGs)** | • As a knowledge engineer, design layer architecture based on the data acquisition layer, federation ontology layer, mapping layer, virtual graph layer, and user application layer | • Architectural design artifact. |
| **Conceptual Modelling** Symbolic-level Artifact (Explicit knowledge) | • As a knowledge engineer, transfer tacit knowledge (e.g., cardiovascular disease business model) into the conceptual metadata model (machine-readable format) using owl languages. | • Conceptual meta models.  
• Define transformation rules in “Cellfie” plug-in  
• Define business rules and write in SWRL plug-in |
| **Deployment, Evaluation and Testing** | • Construction of federated ontology metadata models using MKGs approach.  
• Construction of federated MKGs  
• Construction of federated KGs  
• Evaluation and testing using DL and SPARQL query languages.  
• Test business rules in knowledge graph databases, e.g., “Ontotext” and “Neo4J”. | • Federated ontology in owl.  
• Data interoperability resolution.  
• Execution results to justify competence questions.  
• Execution of business rules in KGs. |

3.3. Activity 4: Anatomy of Federated Knowledge Graph Framework Using Materialized Approach-An Architectural Artifact

This phase describes the anatomy of the proposed framework that is referred to as federated knowledge graphs (FKGs) and also known as “Ontology-based data access (ODBA)”. This proposed FKGs is inspired by [30] work and categorized into five main categories: 1) Data acquisition layer from different sources; 2) Data mapping layer; 3) Ontology federation layer; 4) Federated virtual knowledge graph layer; 5) User application
First, the Data acquisition (DA) layer is dedicated to the data acquisition from different data channels and transforms into their respective data models following different data standards. In this layer, the authors develop a small-scale database related to the patient, which contains cardiovascular disease-related data. This small-scale database comprises patient-related tables, including clinicians, diagnosis, encounter history, medications, and patient details (demographic information).

Second, in the Data mapping (DM) layer, the authors develop the database file with extension (".db"). Similarly, the authors developed patient-related blood pressure test results in an Excel sheet with an extension (".xlsx"), and patient lab tests results file with an extension (".csv"), and lab test results also shown in XML format with an extension (".xml"), and finally domain ontology model built in OWL language with extension (".owl"). Here, the authors used an ontology editor such as Protege 5.6.1. for performing data mapping and integration operations using the “Cellfie” and “DataMaster” plug-ins. These operations can be performed using data transformation rules in the transformation rule edit panel in Protege editor and get functional support for adding, editing and deleting rules and saving and loading existing rules into the owl file. This activity aims to generate the axioms that trigger the creation of new axioms based on transformation rules that map the data from different sources, especially from Excel and CSV files and the ontology. Similarly, the authors used the “DataMaster” plug-in in Protege 3.5. version. This tool helps to map the content of arbitrary relational databases as a knowledge graph and imports a database with RDF/XML structure in Protege 5.6.1. version.

Third, the Ontology federation (OF) layer is responsible for keeping the data from different sources and performing transformation rules that help transform the data from their respective sources without physical transformation but only their structure stored in the OWL file. This layer gives the images of all the results extracted from the data as mentioned above models and domain ontology. So, this federated ontology is the image of the transformation data from different sources and domain contextual knowledge.

Fourth, the Federated knowledge graphs (FKGs) layer follows two modes: 1) Virtualized approach and 2) Materialized approach. This research targets only materialized knowledge graph design and its construction from the federated ontological metadata models. This layer also incorporates business rules that help to generate axioms with reasoning abilities so that end-users can facilitate recommendations during patient diagnosis and suggestions for further medication in the healthcare context. This layer represents the KGs in RDF format, and ontology (O) is expressed in OWL2QL, which provides domain classes, properties, individuals and data values and is stored in semantic web denouements. It also helps to showcase the mapping phenomenon in RDB to RDF (R2RML) mapping language, which helps to transform or map from a relational database to an RDF structure. It also helps to show results expressed in DL and SPAQRL query languages for better information retrieval and realization. The second materialized approach is beyond the scope of the research, and the authors will consider it in forthcoming publications.

Fifth, the User application layers define the data access and virtualization layer, which facilitates end-users and healthcare stakeholders according to their needs and demands in different data vitalization forms in the form of reports, which helps to improve decision-making, and query results, which helps to patients for improving their personalize health outcomes within the healthcare context. The federated knowledge graph (FKG) architectural artifact can be seen in the following Fig. 2.
3.4. Activity 5: Evaluation and Testing of Digital Artifacts

In this study, the authors created several digital artifacts, offering a tangible demonstration of how ontology-driven techniques and frameworks can effectively address the challenges of data integration, interlinking and interoperability.

3.4.1. Evaluation Cycle

In this phase, the developed digital artifacts from federated knowledge graphs (FKGs) undergoes evaluation utilizing the FEDs (Framework for Evaluation in Design Science Artifacts), incorporating artificial and formative strategies. This evaluation aims to substantiate the narrative of this research, demonstrating how the federated knowledge graph framework aids in enhancing data integration, interlinking and interoperability issues through the federated ontological metadata models approach. Throughout this process, the authors assess multiple digital artifacts at various evaluation stages and offer crucial feedback regarding enhancements to data-driven applications in the contemporary complex information landscape [38].

4. Experimental Results


This phase emphasizes the process of developing a Virtual Semantic View Framework (VSVF), with its specification denoted as “λ,” which comprises three primary steps:
- Modelling of the domain ontology (e.g., cardiovascular ontology (CVO));
- Generation of the local views specifications;
- Generation of the linkage rules [20].

4.1.1. Modelling of the Domain Ontology: Cardiovascular Diseases Ontology
The Cardiovascular Diseases Ontology (CVO) is a comprehensive modular-based approach utilized for constructing multidisciplinary ontology development through merging, aligning, and extending predefined ontologies. This section outlines the generic procedure for building the CVO, elucidating its crucial components, elements, and primary features. It provides detailed descriptions of the predefined and extended concepts. Additionally, this section thoroughly expounds on the validity of the proposed CVO. To enhance comprehension, Fig. 2 presents systematically the main components and development stages of the CVO at the federation layer.

**Design and Conception:** In Fig. 3, the authors detail the construction pipeline of the federated knowledge graph (FKG), emphasizing sequential steps and the system architecture necessary to develop health information systems (HISs). This section also elaborates on the generic architecture of CVO, which unfolds across various phases. The initial phase involves a comprehensive discussion, informed by the desktop research, elucidating the rationale and motivation behind adopting a modular architectural approach. Subsequently, the section presents a detailed overview of the fundamental steps in creating a CVO.

**Modular Ontology:** In the development of Health Information Systems (HISs), modularity plays a foundational and critical role, particularly in the realm of distributed ontology engineering for multidisciplinary applications like the Cardiovascular Diseases Ontology (CVO), proposed as the foundational ontological system for HISs development. To this end, CVO integrates predefined applied ontologies, including the Database ontology and various data models stored in Excel and CSV files, converted into “.rdf” formats. The integration process also encompasses the domain ontology enriched with contextual knowledge, competence ontology, and utilizing ontology design patterns (ODPs)\(^3\). The database ontology, structured around patient database models comprising five tables (*patient infrastructure, roles, patient history, diagnosis, and medications*), is developed within a relational database system such as MySQL Workbench 8.0. An enhanced entity-relationship (EER) diagram depicting this structure is illustrated in Fig. 11. The patient database undergoes a comprehensive conversion into ontology using the DataMaster plug-in within Protege, an ontology development editor. This conversion process is depicted in Fig. 13.

Similarly, various data models stored in Excel and CSV formats are transformed into ontology utilizing MappingMaster, a transformation rules language. MappingMaster DSL introduces a new reference

---

\(^3\) http://ontologydesignpatterns.org/wiki/Main_Page
clause to establish connections between spreadsheet data and ontology. This data is integrated into the ontology by adhering to a formula detailed in Fig. 14.

On another front, the domain ontology, serving as a metadata model, intrinsically aligns with the context of CVO. This ontology delineates the roles fulfilled, the activities governed by various tasks, and the types of services, particularly within the domain of healthcare professionals (HPs) and patients. The CVO ontology finds extensive utility in biomedical and bioinformatics research, aiming to construct other biomedical ontologies by reusing disease ontology terms or IDs or mapping to Disease Ontology IDs (DOIDs) for ontology cross-references, synonyms, or annotations. Disease Ontology is employed in our proposed CVO work. Additionally, the authors integrated a competence ontology into the proposed CVO, delineating various competencies of roles such as cultural competence, educational competence, general competence, occupational competence, and work experience competence. The CVO ontology comprehensively describes multiple services offered in relation to HPs and patients. Furthermore, the authors customized various ontology design patterns and aligned them within the proposed integrated CVO as part of a reuse effort, as depicted in Fig. 5.

- **Merging, Aligning and Extending Ontologies:** Fig. 4. demonstrates a conceptual view of the process involving the phenomenon of importing merging, aligning and extending within the proposed CVO. As previously mentioned, predefined ontologies serve as the foundational building blocks for constructing a proposed ontology that accurately captures the context of the cardiovascular diseases. Various operations are performed to extract all the necessary entities required for this endeavour. In particular, the merging process incorporates external source ontologies into the proposed ontology, thus enabling the reuse of existing work. For instance, Disease Ontology (DOID) is imported as part of this merging process.

![Fig. 4. Systematic approach for the development of Integrated Ontologies Alignment Process](https://github.com/abid-fareedi/Data-Integration-and-Interoperability-Project)

- **Conceptual Overview of Ontology Integration and Alignment:** In Fig. 5, the authors depict the conceptual framework of ontology, merging domain ontology and contextual knowledge into the focused

---

ontology CVO and different data models interlinked with patient data. In this section, the authors have presented a chunk of the patient database model, data models with values and the proposed CVO in graphic format. These graphical depictions establish the taxonomies of various entities and illustrate their hierarchical structure, highlighting the “is-a” relationships between parent class and child class. For instance, under the broader category of “Cardiac Diseases”, we can observe subcategories such as “Chronic Diseases”, which further break down into “Diseases”, and “Coronary Artery Diseases”, which also further break down into the “Stroke” family. We can also observe another branch, which includes “Heart failure”, which is subdivided into “Acute heart failure” and “Chronic heart failure”. Similarly, another branch is related to “Inflammatory Heart Disease”, which is subdivided into “Endocarditis Inflammation”, “Eosinophilic Myocarditis”, “Inflammatory Cardiomegaly”, and “Myocarditis”, among others. The proposed semantic model of CVO combines the different entities and their taxonomies, which contain “is-a” and “has_a” relationships for associating various entities. Here, the authors utilized personalized data properties for specific classes or entities and object properties with association links for better expressiveness and semantic understanding.

Fig. 5. Conceptual Overview of Ontology Integration and Alignment of Cardiovascular Disease Meta Model
- **CVO Validation**: In this research, the authors incorporated validated, consistent and coherent ontologies such as database ontology and Cardiovascular Diseases ontology (CVO), different data models expressed in Excel, CSV and XML formats and seamlessly integrated them into the proposed CVO semantic metadata model. To enhance the authenticity and coherence of our work, the authors conducted ontology consistency tests. These tests were carried out using the “Ontology Debugger” plugin within ontology editing tools such as Protege, as illustrated in Fig. 6.

![Fig. 6. Integrated Cardiovascular Disease Ontology (CVO) and Different Data Interlinking Alignment Test](image)

4.1.2. Virtual Semantic View Framework

Fig. 7 explains a systematic way for generating VSVF resulting from semantic integration over heterogeneous data sources $DS_1, ..., DS_n$ is a triple $\lambda = (O_D, V, L)$, where: 1) $O_D$ represents the domain ontology (e.g., CVO) and responsible for establishing a vocabulary to be shared to describe the data sources; “$V$” represents a set of local view specifications $V_1, ..., V_n$ that describes the data sources $DS_1, ..., DS_n$ using the terms in “$O_D$”. A local view specification “$V_i$” is a tuple or record $(O_{V_i}, M_{V_i})$, where: “$O_{V_i}$” is the ontology of the local view. The vocabulary of “$O_{V_i}$” is a subset of the vocabulary of “$O_D$” whose terms occur in “$M_{V_i}$”; “$M_{V_i}$” is a set of mapping rules that relate terms of vocabulary “$O_D$” with terms “$S_i$”; “$L$” is a set of linkage rules that specify virtual “sameAs” links between resources in different local views. These linkage rules are used to relate heterogeneous data sources that represent the same entity of the real-world phenomenon in the following section.

![Fig. 7. Virtual Semantic View Framework Artifact](image)
4.1.3. Generation of the Mapping Rules

Table 3., delineates a set of mapping rules or principles that underpin the process of converting structured relational data sources, such as relational databases, into ontologies [45]. These mapping rules delineate how components of relational databases, including tables, columns, foreign keys, etc., can be translated into ontology components such as Classes, Properties, Instances, etc. [39].

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule1:</td>
<td>Mapping of tables to OWL classes.</td>
</tr>
<tr>
<td>Rule2:</td>
<td>Handling of bridge tables.</td>
</tr>
<tr>
<td>Rule3:</td>
<td>Mapping of referential integrity relationships to inheritance hierarchy.</td>
</tr>
<tr>
<td>Rule4:</td>
<td>Mapping of non-referential integrity columns into datatype properties.</td>
</tr>
<tr>
<td>Rule5:</td>
<td>Representation of datatype properties host class as Domain (D) and data type as Range (R).</td>
</tr>
<tr>
<td>Rule6:</td>
<td>Mapping of relationships represented by referential integrity columns into object properties.</td>
</tr>
<tr>
<td>Rule7:</td>
<td>Representation of object property host classes as D and R.</td>
</tr>
<tr>
<td>Rule8:</td>
<td>Mapping to tuples to individuals.</td>
</tr>
<tr>
<td>Rule9:</td>
<td>Mapping of column constraints into property cardinalities.</td>
</tr>
</tbody>
</table>

Numerous algorithms and software tools such as RDBToOnto⁵, R2O⁶, D2RQ⁷, KAON2⁸, DB2OWL, DB2OntoModule, DartGrid Semantic, Automapper, MARSON, Ontology Generator (RDB2On), and DataMaster [41] are used to convert a relational database into ontology automatically [40]. Formal ontologies support the automatic recognition and processing phenomenon of such heterogeneous expressions with the ingredient of taxonomy (classes and relationships) and inference rules. Numerous languages are used to develop applied ontologies, such as RDF, OWL which depend on description logics (DL) in different ontology editors such as OntoStudioX, Apollo Mapping⁹, Swoop⁹⁰, Protége¹¹; and TopBraid Composer¹² [40]. The authors used Protége editor because of its widest use and most comprehensive use, scalability and extensible nature.

4.2. Ontology Merging: Data Integration Artifact

The authors followed the footprints of ground theories and ontology engineering principles for the development of core ontologies related to Cardiovascular diseases [42]. The following sections explain the process of building an ontological model development and data integration, and interlinking from the various data sources and include contextual knowledge related to cardiovascular diseases. Fig. 8. illustrates the process of the ontology integration phenomenon in which the authors incorporated data mapped from data sources (e.g., XML, Excel, CSV, and database (e.g., Databasefile.owl)) files into the federated OWL file along with contextual domain ontology (e.g., CardioVascularDiseaseOntology.rdf) to create a federated ontology that seamlessly integrates, interlink, and aligns with heterogeneous data sources. This federated ontology is designed to store data in RDF format and address standardized issues related to metadata to make a common language for machine-readable comprehension. Fig. 9., explains the statistics of the federated ontological model, consisting of 257 classes, 2611 axiom counts, 1811 logical axiom counts, 192 object property counts, 54 data property counts, 247 individual counts and two annotative property counts. Fig. 10. illustrates how the authors used a formal materialized knowledge graphs (MKGs) approach to build the federated ontology to encapsulate diverse data from various sources. This formal MKGs approach is one of the potential solutions to develop standard terminology. It help data integration, interoperability, and reusability with different formats, such as Turtle¹³.

⁵ https://sourceforge.net/projects/rdbtoonto/
⁷ http://d2rq.org/
⁸ http://kaon2.semanticweb.org/
⁹ https://apollomapping.com/
¹⁰ http://www.mindswap/
¹¹ https://protege.stanford.edu/
¹² https://legoegraph.com/topbraid-composer/
¹³ https://www.w3.org/2007/OWL/wiki/ PrimerExampleTurtle
RDF/XML\textsuperscript{14}, OWL/XML\textsuperscript{15}, and JSON-LD\textsuperscript{16}.

4.3. Ontology Generation from Different Data Models and Database Mapping and Data Interlinking

In this module, the authors concentrated on the first module of the proposed framework (illustrated in Fig. 2) and took EHR data source inspired by the literature ([43], [40], [30]) and “Kaggle” and converted it to ontology using a suitable tool (e.g., DataMaster). The significant contribution of this work is to construct a federated ontology that can extract all required information from all locally developed ontologies from different sources in a natural way. Moreover, the federated ontology helps to construct federated knowledge graphs (FKGs) for the user interface at a high level by providing precise information retrieval (IR) from diverse sources against user demands and needs within the healthcare context. The authors presented a framework based on federated ontology in a trial for solving semantic interoperability (SI) challenges in any distributed data landscape (e.g., distributed EHRs). The authors built MySQL-Workbench 8.0\textsuperscript{17} local patient database. Fig. 11. illustrates a screenshot of the MySQL EER diagram. Regarding the Excel sheet file, the authors generated a blood pressure test file and a CSV file that contains the result of some basic vital signs tests required for the patient, as shown in Fig. 12.

Fig. 8. Ontology Integration

Fig. 9. Statistics of Ontology

Fig. 10. Sample of Individual in Cardiovascular Diseases Ontology with Integrated Data

\textsuperscript{14}https://www.w3.org/TR/rdf-syntax-grammar/
\textsuperscript{15}https://www.w3.org/TR/owl-xmlsyntax/
\textsuperscript{16}https://www.w3.org/TR/json-ld11/
\textsuperscript{17}https://www.mysql.com/products/workbench/
4.4. Transformation Mapping Rules Artifact for Data Integration and Handle Data Interoperability Using Cellfie and DataMaster Protege Plug-ins

Regarding the rational database, the DataMaster\(^{18}\) plugin is associated with Protege [41], an ontology editing for managing database resources and constructing automatic ontologies from the MySQL 8.0 database in the Protege 3.5 version. DataMaster can support both frame-based and OWL ontologies. It also can be used with

\(^{18}\) https://protegewiki.stanford.edu/wiki/DataMaster
any relational database with JDBC/ODBC drivers. This plugin is not a database back-end and is used as a database viewer. It is used to import legacy data as a set of custom-designed database tables into Protege before doing additional knowledge acquisition or knowledge modelling exercises [43]. Similarly, the Cellfie\(^19\) plugin, a Protégé Desktop plugin for importing spreadsheet data into OWL ontologies.

![Fig. 13. The database import process of DataMaster Plugin](image13)

![Fig. 14. A Screenshot of importing Excel spreadsheet data model in Cellfie Protege plugin with Transformation Rules and generated axioms to populate in Cardiovascular diseases ontology](image14)

Fig. 13., provides an overview of the ontology construction process, highlighting data importation from a MySQL database. After the database mapping was completed, the Protege editor was used to verify it through reasoners like Pellet\(^20\). Subsequently, the process moved to the ontology population, which involves furnishing the ontology with its individual instances. For populating the ontology, the authors employed a method known as ontology mapping, as described in Al-Salhi and Abdullah [44]. They used a specific strategy to introduce individuals, their corresponding classes, and values for datatype properties into the "MappingMaster" ontology. MappingMaster is an open-source library that can be used to transform spreadsheet content into OWL ontologies. This approach utilizes a domain-specific language (DSL), and the DSL is based on the Manchester

\(^19\) [https://github.com/protegeproject/cellfie-plugin](https://github.com/protegeproject/cellfie-plugin)

\(^20\) [https://www.w3.org/2001/sw/wiki/Pellet](https://www.w3.org/2001/sw/wiki/Pellet)
OWL syntax, commonly referred to as OWL-DSL. MappingMaster introduced a new reference clause to establish connections between spreadsheet data and ontology. During this phase, Excel spreadsheets generated in the data collection process were instrumental. These sheets contained information about individuals, associated classes, and datatype properties. MappingMaster DSL was employed to integrate this data into the ontology, adhering to a formula detailed in Fig. 14. To reference any cell in a spreadsheet, the symbol "@" was used, such as "@A2," which points to cell "A2" and adds its content to the ontology, as specified in Al-Salhi and Abdullah [44].

5. Evaluation and Testing

5.1. Testing Data Heterogeneity Using SPARQL and Description Logics (DL) Queries to Evaluate Success Scenario

**Success Scenario:** Fig. 15 shows the evaluation of the success scenario. The junior practitioner role (e.g., Triage Nurse1) is assigned to Person “A,” who is involved in hospital care management procedures and performs different tasks in the hospital. Person “A” carries out several activities, such as completing the physical examination to determine medical problems for patient treatment planning. Person “A” has cultural, occupational, educational and general competencies (they are not shown for brevity). To achieve metadata standardisation and validate competency-related queries, the authors employed two essential tools: the SPARQL Query Tab[21] and the DL Query Tab[22] within Protege 5.6.1. These tools were instrumental in verifying competency questions (CQs) from the federated ontological model repository and providing a sound rationale for the for the above-mentioned success.

![Fig. 15. Evaluation of Heterogeneous Data Result from different Data Models and Database](image)

5.2. Federated Knowledge Graph Construction Artifact and Evaluation Procedure to Retrieve Customized Data and Visualization in GraphDB

The section describes a state-of-the-art mechanism of data retrieval from the federated knowledge graphs (FKGs), which incorporates diverse data from diverse data sources and diverse data structures in nature. This FKGs is developed in Ontotext[23] Graph Database, which is one of the famous graph databases such as

---

21 [https://www.w3.org/TR/sparql11-query/](https://www.w3.org/TR/sparql11-query/)

22 [https://protegewiki.stanford.edu/wiki/DLQueryTab](https://protegewiki.stanford.edu/wiki/DLQueryTab)

23 [https://www.ontotext.com/](https://www.ontotext.com/)
Neo4J\textsuperscript{24}. Ontotext Graph Database stands out as a robust and resilient graph database, excelling in RDF and SPARQL support while seamlessly handling the demands of replication clustering. Its effectiveness has been validated through numerous enterprise use cases where data loading and query performance resilience are imperative. In the following construction process, the authors imported an RDF data file (that is, the combination of domain ontology and other data-mapped ontologies) and stored in the “Data Interoperability Repository” with a total of 180 statements (tuples or records) 76 explicit, 104 inferred, with a “2.37” expansion ratio in the following Fig. 16.

Fig. 16. A Visual Graph Result Artifact Retrieved from a Federated Virtual Knowledge Graph

6. Discussion

In this section, the authors emphasized the importance of the information-sharing phenomenon with a common understanding that there is a crucial need for seamless data transformation to develop federated information systems within the healthcare sector. This data transformation from different data sources or data models is based on data mapping, data integration and interlinking approaches such as Enterprise Applications Integration (EAI) [46], Enterprise Information Integration (EII) [47], and Schema Management (SM) [48]. However, data transformation is a pivotal task, and the authors explored the desktop research review to analyze which data merging and data interlinking techniques are helpful to support the narrative of this research. This research presents a federated knowledge graph framework with its materialized knowledge graph approach, which can be seen in Fig. 2. The materialized knowledge graph approach is grounded with systematic steps to achieve federated knowledge graph (FKGs) as a multidisciplinary data repository to build health information systems (HIS). The FKGs are considered a potential solution to address persistent data integration and interoperability challenges using transformative data mapping and data interlinking techniques to fulfill today’s complex information landscape.

To support the materialized knowledge graph approach, authors tailored design science research methodology to develop design artifacts embedding with a core domain-oriented ontological metadata model focused on cardiovascular diseases. The cardiovascular disease ontology (CVO) follows a comprehensive modular-based approach used to construct multidisciplinary ontology development by merging, aligning, and extending predefined ontologies. Modular ontology engineering is based on database ontology, different data models stored in Excel and CSV files and converted into ontology, competence ontology and domain-oriented ontology,

\textsuperscript{24} https://neo4j.com/
e.g., cardiovascular diseases. The database ontology follows relational database schema (patient information, medications, roles, diagnosis, patient history) related to the patient data artifacts. In this research, authors utilized the DataMaster plug-in for mapping relational databases or schema into the ontology. This schema mapping involves mapping the structure and semantics of the data elements between different schemas or data models. It helps ensure compatibility and consistency when integrating data from heterogeneous sources or models. Similarly, instance matching involves identifying and linking similar or related data instances across datasets. The authors incorporate different data models stored in the Excel and CSV files. These files are mapped into the ontology and stored in structured “.rdf” formats. For this purpose, the authors used a Cellfie plug-in. They followed a specific strategy to introduce individuals, their corresponding classes, and values for datatype properties into the ontology called “MappingMaster.”

This approach utilizes a domain-specific language (DSL), and the DSL is based on the Manchester OWL syntax, commonly referred to as OWL-DSL. MappingMaster introduced a new reference clause to establish connections between spreadsheet data and ontology. During this phase, Excel spreadsheets generated in the data collection process were instrumental. These sheets contained information about individuals, associated classes, and datatype properties. MappingMaster DSL was employed to integrate this data into the ontology, adhering to a formula detailed in Fig. 14. The authors also developed CVO ontology and competence ontology see in detail in Fig. 5. The authors developed federated ontology. They encapsulated diverse data from different sources to build FKGs.

By adopting FKGs, organizations can streamline data integration efforts, improve cross-system data harmonization, and foster seamless data exchange with metadata standardization among diverse information sources. This research highlights the advantages of FKGs in enhancing data connectivity with semantic alignment, seamless data transformation, facilitating on-demand data retrieval to support optimization of query performance, scalability, and more effective decision-making processes within healthcare settings. In this research, ontology mapping is considered a pivot and successful method for many reasons, especially in terms of semantic interoperability, data integration, standardization and reusability:

- Semantic interoperability (SI) poses challenges that can be effectively addressed through ontology mapping techniques. These techniques empower systems to comprehend and interpret the significance of data elements spanning diverse domains. By harmonizing the semantics of ontologies, they foster interoperability and communication among heterogeneous systems.
- Data integration (DI) and interlinking can be effectively accomplished through ontology mapping techniques. These techniques facilitate the integration of data from various sources by establishing mappings between concepts and relationships. These techniques enable the consolidation and analysis of data from multiple models, leading to fresh insights and knowledge extraction.
- Standardization and reusability phenomena can be achieved handsomely using the ontology mapping technique that promotes the adoption of standardized vocabularies and ontologies that enhance data consistency, reusability, and interoperability. Aligning with established standards ensures that data can be easily shared and exchanged across different systems and applications in various domains, e.g., healthcare.

This research introduces a systematic materialized approach mechanism for developing Federated Knowledge Graphs (FKGs), serving as the foundation for streamlining the data integration process and enhancing cross-system data harmonization. The materialized knowledge graph (MKG) approach facilitates seamless data exchange with standardized metadata among diverse information sources. Initially, multidisciplinary data models are converted into a standardized format (.rdf or .owl) to create a federated ontology, amalgamating various data models and encapsulating domain-oriented ontology based on domain knowledge, such as cardiovascular diseases. This federated ontology is a prerequisite for establishing the FKGs, which could serve as the basis for developing Health Information Systems (HIS). The conversion of data models into ontology was accomplished using the Cellfie plugin within an ontology development editor such as Protege.

Similarly, to convert the patient database into an ontology or receive the data in a structured format, the authors utilized the DataMaster plugin in an ontology development editor such as Protege. The authors can receive standardized data from different information sources at the federated data layer. This federated layer helps streamline data integration and improve cross-system data harmonization, which helps develop various data-driven applications. This research also highlights the value-added creation advantages of FKGs in enhancing data connectivity and interlinking with semantic alignment, seamless data transformation, facilitating on-demand data retrieval to support optimization of query performance, scalability, and more effective decision-making processes within healthcare use cases.

This novel contribution to the research facilitates the development of various data-driven applications, including Federated Knowledge Graphs (FKGs) and Semantic Data Lakes. FKGs integrate a diverse range of
data content, thereby aiding in developing Health Information Systems (HISs) and various AI-based systems, such as conversational agents, representing a new class of information systems. Similarly, this contribution supports the development of semantic data lake applications, incorporating the Ontology-based Data Access (OBDA) concept. The data lake ingests data from different source systems in their original format in near real-time, and the ingested source data is mapped to an existing ontology to enable OBDA [49].

7. Conclusion and Future Directions

This research outlines a compelling approach to address critical challenges in the healthcare industry, particularly the need for shared information with a common understanding. The authors presented the Federated Knowledge Graphs (FKGs) Framework with a materialized knowledge graph approach (MKGs) as an innovative solution, and the research demonstrates its transformative potential in improving data integration and interoperability within healthcare settings. Adopting a Design Science Research Methodology (DSRM) tailored to develop design artifacts with a focus on Cardiovascular diseases (CVD) is a strategic approach to ensure that the framework is theoretically sound and practically applicable within a specific domain. The emphasis on semantic alignment, seamless data transformation, and on-demand data retrieval through Federated Knowledge Graphs (FKGs) highlights their significance in enhancing data connectivity and harmonization. The benefits, such as improved query performance, scalability, and more effective decision-making processes, are promising for the healthcare industry, where data accuracy and accessibility are paramount.

The research outlines a well-structured approach to improving data integration and interoperability in healthcare using FKGs. It underscores the potential of this framework to bring about positive transformations in the industry, ultimately leading to better healthcare outcomes and decision-making. This research could serve as a valuable resource for healthcare professionals and organizations seeking solutions to data management challenges in the modern healthcare landscape.

In the future, there will be a significant focus on constructing virtual knowledge graphs utilizing the "Ontop" virtual knowledge graph system. This system, known for its robust capabilities in semantic data integration and query answering over heterogeneous data sources, offers promising opportunities for advancing knowledge graph construction. The Ontop virtual knowledge graph system seamlessly integrates data from various heterogeneous sources, including relational databases, RDF stores, and web services, into a unified virtual knowledge graph representation. It accomplishes this through ontology-based data access (OBDA) techniques, which involve mapping data sources to an ontology, allowing users to query the data as if it were a single, integrated knowledge graph.

One of the key advantages of the Ontop system is its ability to provide a unified view of data across different domains and formats, enabling users to query and analyze data effectively without the need for complex data integration processes. This approach helps to improve retrieval time and cost saving and improves data interoperability and decision-making capabilities.

Additionally, the Ontop system supports advanced query capabilities, including semantic reasoning and inference, which allow users to derive new insights and knowledge from the integrated data sources. This can be particularly beneficial in domains such as healthcare, where complex relationships and dependencies between data elements need to be analyzed to make informed decisions. Adopting the Ontop virtual knowledge graph system holds great promise for future knowledge graph construction, enabling organizations to leverage their existing data assets more effectively and derive more excellent value from their data investments. Continued research and development in this area are expected to enhance the capabilities of the Ontop system further and drive its adoption across various industries and domains.

References