### **Evaluation of a Semantic Web-Based System Bridging FHIR and OMOP CDM Using Clinical Phenotyping Algorithms**

**Nan Huo, MD, PhD1, Rui Huang, MS1, Danel Stone, BS1, Guohui Xiao, PhD2, Luke V. Rasmussen, MS3, Guoqian Jiang, MD, PhD1, Robert R. Freimuth, PhD1**

**1Mayo Clinic, Rochester, MN, USA; 2Southeast University, Nanjing, China; 3Northwestern University, Chicago, IL**

**Abstract**

**Background**

HL7 FHIR and OMOP CDM are essential frameworks for healthcare data exchange and observational research, respectively. Integrating these standards is challenging due to differences in data models and query logic, leading to potential discrepancies. To address this, we developed the FHIR-Ontop-OMOP system, which enables real-time querying of OMOP data and converts results into FHIR format, enhancing interoperability between clinical and research datasets.

**Method**

The FHIR-Ontop-OMOP system integrates FHIR, Ontop, and OMOP CDM, enabling OMOP-compliant relational databases to be queried with SPARQL while retaining their relational structure. To evaluate the system, we selected five phenotypes from PheKB: Resistant Hypertension, Asthma, Hypothyroidism, Herpes Zoster, and Type II Diabetes, focusing on structured, clinically validated phenotypes with diverse data elements. We used an OMOP CDM-based version of the MIMIC-III dataset, converted from the original format using an open-source ETL tool. Complex SPARQL queries executed via FHIR-Ontop-OMOP were compared to SQL queries on OMOP CDM, focusing on data points like patient counts, medication records, diagnostic codes, and event sequences. The accuracy of data transformation was assessed by examining consistency in query results across both platforms.

**Results**

All five phenotypes yielded identical patient counts across both SPARQL and SQL queries, ensuring accuracy in data transformation. Resistant Hypertension was divided into type 1 (n=271) and type 2 (n=221), with exact matches between methods. Asthma yielded n=350 patients, and hypothyroidism counts were also consistent. Clinical data, including medication, lab results, and chronological events, were faithfully mapped, underscoring the system’s accuracy in handling complex data types.

**Conclusion**

Our findings confirm the FHIR-Ontop-OMOP system’s capability for accurate, real-time data transformation between FHIR and OMOP CDM, supporting robust interoperability. Future work will focus on simplifying implementation to promote adoption, enhancing its utility for data integration in clinical and research settings.

**Introduction**

FHIR (Fast Healthcare Interoperability Resources) is an advanced healthcare data exchange standard developed by Health Level Seven International (HL7) to address the complexities and inefficiencies associated with healthcare information exchange.[1] As a modern framework, FHIR is designed to facilitate the seamless interoperability of electronic health records (EHRs) across different healthcare systems, thereby improving the quality, accessibility, and coordination of patient care. Its adoption is expected to continue growing as healthcare organizations increasingly recognize the value of interoperable, standardized data exchange in improving patient outcomes and operational efficiency.[2]

The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) is a comprehensive and standardized data framework developed to facilitate large-scale observational research in healthcare.[3] It is a pivotal tool in modern healthcare research, providing a standardized and scalable framework for the integration and analysis of observational data. Its widespread adoption has transformed the landscape of real-world evidence generation, enabling more robust and reliable insights into the effectiveness and safety of medical interventions across diverse patient populations. However, since most clinical data are stored in relational databases, there is a pressing need to provide FHIR-based data access and query services over these databases.[4] This approach would facilitate standards-based semantic data integration, sharing, and discovery within the broader scientific research community, bridging the gap between existing data infrastructure and the demands of modern healthcare and research.

Various methods exist for connecting FHIR and OMOP, including the original Google Data Harmonization proof-of-concept project,[5] UNC CAMP FHIR,[6] custom FHIR-to-OMOP ETL processes,[7] SQL-based transformations, and MENDS-on-FHIR.[4] However, these approaches often face limitations in achieving real-time integration. For example, patient care often demands timely access to the most current data. Traditional ETL (Extract, Transform, Load) processes may introduce lag times that are not acceptable in a clinical environment where decisions must be made swiftly and accurately. With real-time integration, clinicians have immediate access to data that reflects the most recent diagnoses, lab results, and treatment plans, allowing for more informed and effective decision-making.[8] Therefore, we developed the Semantic Web-based FHIR-Ontop-OMOP system to address this gap.

The FHIR-Ontop-OMOP system enables real-time, dynamic querying of OMOP CDM data and converts the results into FHIR format, thereby facilitating seamless data exchange between research databases and clinical systems.[9] For example, real-time integration enables continuous monitoring of patient populations for clinical trials or epidemiological studies.[10] This allows researchers to track changes as they occur, adjusting study parameters or identifying emerging trends more quickly.[11] The ability to query OMOP data in real-time and convert it into FHIR format ensures that insights derived from the data are not only timely but also actionable, benefiting both clinical practice and scientific research.[12] To sum up, this system offers significant advantages, including enhanced interoperability, real-time data access, cost efficiency, flexibility, and adherence to industry standards. By enabling the seamless integration of OMOP CDM data with FHIR-based systems, FHIR-Ontop-OMOP supports both clinical and research activities, enhancing the ability to derive valuable insights from healthcare data while ensuring regulatory compliance and optimizing resource use.[13]

In this system, Ontop, an open-source virtual knowledge graph platform, plays a central role in the framework.[14] It allows for the mapping of relational data stored in OMOP CDM databases to RDF format, a semantic web standard. This platform is essential for translating the OMOP-compliant relational database structure into the semantic web representation needed for interoperability. In addition, the RDF, as a widely recognized and commonly used format for publishing structured data on the web, enables access to data without requiring knowledge of its internal storage structure. It also supports schema evolution without requiring modifications to services that utilize the data.

A key feature of systems that manage RDF data is the ability to provide a web service, known as a “SPARQL endpoint,” which allows data to be queried using the SPARQL language. RDF and ontologies are foundational to the Semantic Web, with RDF serving as the universal language for machine-processable information, while ontologies offer formal definitions that help both machines and humans understand the intended meaning of the data. Therefore, we used the Virtual Knowledge Graph (VKG) technology to set up a SPARQL endpoint.[9] The VKG approach implements the query service through query rewriting techniques, avoiding the need to materialize triples (i.e., it does not require all data to be pre-converted into RDF format).[15-17] After the VKG specification was developed, we used Ontop's command line interface to set up a SPARQL endpoint, allowing end users to interact with the endpoint using standard SPARQL tools without needing to know whether the endpoint is virtual or not.[9] The Ontop we used is the most advanced open-source VKG system, supporting SPARQL 1.1 queries, R2RML (RDB to RDF Mapping Language) mappings, as well as OWL2QL (Web Ontology Language 2 – Query Language Profile) and RDFS (RDF Schema) ontologies, and it is compatible with all major relational databases.[14 18 19] The Ontop toolkit includes the Protégé Ontop Plugin, which is used for developing VKG specifications. In our FHIR-Ontop-OMOP system, RDF is the data model, Ontop is the platform that enables RDF-style querying on relational databases, and SPARQL is the query language used to retrieve and manipulate RDF data, facilitated by Ontop’s translation of SPARQL to SQL. More technical details about the FHIR-Ontop-OMOP system have been published elsewhere.[9]

Although the FHIR-Ontop-OMOP system offers several key advantages that make it a powerful solution for integrating and querying healthcare data across different standards and platforms, it still faces several challenges and limitations that need to be addressed to fully realize its potential. These include mapping difficulties, query performance and scalability.[9] For example, when translating between the FHIR and OMOP data models, mapping difficulties(mapping inconsistencies) may occur due to the changes in the structures and representations to which the mapping is applied. One of the key issues lies in the complexity of creating and maintaining accurate mappings between these models. OMOP CDM and FHIR are designed for different purposes, with OMOP focusing on observational research and FHIR prioritizing clinical data interoperability.[20] As a result, there are often mismatches in the way concepts, relationships, and attributes are represented, which can make it difficult for users to define precise mappings. In addition, translating SPARQL queries into SQL through Ontop can introduce performance bottlenecks, especially when dealing with large datasets or complex queries. The additional layer of abstraction can lead to slower query execution times compared to native SQL queries. Moreover, as the size of the data and the complexity of queries increase, the system might struggle to maintain performance. Scaling the system to handle large-scale data in real-time could be challenging, particularly in environments with high data throughput requirements.

In this study, we systematically evaluated the FHIR-Ontop-OMOP system by examining its performance, functionality, interoperability, and usability using five previously-defined electronic phenotyping algorithms. We also assessed the system’s capability to execute complex SPARQL queries and accurately translate them into SQL queries that can be executed on OMOP CDM databases.

**Materials & Methods**

The FHIR-Ontop-OMOP system is an innovative integration framework designed to bridge the gap between traditional relational healthcare databases and modern semantic web technologies. It leverages the strengths of three key components: FHIR (Fast Healthcare Interoperability Resources), Ontop, and the OMOP (Observational Medical Outcomes Partnership) Common Data Model. By using Ontop's virtual knowledge graph capabilities, this system allows data stored in OMOP-compliant relational databases to be queried using SPARQL, a query language for RDF (Resource Description Framework), while maintaining the underlying data in its original relational format. The FHIR-Ontop-OMOP system facilitates seamless access to clinical data across different platforms, enabling advanced data integration, sharing, and analysis within the healthcare domain. This approach is particularly valuable for enabling interoperability between diverse health data sources and supporting the development of explainable AI applications in clinical research and healthcare analytics.

The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) [relational databases](https://www.sciencedirect.com/topics/computer-science/relational-database). The system consists of the following modules (from the bottom up): 1) an input module that takes input from the FHIR model ontology, the OMOP data repository, and OMOP-FHIR mappings represented by a mapping [template](https://www.sciencedirect.com/topics/medicine-and-dentistry/dna-template); 2) a CKG generation module that relies on the Ontop system to generate a virtual CKG; and 3) a semantic query module that establishes SPARQL endpoints with reasoning capability.

OMOP CDM is an open community data standard, designed to standardize the structure and content of [observational data](https://www.sciencedirect.com/topics/computer-science/observational-data) and to enable efficient analyses that can produce reliable evidence. The OMOP CDM v5.4 is defined as a collection of standardized [relational table](https://www.sciencedirect.com/topics/computer-science/relational-table) schemas in six categories: [clinical data](https://www.sciencedirect.com/topics/computer-science/clinical-data) (e.g., Person, Condition\_occurrence, Drug\_exposure), health system (e.g., Care\_site), vocabularies (e.g., Concept, Vocabulary, Domain), health economics, derived elements, and metadata.

***Computable phenotyping***

EHR-driven phenotyping is a 3-step, iterative process: defining the phenotype (authoring), executing the phenotype against some data repository (execution), and evaluating or validating the correctness of the resulting cohort (validation).[21 22] Informaticists perform the execution step, sometimes in collaboration with database analysts, to extract the cohort from the data source using SQL or other code custom-built for the specific data source. Validation is typically done by experts who review the entire patient medical record. The Phenotype Knowledgebase website, PheKB (https://www.phekb.org/), have identified many requirements for the use of computable artifacts to automate EHR-driven phenotyping. The PheKB was initiated in 2012 and has been continuously contributed to by various research teams, most notably by researchers involved in the eMERGE Network.[23]The repository contains 88 published phenotypes in various stages of development. These include: using structured and standardized data representations, using human-readable and computable representations for cohort criteria, and providing interfaces for external software with backwards compatibility. PheKB was functionally designed to enable such a workflow and has purposefully integrated tools and standards that guide the user in efficiently navigating each of these stages, from early-stage development to public sharing and reuse. PheKB has tools to enable cross-site collaboration for algorithm development, validation, and sharing for reuse with confidence.

***Phenotype Selection***

From the full collection of phenotypes in PheKB, we included those were publicly available and marked with a status of “FINAL”. We reviewed these phenotype definitions (descriptions, artifacts, and metadata) and only included those that used some structured data element (i.e., were not entirely natural language processing [NLP]-based), were an actual phenotype definition (e.g., did not simply serve as a repository to submit data), and were used in a published research study. These criteria were chosen to ensure that our analysis was conducted using only completed and clinically validated phenotype definitions. In addition, each phenotype in PheKB has a dedicated page that contains metadata curated by the phenotype authors, such as the authors’ names and research network affiliation, and the demographics to which the phenotype applies. Each phenotype optionally also includes one or more implementation reports, which provide a summary of the results for a specific implementation of the phenotype definition at a single institution. Based on the complexity of the algorithms and the prevalence of the diseases, we selected five phenotypes in our study: resistant hypertension, asthma, hypothyroidism, herpes zoster, and type II diabetes. Details of the phenotypes can be found in S1-5.

***Dataset***

For our system evaluation, we utilized an OMOP CDM-based version of the MIMIC-III (Medical Information Mart for Intensive Care) dataset, a freely accessible critical care database.[24] MIMIC-III includes a wide range of data such as vital signs, medications, laboratory measurements, clinical observations and notes, fluid balance, procedure and diagnostic codes, imaging reports, hospital length of stay, and survival data. We employed an open-source MIMIC-OMOP ETL tool to convert the MIMIC-III dataset into the OMOP CDM format for our analysis. [REF]

***Query Languages***

SQL queries for the five phenotype algorithms were adjusted based on their released SQL code to utilize a variety of tables, columns, and data types across the OMOP model. For example, the phenotyping algorithm we are using for hypothyroidism is the latest published [version from Columbia University](https://phekb.org/implementation/hypothyroidism-implementation-columbia) on the PheKB page. Although it is the most recent implementation, it is 7 years old. The OMOP CDM version in SQL is outdated and does not match the OMOP CDM version we have in our database. Specifically, the database we are using to run the phenotyping algorithm is MIMIC III patient data in the form of OMOP CDM v5.3, and it is running on a PostgreSQL database system. Thus, we need to update the elements in the published SQL query’s OMOP CDM v4 to match the 5.3 version standard. The OMOP CDM SQL query posted on PheKB’s page is utilizing CDM v4, which contains 16 tables, while the CDM v5.3 we are using contains 37 tables (Figure 1). This requires us to identify and map the corresponding elements from one table to another if they are not aligning, and to check whether the datatype of each element is properly assigned. For example, lab-related values were stored in the observation table in OMOP v4, but in the measurement table in OMOP 5. The team worked on the conversion and was able to obtain logical results from the converted SQL queries.  A series of SPARQL queries were also written to execute the same rule-based logic as the original SQL queries that extracted the EHR-based cohort.

***Evaluation Design***

For this study, we assessed whether the SPARQL query for each phenotype retrieved matching results to its corresponding SQL query. We note that each algorithm is logically separated into sub-queries based on specific data elements (see Table 2 for an example for Resistant Hypertension). We compared counts returned for each sub-query, as well as for the overall query. Additionally, we compared row-level data from the SPARQL query to that returned by SQL to check concordance. Missing data points or discrepancies were noted and investigated. The correctness of the retrieved data was further verified, ensuring that the data aligned with the clinical definitions and parameters set out for each phenotype.

**Results**

Five phenotypes from PheKB were included in our study: Resistant Hypertension, Asthma, Hypothyroidism, Herpes Zoster, and Type II Diabetes. Table 1 shows the five EHR-driven phenotypes that we used to test the faithfulness of data transformation from the OMOP CDM to the Clinical Knowledge Graphs (CKGs) in FHIR RDF. Table 2-6 shows the results of running our demonstration queries against MIMIC III data, using SQL directly against OMOP and SPARQL via the FHIR-Ontop-OMOP system. The counts for all queries are identical, ensuring faithful transformation of the tested domains. For example, the number of type 1 (N=271) and type 2 (N=221) resistant hypertensive patients were exactly same between FHIR RDF and OMOP CDM using SPARQL and SQL queries. Similarly, the patients with hypothyroidism were also same between FHIR RDF and OMOP CDM.

**Discussion**

The systematic evaluation of the FHIR-Ontop-OMOP system marks a significant step towards achieving interoperability between the FHIR and OMOP CDM standards, a long-desired goal within the standardization and research communities. The collaboration between HL7 and OHDSI to create a unified standard data model underscores the importance of such efforts in enhancing data sharing in the healthcare and research industries. The FHIR-Ontop-OMOP system, which leverages the Semantic Web, demonstrates the feasibility of real-time, automated conversion in healthcare applications enabled by semantic and structural mappings between FHIR and OMOP CDM.

Validating the faithfulness and conformance of these mappings were critical to ensuring that the FHIR-Ontop-OMOP system could reliably transform data from OMOP CDM into FHIR RDF. The results of our evaluation, which used complex EHR-based phenotyping algorithms, demonstrate a high degree of consistency between the outputs of SQL queries executed directly against the OMOP CDM and SPARQL queries processed via the FHIR-Ontop-OMOP system. For example, the consistent final counts of patients based on the five phenotyping algorithms between the outputs of SQL and SPARQL queries underscore the accuracy and reliability of the FHIR-Ontop-OMOP system in accurately transforming and querying OMOP data within the FHIR RDF framework. The fact that complex phenotyping algorithms—typically involving intricate logic and multiple data elements—yielded identical results across both query methods indicates that the system effectively preserves the semantic integrity and structure of the original OMOP data. These findings suggest that the FHIR-Ontop-OMOP system is well-suited for applications requiring robust data interoperability between FHIR and OMOP, particularly in scenarios where precise data transformations are critical, such as clinical research, decision support, and healthcare AI. Moreover, the ability of the system to handle complex queries with fidelity reinforces its potential as a valuable tool for advancing standardized data integration and analysis in diverse healthcare settings.

Just as the FHIR and OMOP data models each have strengths and limitations, SQL and SPARQL do as well. SQL is a widely used query language that excels in managing and manipulating relational databases. Its primary strength lies in its ability to efficiently retrieve, insert, update, and delete data, which is critical for performing complex operations across multiple databases. In contrast, SPARQL, a query language specifically designed for querying RDF data, operates quite differently from SQL. RDF represents data as triples (subject, predicate, object), and SPARQL is uniquely suited for conducting semantic queries and pattern-matching operations within this framework. For instance, when analyzing hypothyroidism among pregnant patients, pregnancy data is stored in two relationships: the patient (subject), has\_pregnancy\_start\_date (predicate), and the actual value of the pregnancy start date (object); and the patient (subject), has\_pregnancy\_end\_date (predicate), with the corresponding end date as the object (Figure 1). To extract medication usage information and compare it with pregnancy dates, the named graph allows efficient querying by referencing the stored relationships. However, when multiple data points need to be stored, managing the corresponding named graphs can become cumbersome, leading to potential difficulties in maintaining data integrity and ensuring query efficiency.

Moreover, we encountered several challenges when attempting to convert SQL-based phenotyping algorithms into SPARQL queries. The SQL queries often required transitional query results, which were saved in temporary tables for further processing. This workflow is difficult to replicate in SPARQL, as it is not inherently designed for creating or querying structured tables in the same manner as SQL. For instance, in a specific use case where the goal was to identify patients who had taken hypothyroidism medications while excluding those with concurrent pregnancy factors, the SQL query first generated a table to store patient IDs, along with the maximum and minimum pregnancy record dates. This table was then used to compare pregnancy dates against medication prescription dates, determining whether hypothyroidism medications were prescribed within a six-month window before and a one-year window after pregnancy. Reproducing this logic in SPARQL required the use of GraphDB to create a named graph, storing patient pregnancy records to facilitate similar comparisons (Figure 2).

**Limitations**

This study is limited by several factors. One notable limitation of this study is the use of a single MIMIC-III OMOP CDM instance for evaluation. While this provided valuable insights into the system’s functionality, it does not account for the variability that may exist when using different clinical data repositories. The reliance on a single dataset limits the generalizability of the findings to broader healthcare settings. Future work will involve a more rigorous evaluation that incorporates multiple clinical data repositories from diverse healthcare environments. This will allow us to assess the system’s performance and scalability in various real-world scenarios and further refine its capabilities. As the next step, we plan to identify multiple clinical data repositories in OMOP CDM for more rigorous evaluation, including demonstrating distributed analytics and AI applications enabled by the system. Second, this study may not fully reflect the system’s performance in real-world clinical environments, where high-volume and concurrent queries are the norm. Handling large-scale datasets or executing complex queries could introduce potential bottlenecks that were not evident in our evaluation. For instance, while the five phenotyping algorithms used in this study were limited to identifying qualified patients, real-world clinical scenarios often involve much more complex data processing and decision-making tasks. To better understand the system’s scalability and reliability, future evaluations should incorporate performance testing under high-throughput conditions that more accurately simulate the demands of a clinical setting.

**Conclusion**

In conclusion, this evaluation demonstrated that the FHIR-Ontop-OMOP system can accurately and reliably transform OMOP CDM data into FHIR RDF. Important challenges and limitations were encountered that represent areas of future work, including the need to lower barriers to adoption. Nonetheless, real-time query and transformation between data models is an essential capability to advancing healthcare interoperability and enabling AI-driven healthcare applications.

Furthermore, it is likely that the usage of the phenotyping algorithms will extend beyond professionals proficient in SQL and SPARQL. In this context, SPARQL may be preferable due to its ease of readability and writability, especially once the RDF triple structure is understood. SPARQL also offers greater flexibility in querying data, and as a standard maintained by the W3C (World Wide Web Consortium), it ensures compatibility and interoperability among various data management systems that support RDF. This makes SPARQL an invaluable tool for data integration projects across platforms that rely on semantic relationships.

As the FHIR-Ontop-OMOP system continues to evolve, further collaboration with HL7 and OHDSI, along with ongoing enhancements to the mapping process, will be crucial to for fully realizing its potential in supporting large-scale data integration and advancing precision medicine.

**Authorship contribution statement**

**Nan Huo:** Conceptualization, Methodology, Writing – original draft, Data curation, Software.

**Rui Huang:** Validation, Methodology, Data curation, Writing – original draft.

**Daniel J. Stone:** Validation, Methodology, Data curation, Writing – review & editing.

**Luke Rasmussen:** Validation, Writing – review & editing.

**Robert R. Freimuth:** Funding acquisition, Methodology, Conceptualization, Writing – original draft, Supervision.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgments**

This work was supported in part by funding from the NIH FHIRCat (R01 EB030529)

**References**

1. Vorisek CN, Lehne M, Klopfenstein SAI, et al. Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review. JMIR medical informatics 2022;**10**(7):e35724

2. Posnack S, Barker W. The Heat is On: US Caught FHIR in 2019. Health IT Buzz. 2021, 2023.

3. Künnapuu K, Ioannou S, Ligi K, et al. Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. JAMIA open 2022;**5**(1):ooac021

4. Essaid S, Andre J, Brooks IM, et al. MENDS-on-FHIR: Leveraging the OMOP common data model and FHIR standards for national chronic disease surveillance. JAMIA open 2024;**7**(2):ooae045

5. Google HCLS Data Harmonization. 2023.

6. Pfaff ER, Champion J, Bradford RL, et al. Fast healthcare interoperability resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. JMIR medical informatics 2019;**7**(4):e15199

7. Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. International Journal of Medical Informatics 2023;**169**:104925

8. Henke E, Peng Y, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An extract-transform-load process design for the Incremental Loading of German Real-World Data based on FHIR and OMOP CDM: Algorithm Development and Validation. JMIR Medical Informatics 2023;**11**:e47310

9. Xiao G, Pfaff E, Prud'hommeaux E, et al. FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common data Model. Journal of biomedical informatics 2022;**134**:104201

10. Turki H, Jemielniak D, Taieb MAH, et al. Using logical constraints to validate statistical information about disease outbreaks in collaborative knowledge graphs: the case of COVID-19 epidemiology in Wikidata. PeerJ Computer Science 2022;**8**:e1085

11. Balch JA, Ruppert MM, Loftus TJ, et al. Machine learning–enabled clinical information systems using fast healthcare interoperability resources data standards: scoping review. JMIR Medical Informatics 2023;**11**:e48297

12. Shang Y, Tian Y, Lyu K, et al. Electronic Health Record–Oriented Knowledge Graph System for Collaborative Clinical Decision Support Using Multicenter Fragmented Medical Data: Design and Application Study. Journal of Medical Internet Research 2024;**26**:e54263

13. Turki H, Rasberry L, Taieb MAH, et al. FHIR RDF-Why the world needs structured electronic health records. Journal of Biomedical Informatics 2022;**136**:104253

14. Calvanese D, Lanti D, De Farias TM, Mosca A, Xiao G. Accessing scientific data through knowledge graphs with Ontop. Patterns 2021;**2**(10)

15. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Scientific reports 2017;**7**(1):5994

16. Santos A, Colaço AR, Nielsen AB, et al. Clinical knowledge graph integrates proteomics data into clinical decision-making. bioRxiv 2020:2020.05. 09.084897

17. Gong F, Wang M, Wang H, Wang S, Liu M. SMR: medical knowledge graph embedding for safe medicine recommendation. Big Data Research 2021;**23**:100174

18. Calvanese D, Cogrel B, Komla-Ebri S, et al. Ontop: Answering SPARQL queries over relational databases. Semantic Web 2017;**8**(3):471-87

19. The virtual knowledge graph system ontop. International Semantic Web Conference; 2020. Springer.

20. Jiang G, Kiefer RC, Sharma DK, Prud'hommeaux E, Solbrig HR. A consensus-based approach for harmonizing the OHDSI common data model with HL7 FHIR. Studies in health technology and informatics 2017;**245**:887

21. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. Science translational medicine 2011;**3**(79):79re1-79re1

22. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. Journal of the American Medical Informatics Association 2013;**20**(e1):e147-e54

23. Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. Journal of the American Medical Informatics Association 2016;**23**(6):1046-52

24. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Scientific data 2016;**3**(1):1-9