Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230203

The PrescIT Knowledge Graph: Supporting ePrescription to Prevent Adverse Drug Reactions

Achilleas CHYTAS[a,1], Vlasios DIMITRIADIS[a], Giorgos GIANNIOS[b], Margarita GRAMMATIKOPOULOU [b], George NIKOLAIDIS[c] , Jenny PLIATSIKA[c], Martha ZACHARIADOU[c], Haralampos KARANIKAS[d], Ioannis KOMPATSIARIS [b], Spiros NIKOLOPOULOS [b] and Pantelis NATSIAVAS [a,1]

[a] *Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece*
[b] *Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece*
[c] *Ergobyte SA, Thessaloniki, Greece*
[d] *Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece*
ORCiD ID: Achilleas CHYTAS https://orcid.org/0000-0001-8486-011X

Abstract. Adverse Drug Reactions (ADRs) are an important public health issue as they can impose significant health and monetary burdens. This paper presents the engineering and use case of a Knowledge Graph, supporting the prevention of ADRs as part of a Clinical Decision Support System (CDSS) developed in the context of the PrescIT project. The presented PrescIT Knowledge Graph is built upon Semantic Web technologies namely the Resource Description Framework (RDF), and integrates widely relevant data sources and ontologies, i.e., DrugBank, SemMedDB, OpenPVSignal Knowledge Graph and DINTO, resulting in a lightweight and self-contained data source for evidence-based ADRs identification.

Keywords. Adverse Drug Reactions, Drug Safety, Clinical Decision Support Systems, ePrescription

1. Introduction

Adverse Drug Reactions (ADRs) have been identified as a major public health issue as they lead to huge healthcare costs and they can also be considered a significant causal factor for health morbidity and mortality [1]. Indicatively, it has recently been quantified that more than 20% of hospitalizations in a random selection of multiple hospital records is related with ADRs.

As Artificial Intelligence (AI) and other relevant technical paradigms have emerged, they have also been widely investigated to support Drug Safety (DS) [2,3]. Focusing on symbolic AI, i.e., the branch of AI which is oriented on "rule-based" knowledge schemes

[1] Corresponding Authors: Achilleas Chytas and Pantelis Natsiavas, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou – Thermi Rd, GR57001, Thessaloniki, Greece; E-mails: achytas@certh.gr, pnatsiavas@certh.gr.

and automatic reasoning upon them, Knowledge Engineering (KE) has been highlighted as a key scientific domain used to support relevant decision support systems [4]. KE includes the use of Natural Language Processing (NLP) to extract knowledge from free-text, the use of ontologies and specific data formalisms to represent knowledge - typically in the form of Knowledge Graphs (KGs) - and the use of reasoning algorithms to infer new knowledge upon the explicit statements stored in the respective knowledge base.

Clinical Decision Support Systems (CDSS) have also been developed and are currently used to support the prevention of potential ADRs during various aspects of the clinical practice (e.g., ePrescription, clinical orders etc.). Typically, CDSSs are integrated in larger healthcare systems like Electronic Health Record (EHR), Computerized Physician Order Entry (CPOE), ePrescription systems etc. The PrescIT project[2] is a nationally funded research and development initiative aiming to develop a CDSS platform to support safe ePrescription via the prevention of ADRs. To this end, a CDSS is developed and will be pilot tested in the clinical context – the consortium includes three clinical partners. PrescIT employs KE as one of its main technical paradigms and the deployment of a KG, as well as other technical components, e.g., a dynamic workflow module using Business Process Management Notation (BPMN) to employ clinically validated Therapeutic Prescription Protocols [5,6].

This paper focuses on the description of the PrescIT KG, the main module used to deploy the rules upon which the CDSS builds its "alerts" to prevent potential ADRs.

2. Methods

A modular architecture is employed for the PrescIT CDSS, according to which each module can be considered a standalone service integrated with other modules via HTTP calls. The main module elaborated in this paper is the PrescIT KG, consisting of several openly available data sources. The core of the system are the 4 major knowledge sources which are represented as KG using OWL/RDF as the main data formalism:

1. SemMedDB: A knowledge base containing information extracted from thousands of PubMed papers via NLP [7].
2. OpenPVSignal KG: More than 100 pharmacovigilance signal reports published by Uppsala Monitoring Centre, using OpenPVSignal as the main ontological model to represent them [8].
3. DrugBank: An up-to-date, widely used and free-to-access online database containing information on a variety of drugs and drug targets. It combines general useful information regarding drugs such as chemical, pharmacological and pharmaceutical data [9].
4. The DINTO ontology: An RDF based knowledge source integrating various data sources containing information about drug-drug interactions [10].

OpenPVSignal KG was created via a manually curated process with various stages of quality control, while the SemMedDB/DrugBank KGs were populated via scripts with a subset of the data retrieved from their respective data sources. This process of converting already existing data in RDF/OWL format, required significant engineering work directly related with each data source's specifics. Indicatively, for SemMedDB the original data are presented in a relational triple-based format (SQL) i.e., subject-predicate-object (s, p, o). In order to extract the desired information, the fields

[2] https://www.prescit.com

"PREDICATE", "SUBJECT_SEMTYPE" and "OBJECT_SEMTYPE" were filtered. For PREDICATE only the "CAUSES" records were selected, for SUBJECT_SEMTYPE all terms that may allude to drugs were selected, whilst for OBJECT_SEMTYPE all terms that may refer to an ADR were selected. In the end, the selected (s, p, o) were equivalent to a pattern logic like similar to "biochemical substance" "causes" "condition", also pointing to the relevant article's PubMed id. It should be noted that the SemMedDB KG also interlinks with other widely used terminologies (e.g., ATC, MedDRA, etc.) via the concept unique identifier, part of the UMLS Metathesaurus.

A simple ontological model, which also has the potential to minimize reasoning execution times, was selected for the KG design since its goal is to be utilized mostly as a data source for ADRs, DDIs and their evidence. The KG is annotated and interlinked with other data sources ontologies, not only unifying the PrescIT KG, but also for providing additional querying options to the end users (i.e., either name of a disease or MedDRA/SNOMED codes etc.). For each such aspect of the KG, the most suitable data sources were selected.

3. Results

The PrescIT KG is hosted in two triple stores, based on Virtuoso and Ontotext GraphDB platform (fig. 1). The data are consumed using a set of SPARQL queries. Indicative queries can be outlined as follows: "List known/suspected ADRs for drug A", "Given drugs A and B, list all known/suspected ADRs" etc. The produced responses include the evidence in which each data source base these claims.

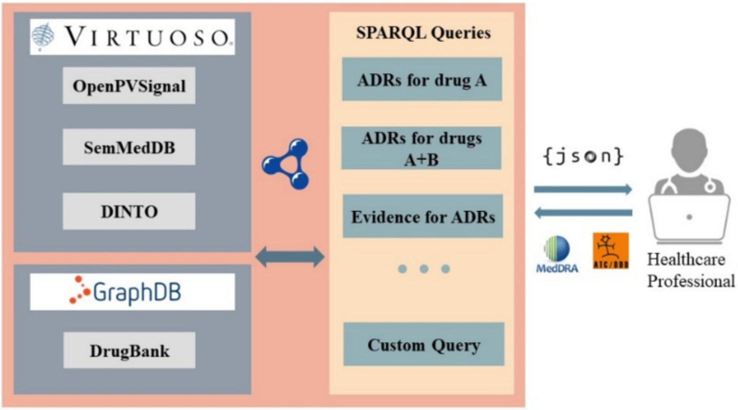

Figure 1. PrescIT architecture

To avoid direct access and enforce technical security controls, in the case of the Virtuoso database, it is exposed via an additional proxy server which was developed using FLASK, a python-based web application framework, while a similar approach is applied for the GraphDB data which are accessible via a REST API. Predefined SPARQL queries were encapsulated and exposed via specific endpoints while an additional endpoint was created that enabling the submission of custom SPARQL queries. The predefined query endpoints facilitated the data utilization without prior knowledge of the KG "schema" (TBox).

The query results are meant to be consumed by the PrescIT CDSS' front-end where they are consolidated and presented to the end-user, typically a healthcare professional. The total amount of triplets across the entire PrescIT KG is 22.5M (16M triplets from DrugBank, 108k from OpenPVSignal, 5.4M from SemMedDB+UMLS and 1M from DINTO). The KG is available upon request.

4. Discussion

Despite the hype of AI and data science, relevant technical developments have not yet been widely adopted in the clinical context. Hence, there is an increasing need for the development of new tools aiming to integrate "intelligent" technical paradigms to support drug safety.

However, the integration of "intelligent" tools in the clinical context comes with several challenges. Testing, validation and certification of such systems come with ethical, administrative, legal and technical difficulties [11]. To this end, there is an active discussion regarding how AI could be "trustable" in terms of supporting clinical operations. The PrescIT project will soon enter the pilot testing phase, evaluating the impact of the proposed CDSS and the relevant challenges, currently the technical aspects are under validation while the clinical evaluation will take place the following months.

Regarding the implementation and integration of the PrescIT KG, these challenges can be summarized as follows:

- Testing, validation and integration in the clinical environment: Clinical environments are complex and difficult to be standardized as flows of information could heavily vary, even in the same hospital. Three clinical partners have been involved during the KG development to support the definition of the relevant data sources and the PrescIT CDSS design and evaluation as a whole. For an effective integration, the PrescIT CDSS provides various flexible levels of integration with local EHRs via vendor "neutral" interfaces (e.g., SPARQL and REST APIs). Beyond the technical integration difficulties, challenges such as KG's usability and information comprehension, i.e., how well are the "alerts" received etc., due to various issues (e.g., cultural, language barriers etc.) are going to be evaluated during the pilot phase.
- Reasoning and data volume: A technical challenge directly related with the use of the KG, has to with the ability to use a "reasoner", i.e., software which is able to infer RDF statements beyond the ones which are explicitly stated in the KG. This reasoning process is computationally intensive and it has been identified as significant performance bottleneck. In order to avoid this, it was decided that no "reasoner" will be used and the "intelligence" required (e.g., the need to identify relevant RDF individuals based on subclass relationships) will be integrated in the respective SPARQL queries.

5. Conclusion

The PrescIT project aims to support ePrescription process via an integrated CDSS facilitating the prevention of potential ADRs. The PrescIT CDSS is based on the use of an RDF-based KG and its integration in the clinical environment comes with several challenges, also discussed in the paper while the impact of the proposed system and potential gaps are going to be evaluated during the pilot phase. As part of the future work there are plans of extending the KG with biochemical and pathway information which might be clinically relevant. More specifically, pharmacogenomics has been identified as a potential use case for the PrescIT KG.

Acknowledgements

This research has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T2EDK-00640).

References

[1] Formica D, Sultana J, Cutroneo PM, Lucchesi S, Angelica R, Crisafulli S, et al. The economic burden of preventable adverse drug reactions: a systematic review of observational studies. Expert Opin Drug Saf. 2018 Jul;17(7):681–95.

[2] Bates DW, Levine DM, Salmasian H, Syrowatka A, Shahian DM, Lipsitz S, et al. The Safety of Inpatient Health Care. The New England Journal of Medicine. 2023 Jan 11;388(2):142–53.

[3] Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. Trends in Pharmacological Sciences. 2019 Sep 1;40(9):624–35.

[4] Natsiavas P, Malousi A, Bousquet C, Jaulent MC, Koutkias V. Computational Advances in Drug Safety: Systematic and Mapping Review of Knowledge Engineering Based Approaches. Frontiers in Pharmacology. 2019 May 17;10:415.

[5] Natsiavas P, Stavropoulos TG, Pliatsios A, Karanikas H, Gavriilidis GI, Dimitriadis VK, et al. Using Business Process Management Notation to Model Therapeutic Prescription Protocols: The PrescIT Approach. Public Health and Informatics: Proceedings of MIE 2021. 2021 Jul 1;1089–90.

[6] Karanikas H, Papadakis M, Thireos E. Development Of Prescription E-Protocols For Medicines And Integration On The Greek National E-Prescription System. Value in Health. 2015 Nov 1;18(7):A385.

[7] Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindflesch TC. SemMedDB: a PubMed-scale repository of biomedical semantic predications. Bioinformatics. 2012 Dec 1;28(23):3158–60.

[8] Natsiavas P, Boyce RD, Jaulent MC, Koutkias V. OpenPVSignal: Advancing Information Search, Sharing and Reuse on Pharmacovigilance Signals via FAIR Principles and Semantic Web Technologies. Frontiers in Pharmacology [Internet]. 2018 [cited 2023 Jan 13];9. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2018.00609

[9] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research. 2006 Jan 1;34(suppl_1):D668–72.

[10] Herrero-Zazo M, Segura-Bedmar I, Hastings J, Martínez P. DINTO: Using OWL Ontologies and SWRL Rules to Infer Drug–Drug Interactions and Their Mechanisms. J Chem Inf Model. 2015 Aug 24;55(8):1698–707.

[11] Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. npj Digital Medicine. 2020 Dec 21;3(1):107.