Letter to the Editor: FHIR RDF - Why the world needs structured electronic health records

Houcemeddine Turki[a], Lane Rasberry[b], Mohamed Ali Hadj Taieb[c], Daniel Mietchen[d], Mohamed Ben Aouicha[e], Anastassios Pouris[f], Yamen Bousrih[g]

[a,c,e] Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Sfax, Tunisia

[a] turkiabdelwaheb@hotmail.fr [c] mohamedali.hadjtaieb@fss.usf.tn [e] mohamed.benaouicha@fss.usf.tn

[b] School of Data Science, University of Virginia, Charlottesville, Virginia, United States of America

[b] rasberry@virginia.edu

[d] Ronin Institute for Independent Scholarship, Montclair, New Jersey, United States of America

[d] daniel.mietchen@ronininstitute.org

[f] Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa

[f] anastassios.pouris@up.ac.za

[g] VERMEG France, Paris, France

[g] yamenbousrih@gmail.com

Abstract: This comment discusses the benefits of representing and reusing the information in Electronic Health Record databases as knowledge graphs in the RDF format based on the FHIR RDF specification. As a structured representation of clinical data, FHIR RDF-based electronic health records allow a simpler and more effective integration of biomedical information using semantic alignment, queries, interoperability, and federation to provide better support for health practice and research.

Keywords: Electronic Health Records, FHIR RDF, Semantic EHR, Structured EHR, Clinical Data Integration, Data reuse

Comment

We have recently read with a lot of interest the work of Xiao et al. about using the FHIR RDF specification to develop structured electronic health records (structured EHRs) (Xiao, et al., 2022). It was interesting to see FHIR RDF as a standard ontology that allows the structured representation of clinical information and consequently enables the validation of biomedical data using shape-based methods and logical constraints (Prud'Hommeaux, et al., 2021) and the reasoning over clinical knowledge using machine learning algorithms as well as SPARQL as a query language (Xiao, et al., 2022). It was also important to see how external resources, particularly the OMOP Vocabulary concepts, are used to describe clinical information about patients (Xiao, et al., 2022). However, there are multiple motivations for the development of electronic health records as knowledge graphs in the Resource Description Framework (RDF) format that have not been highlighted in this research work and that can solve several major challenges related to clinical practice. These motivations imply clinical data federation, analysis, and interoperability.

Structured EHRs can be easily federated with other biomedical knowledge resources, particularly open knowledge graphs like Wikidata (Turki, et al., 2022a) and Open Biomedical Ontologies (Konopka, 2015). As biomedical knowledge resources include background knowledge for clinical research and practice, they can be easily mirrored and then reused to drive evidence-based clinical decisions for patient management, including disease diagnosis and prognosis, risk factor identification, medical prescription, and drug interaction identification (Callahan, Tripodi, Pielke-Lombardo, & Hunter, 2020). External identifiers of diseases, symptoms, and drugs in open knowledge graphs like Wikidata can be very useful to identify resources related to the condition of the patients: just like relevant publications can be found in bibliographic databases like PubMed (Turki, et al., 2022b), records about other concepts resources like datasets, software, cell lines, biomarkers or ontologies can be located in biomedical databases like Orphanet (Ito, et al., 2015) and online encyclopedias like Wikipedia (Turki, et al., 2022a). This will permit easier access to medical information by health practitioners and facilitate medical education for patients, medical students and the general public. The use of more specialized systems like QR codes (Diazgranados & Funk, 2013) or text-to-speech systems (Baumann, Köhn, & Hennig, 2019) can facilitate access to biomedical knowledge for specific demographics, e.g. illiterate or disabled patients[1]. Another advantage of reusing concepts from other knowledge resources is that these items can be easily aligned to a multilingual ontology or knowledge graph, particularly Wikidata, and then translated into multiple languages (Turki, et al., 2022a; Rasberry, et al., 2022), allowing the easy reuse of electronic health records across countries, especially when important clinical capacities are found abroad.

Furthermore, the SPARQL endpoint of structured EHRs is not only used for the identification of patients having a particular characteristic (Xiao, et al., 2022) and the validation of biomedical data (Turki, et al., 2022c) but also for increasing the automation of health management and scholarly research. This is possible thanks to the development of specific SPARQL queries that provide insights from the EHR database and then to the embedding of the results of the queries into a dashboard (Turki, et al., 2022a). These dashboards can cover hospital administration, epidemiology, and clinical trial monitoring and are automatically updated as the electronic health records are enriched (Turki, et al., 2022a). The results of the dashboards can be visualized in a variety of layouts including bar charts, graphs, pie charts, and tables, and finally discussed through a federated SPARQL query that compares the outcomes of the real-time analysis with the findings of related scholarly publications as retrieved from the Open Research Knowledge Graph, a large-scale ontological database providing a structured representation of the outputs of research papers (Jaradeh, et al., 2019).

Moreover, as a standard for the data modeling of patient clinical information, FHIR RDF also enhances system interoperability (Girardi, De Gennaro, Colizzi, & Convertini, 2020). Such an interaction between FHIR RDF and other systems can be ensured using a distributed architecture, particularly Blockchain, for data security purposes (Girardi, De Gennaro, Colizzi, & Convertini, 2020). Interoperability involves the alignment between EHR systems so that any electronic health record can be readable by any EHR software (Saripalle, Runyan, & Russell, 2019). This will favor the exchange of clinical information between stakeholders for better and quicker support for patients when they are located outside their home region or when advanced clinical advice is required from medical scientists located elsewhere. Interoperability includes the alignment between the properties of electronic health records and the outputs of medical Internet of Things devices such as laboratory equipment, health monitoring systems, radiology instruments, electrocardiographs, and electroencephalographs (Girardi, De Gennaro, Colizzi, & Convertini, 2020). This clears the way for the automatic enrichment of electronic health records by the results of examinations, allowing a multimodal analysis of digital information for a better characterization of diseases using machine learning techniques (Huang, Pareek, Seyyedi, Banerjee, & Lungren, 2020).

To sum up, shifting to the use of electronic health records in RDF format driven by semantic resources can be a huge leap into the beginning of a new category of sustainable and explainable artificial intelligence for clinical practice that benefits not only from the latest advances in machine learning but also on semantic web technologies for healthcare data integration and reasoning. That is why we invite the biomedical informatics community to explore in more breadth and depth the integration of FHIR RDFs specification into further development of EHR systems.

[1] An example is Sawtpedia (http://sawtpedia.wiki) generating a QR code that redirects to the audio transcription or the text-to-speech output of the Wikipedia page about a particular entity in the user language.

Funding

This output is funded by the *Adapting Wikidata to support clinical practice using Data Science, Semantic Web and Machine Learning* Grant supported by the Wikimedia Foundation, the non-profit organization maintaining Wikipedia, Wikidata, and other projects. The contributions of Houcemeddine Turki, Mohamed Ali Hadj Taieb, Mohamed Ben Aouicha, and Nejmeddine Hentati are supported by the Ministry of Higher Education and Scientific Research in Tunisia (MoHESR) in the framework of Federated Research Project PRFCOV19-D1-P1.

References

Baumann, T., Köhn, A., & Hennig, F. (2019). The Spoken Wikipedia Corpus collection: Harvesting, alignment and an application to hyperlistening. *Language Resources and Evaluation, 53*(2), 303-329. doi:10.1007/s10579-017-9410-y

Callahan, T. J., Tripodi, I. J., Pielke-Lombardo, H., & Hunter, L. E. (2020). Knowledge-Based Biomedical Data Science. *Annual Review of Biomedical Data Science, 3*, 23-41. doi:10.1146/annurev-biodatasci-010820-091627

Diazgranados, M., & Funk, V. (2013). Utility of QR codes in biological collections. *PhytoKeys, 25*, 21–34. doi:10.3897/phytokeys.25.5175

Girardi, F., De Gennaro, G., Colizzi, L., & Convertini, N. (2020). Improving the Healthcare Effectiveness: The Possible Role of EHR, IoMT and Blockchain. *Electronics, 9*(6), 884. doi:10.3390/electronics9060884

Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine, 3*(1), 136. doi:10.1038/s41746-020-00341-z

Ito, M., Nakagawa, S. I., Mizuguchi, K., & Okumura, T. (2015, June). Integration of Disease Entries Across OMIM, Orphanet, and a Proprietary Knowledge Base. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 120-130). Springer, Cham. doi:10.1007/978-3-319-19066-2_12

Jaradeh, M. Y., Oelen, A., Farfar, K. E., Prinz, M., D'Souza, J., Kismihók, G., Stocker, M., & Auer, S. (2019). Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. *K-CAP '19: Proceedings of the 10th International Conference on Knowledge Capture* (pp. 243-246). Marina Del Rey, California, United States of America: ACM. doi:10.1145/3360901.3364435

Konopka, B. M. (2015). Biomedical ontologies—A review. *Biocybernetics and Biomedical Engineering, 35*(2), 75-86. doi:10.1016/j.bbe.2014.06.002

Prud'Hommeaux, E., Collins, J., Booth, D., Peterson, K. J., Solbrig, H. R., & Jiang, G. (2021). Development of a FHIR RDF data transformation and validation framework and its evaluation. *Journal of Biomedical Informatics, 117*, 103755. doi:10.1016/j.jbi.2021.103755

Rasberry, L., Tibbs, S., Hoos, W., Westermann, A., Keefer, J., Baskauf, S. J., Anderson, C., Walker, P., Kwok, C., & Mietchen, D. (2022). WikiProject Clinical Trials for Wikidata. *medRxiv*. doi:10.1101/2022.04.01.22273328

Saripalle, R., Runyan, C., & Russell, M. (2019). Using HL7 FHIR to achieve interoperability in patient health record. *Journal of biomedical informatics, 94*, 103188. doi:10.1016/j.jbi.2019.103188

Turki, H., Dossou, B. F., Emezue, C. C., Hadj Taieb, M. A., Ben Aouicha, M., Ben Hassen, H., & Masmoudi, A. (2022b). MeSH2Matrix: Machine learning-driven biomedical relation classification based on the MeSH keywords of PubMed scholarly publications. *Proceedings of the 12th International Workshop on Bibliometric-enhanced Information Retrieval co-located with the 44th European Conference on Information Retrieval (ECIR 2022)* (pp. 45-60). Stavanger, Norway: CEUR-WS.org. Retrieved from http://ceur-ws.org/Vol-3230/paper-07.pdf

Turki, H., Hadj Taieb, M. A., Shafee, T., Lubiana, T., Jemielniak, D., Ben Aouicha, M., Labra Gayo, J. E., Youngstrom, E. A., Banat, M., Das, D., & Mietchen, D. (2022a). Representing COVID-19 information in collaborative knowledge graphs: The case of Wikidata. *Semantic Web, 13*(2), 233-264. doi:10.3233/SW-210444

Turki, H., Jemielniak, D., Hadj Taieb, M. A., Labra Gayo, J. E., Ben Aouicha, M., Banat, M., . . . Mietchen, D. (2022c). 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, 8*, e1085. doi:10.7717/peerj-cs.1085

Xiao, G., Pfaff, E., Prud'Hommeaux, E., Booth, D., Sharma, D. K., Huo, N., Yu, Y., Zong, N., Ruddy, K. J., Chute, C. G., & Jiang, G. (2022). FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common data Model. *Journal of Biomedical Informatics, 134*, 104201. doi:10.1016/j.jbi.2022.104201