

Demonstration: The MeDaX-KG on FHIR

Ilya Mazein^{1,*†}, Tom Gebhardt^{1,†}, Sebastian Berthe², Kai Fitzer², Ron Henkel¹, Markus Mandalka², Seyed Reza Mazlooman², Lea Michaelis¹, Dagmar Waltemath^{1,2}, Benjamin Winter¹ and Judith A.H. Wodke^{1,*}

¹Medical Informatics Department, Institute of Community Medicine, University Medicine Greifswald

²Core Unit Data Integration Center, University Medicine Greifswald

1. Motivation

Aggregation of high-quality health data is costly and time-consuming. The Medical Informatics Initiative (MII) in Germany is dedicated to providing health data for secondary use in research, thus fostering innovation in biomedical research and the improvement of health care. A key component of the MII are the data integration centers (DIZ), established at all German university clinics for research data management and provision.

The MeDaX project aims for bioMedical Data eXploration based on graph technologies. We implement a pipeline to generate knowledge graphs (KG) from health research data [1]. Together with the DIZ of the University Medicine Greifswald, we set up the first clinical prototype of our MeDaX-KG. In parallel, we prepare the first stable release for our pipeline code. With this demonstration, we want to gather feedback from the semantic web community to further improve the pipeline.

2. Data formats

In Germany, health research data is stored and exchanged in the FHIR (Fast Healthcare Interoperability Resources) format. We generically transform FHIR data into a graph structure, optimise that structure, and integrate the resulting graph with ontological information via BioCypher, a modular framework for harmonised graph creation [2]. In comparison to our non-clinical prototype [1], we switch transformation now using the Python package networkX instead of Neo4j and its CyFHIR plug-in, thus eliminating dependency on proprietary formats. The BioCypher framework, among other features, provides access to different output adapters, including RDF (Resource Description Framework) and Neo4j. For visualisation in the clinical setting, we are using Neo4j and the SemSpect plugin. The labelled property graph format provides a desired reduction of graph complexity in comparison to RDF.

3. Challenges and solutions

The FHIR interoperability standard was designed for fast data exchange and is less suited for data analysis and visualisation due to its flat file format, usually a nested json. We develop the MeDaX pipeline to generically transform FHIR data into a graph database (GDB), established as the tool of choice when dealing with heterogeneous and complex interconnected data across scientific disciplines [3]. The pipeline generates a graph structure, optimises it and creates a generic data schema for nodes and relationships. It then automatically expands the manually predefined data schema with unspecified node and relationship types from the generic schema, informing users of the changes through logs.

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025

*Corresponding authors.

†These authors contributed equally.

✉ ilya.mazein@uni-greifswald.de (I. Mazein); judith.wodke@uni-greifswald.de (J. A.H. Wodke)

ORCID 0009-0000-1130-8332 (I. Mazein); 0009-0009-9712-060X (J. A.H. Wodke)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The resulting schema is required for KG creation with BioCypher. Privacy regulations limited the use of real-world data, leading to the use of Synthea-generated FHIR data for developing algorithms that not only generically transform the FHIR data into a graph structure, but also generically optimise this structure and update the predefined data schema to account for possibly unknown data types. Our first clinical prototype at the DIZ of the University Medicine Greifswald aims to provide a visualisation and exploration tool for our primary target groups: clinical researchers and technical experts at DIZ.

4. Conclusion and perspectives

Our graph-based solution will allow scientists to ease interaction with clinical data for research, providing a unified and intuitive way of accessing data. It will support establishing a standardised common infrastructure for managing clinical records across Germany, allowing medical practitioners to organise their information and answer work-related questions more efficiently with the help of structured GDB queries.

Currently, we are working on a user interface with different features for visualisation and exploration of the data. The feedback gathered during the demonstration will support this task. Through the integration of additional data sources, we expect to significantly increase the opportunities to generate new knowledge from already existing data.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] I. Mazein, T. Gebhardt, F. Zinkewitz, L. Michaelis, S. Braun, D. Waltemath, R. Henkel, J. A.H. Wodke, Medax: A knowledge graph on fhir, *Studies in Health Technology and Informatics* (2024). doi:10.3233/SHTI240423.
- [2] S. Lobentanzer, P. Aloy, J. Baumbach, et al, Democratizing knowledge representation with BioCypher, *Nature Biotechnology* (2023). doi:10.1038/s41587-023-01848-y.
- [3] I. Mazein, A. Rougny, A. Mazein, R. Henkel, L. Gütebier, L. Michaelis, M. Ostaszewski, R. Schneider, V. Satagopam, L. Jensen, D. Waltemath, I. A.H. Wodke J, Balaur, Graph databases in systems biology: a systematic review, *Briefings in Bioinformatics* 25(6) (2024). doi:10.1093/bib/bbae561.