



– Patient-level outlier detection using knowledge graphs

Join us in our exploratory journey to leverage healthcare data with knowledge graphs to detect outliers, improve data quality and even generate insight.

Apply until 16 April 2023 / Xplorers Camp on 08 May 2023

Question to be solved

- How can we improve the mapping of semantic concepts (e.g. patient – drug relationship) with Knowledge Graphs (KG)?
- How do we characterize outliers within patient-level data using knowledge graphs?
 - Identify potential data quality issues
 - Generate hypotheses about patient subpopulations

General background

Medical informatics provides the contextual bridge for healthcare data from source to insight, through a combination of people, process, and technology. With this project you will join a diverse group of clinical and technical experts who scale Roche's ability to leverage healthcare data across the healthcare value chain.

In healthcare, the data is generally stored in relational databases; however, taking a KG approach can leverage before unforeseen relationships, detect outliers and even improve the mapping of semantic concepts. However, one of the first steps is constructing a harmonized KG from disparate data sources and being able to validate the quality of the graph.

This exploratory project will dive deeper into knowledge graphs of patient-level healthcare data with a focus on state of the art outlier detection. Outliers may reflect (but not limited to) patients of interest, novel disease groups, disease subtypes, or data quality issues.

Data types & technologies

- Data standards
 - Common data models (e.g., FHIR, OMOP)
 - Standardized vocabularies/terminologies (e.g., ICD-10, LOINC, RxNorm, SNOMED, UMLS)
 - Graphs
 - Semantic web technologies (e.g., SPARQL, RDF)
 - Labeled property graphs
 - Graph algorithms
 - Programming
 - Python and associated ML/DL libraries (e.g., pandas, PyTorch, etc.)
 - Data Eng
 - ETL/ELT pipelines using Airflow or similar
 - CI/CD
 - Version control
 - Data science / analytics
 - Graph neural networks / graph embeddings
-

Supporting material or links

- [FHIR-Ontop-OMOP Building clinical knowledge graphs in FHIR RDF with the OMOP Common data Model](#)
 - [Graph representation learning in biomedicine and healthcare](#)
 - [Constructing knowledge graphs and their biomedical applications](#)
-

Needed skills

- Interested / experienced in Medical Informatics
 - Familiar with some of the above-mentioned data types & technologies
 - Fast-learner, passionate about continually adapting your skills and knowledge, agile, curious and IT savvy
 - Strong interpersonal, analytical and intercultural awareness
-

Mentors



Fabio Eglin
RWD Analytics Engineer



Andrew Nguyen
Section Lead, Medical Informatics

Form of cooperation

Preferred scale: 6-12 months full-time (flexible models are also possible)

Possible format: Working student, internship or master thesis

How to present your idea

Please demonstrate your approach to the problem using 3 to 5 slides. We do not expect a bullet-proof solution, we are rather interested in the way you would tackle the given challenge.