PHUSE EU Connect 2023

Paper CM01

Knowledge Graphs for Clinical Trials

Eva Kelty, Capish, Malmö, Sweden
Anders Högberg, Capish, Malmö, Sweden

ABSTRACT

As clinical trials are conducted vast amounts of data is generated, this includes data describing for example personal attributes, medical conditions, treatments, and procedures conducted together with outcome measurements.

Often outcomes/variables differ between different trials but for each specific patient data forms a web of information that is connected and may change over time, a mind-map. Property graph databases are an efficient way of managing this information where measurements and the relations between the measurements are explicit. By appending metadata to this structure, a knowledge graph is created.

In the presentation, we will discuss how the knowledge graph is a FAIR way to store harmonized and standardized clinical trial data, but also how it can be used to explore and understand data objectively. We will also show how complex questions can be answered instantly by utilizing indexing structures in a graph database.

INTRODUCTION

The R&D performance of pharmaceutical companies is dependent on a good understanding of all relevant data, not only for each study, but also for entire study programs. This sets requirements on *flexibility* and *scalability* of a database solution, to avoid the need to remodel the data repository as new data is to be incorporated [1]. Interoperability is an important component to achieve this e.g., by using a consistent approach to harmonization and integration of data, through standardization and implementation of common terminologies [2].

Furthermore, data should be modelled in an objective manner that maintains the *context* in which the data was collected. The context carries the meaning or the purpose of the data, which enables users to properly communicate, explore, analyze, draw conclusions and make decisions [3].

GRAPH DATABASE

Graph databases, as the name implies, represent data in a graph format, i.e., a network of information/ a mind-map. The main difference to traditional relational databases is that each data point (node, vertex) is explicitly connected to other data points, via relationships (edges) [1].

A property graph database has a possibility to hold an internal structure of its components, achieved by adding properties, types and labels to the nodes and relations. How the nodes and their relationships are modelled in a property graph depends on the purpose of the database. While nodes and relationships are equal partners in a graph database, models can choose to focus on either nodes or relationships. The model to choose depends on the questions you want to ask and whether the aim is to analyze the graph as a whole for analytical purposes or whether local transactions are the primary concern.

ONTOLOGY

An ontology is a formal representation of a knowledge domain, describing its entities (e.g. events and processes) as well as the relationships connecting these entities. Ontologies are often used to add semantics to the data in a graph, i.e.. the meaning and understanding of the data. No graph database can in itself provide semantics to its components. Instead, semantics must be provided explicitly, often by the use of an ontology that describes the components of the graph [1].

One example of an ontology is an information model, defining how the nodes and relations of the graph should be modelled, as well as a terminology, describing and defining all the components of the data. The model is based on a network of information units, so-called “Holons”, and their relations. Data that is tightly connected are stored together in the Holon. Examples of such data could be variables belonging to an individual event, process, or entity, e.g., a measurement result stored together with its unit, when the sample was collected, the date it was analyzed and the assessment of the result.

KNOWLEDGE GRAPH

When the property graph database is combined with the structure an ontology provides and standardized/harmonized data a mind-map of information is created. The populated Holons and the relations between them make up a knowledge/information graph that combines data and metadata in each information unit.

Working with a Knowledge Graph

A knowledge graph can be queried and cut in different ways. For example, if an adverse event of interest has been identified it is possible to easily identify all related information either by traversing the graph or by utilizing indices. For example, by applying an indexing structure it is possible to identify the patient(s) connected to the adverse event and from there find all information related to the patient(s) of interest.

In other words, reflective logic is used to formulate queries with a centricity, called reflection point, and to find all information for Holons of that centricity that fulfil the query. In clinical trial data a typical reflection point is the patient, but the study or site are also reflection points that are commonly of interest. The result from a search with reflective logic will contain all Holons that are related to all Holons of the Holon type specified as the reflection point that match the query criteria [4].

The presentation will cover live examples of different ways a knowledge graph can be cut to answer more or less complex queries utilizing a user interface.

FAIR

The ‘FAIR Guiding Principles for scientific data management and stewardship’ were published in Scientific Data in 2016. The intention was to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. A knowledge graph accessible through an interactive web-based interface makes for a solution that is well in line with the FAIR principles.

CONCLUSION

Knowledge/Information graphs is a good way to work with life science data as it makes it possible to combine data and its metadata. The knowledge graph also provides great flexibility as the graph can be built out continuously to include new types of data. Combining data, a property graph database and an ontology provides an effective structure for innovation and collaboration.

REFERENCES

1. A Berg, H Drews and C Dahlbo, “PP05- New Approach to Graph Databases”, PhUSE EU Connect 2018
2. A.E. Thessen, and D.J. Patterson, “Data issues in the life sciences”. ZooKeys, 2011: p. 15-51.
3. I. Fleming and J. Leveille, *”TT10 – The Technology of Context”*, presented at the annual PhUSE conference, 2015.
4. C Dahlbo and E Kelty, “PP14 - A Two-Step Procedure – Select and Act – How Does It Work?”, PhUSE EU CONNECT 2020

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:
Eva Kelty
Capish Nordic AB
Lokgatan 8
211 20 Malmo, Sweden

Email:eva.kelty@capish.com