Diabetes Research Using New Data Models Makes Fresh Research Connections

Neo4j's Emil Eifrem examines how Munich's DZD are studying large datasets and uncovering exciting new insights using graph database technology

The emergence of Big Data and other advances in data science approaches and tools are providing medical researchers with the opportunity for previously unobtainable insight – insights that have the potential to improve all our lives.

This is important, as the latest wave of innovation marks a sharp departure from what has become the standard way of working with data these past 30 years – spreadsheets and relational databases. The reason: neither can cope with the volume, let alone the complexity, of the multiple sources of data today's lab technicians want to explore.

By its very nature, medical data is highly heterogeneous, and so it can be a real challenge to model. And what has also been realised is that it's the relationships in data that we want to explore and uncover.

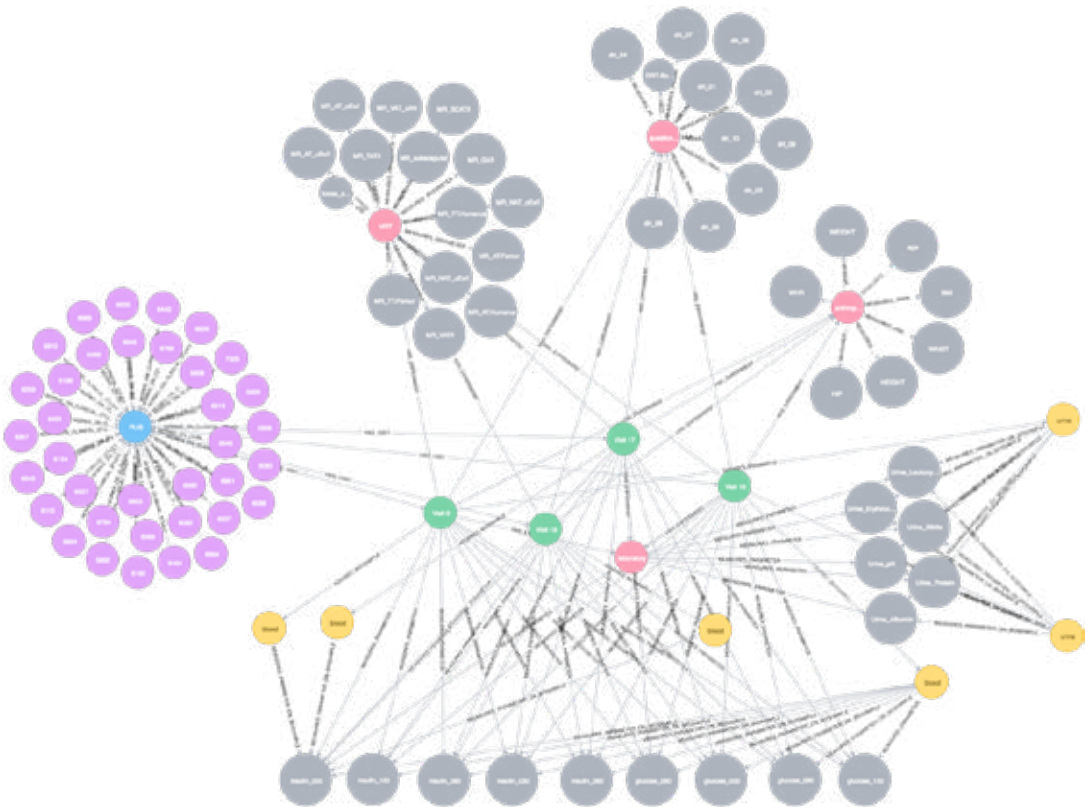
Graph database technology has appeared as a viable and powerful alternative here. One prominent convert to the approach is the German Centre for Diabetes Research, the DZD (Das Deutsche Zentrum für Diabetesforschung e.V.), which is planning to use graph software in combination with techniques such as artificial intelligence (AI) to make connections that no-one else is doing.

A New Way of Working with Data

Based in Munich, DZD brings together scientists from across the Federal Republic to develop effective prevention and treatment measures for diabetes across multiple disciplines and to see what treatment the latest biomedical technologies may offer citizens dealing with the condition.

In order to better understand diabetes' causes, the Centre's experts examine the disease from as many different angles as they can. DZD's research network accumulates a huge amount of data that is in turn distributed across various locations, and its internal IT leadership decided it needed a better way of seeing all this in the round.

DZD has built a 'master database' to consolidate this information and provide their 400-strong team of scientist peers with a holistic view of available information, enabling them to gain valuable insights into the causes and progression of diabetes.



In search of a suitable data tool to build such a system on, Dr Alexander Jarasch, the Centre's Head of Bioinformatics and Data Management, drew on experience gleaned from previous work on a project at Munich's Helmholtz Zentrum. They had used a graph database – a positive experience that prompted him to test graph technology at DZD, specifically Neo4j's software.

Dr Jarasch has now offered a new internal tool, DZDconnect, built-in graph software, that sits as a layer over the various relational databases linking different DZD systems and data silos. DZDconnect is not fully implemented yet, but scientists can already access metadata from clinical studies in the prototype – and are particularly impressed by the visualisation and the easy querying it will make possible, utilising the Neo4j 'Bloom' visualisation.

The idea is to eventually enable DZD's researchers to pose useful questions in a natural language form, such as *How many blood samples have we received from male patients under 69? Which studies are the samples from? Which parameters are measured?*

The promise here is that the more detailed the information, the easier it is to identify relationships and patterns. For the researchers, what had been a mass of abstract data can quickly fall into useful shapes: *Over what period of time was the sugar value measured? Were measurements taken on an empty stomach or after a glucose intake? How has the value changed in the long term? Can the change be attributed to a healthier diet, a drug or a hitherto unknown factor?*

Innovative Data Management

"With graph we were able to combine and query data across various locations," Dr Jarasch enthuses, adding that, "Even though only part of the data has been integrated, queries have already shown interesting connections, which will now be further researched by our scientists".

In the long term, as much DZD data as possible should be integrated into the graph database, Dr Jarasch believes, noting that the next step is to see how human data from clinical research will be complemented with highly standardised data from animal models, such as mice, to find commonalities or other insights.

It's not just graph software that is being employed. AI techniques like machine learning will play a key role going forward, says DZD, with a particular area of interest being building a system able to 'read' scientific texts and integrate them into the database ready for analysis. "Technology makes it easier to view medical issues from different perspectives and across indications," Dr Jarasch points out. "This also makes it possible to identify correlations between various common diseases."

The kind of innovative data management and analysis approach DZD is pioneering could well be the way forward in precision medicine, prevention and treatment of diabetes – and, perhaps, other diseases.

Graph technology's innate ability to discover relationships between data points could have an enormous role to play in medicine and healthcare in the future.

Emil Eifrem

Emil Eifrem is CEO and co-founder of Neo4j. Emil famously sketched out what today is known as the property graph model on a flight to Mumbai in 2000. Since then Emil has devoted his professional life to building and evangelising graph databases, and is a frequent conference speaker and a well-known author and blogger on NoSQL and graph databases.