Case Study

DZD
Deutsches Zentrum
für Diabetesforschung

German Centre for Diabetes Research (DZD)

DZD Greatly Improves Research Capabilities with Graph Technology

INDUSTRY
Bioinformatics / Medical Research

USE CASE
Knowledge Management / Master Data Management

GOAL
Unify past and concurrent diabetes research data for time savings and shared insights

CHALLENGE
Multiple data sources and heterogenous formats without a shared, central database

SOLUTION
DZDconnect graph database for central, intuitive querying of metadata powerful visualization enabling efficient research

RESULTS
– Rapid integration of DZD's clinical trial data
– Uncovering of data relationships to be further explored in new research projects

In order to give German diabetes researchers better access to data, the German Centre for Diabetes Research (DZD) is building a new graph-based knowledge management resource. Based on a Neo4j graph database, the system combines metadata from a wide range of basic research and clinical trials.

The Company

The German Centre for Diabetes Research (DZD), founded in 2009, is funded by the German Federal Ministry of Education and Research and associated federal states. It brings together experts from across Germany to develop effective prevention and treatment measures for diabetes across multiple disciplines and to benefit from the latest biomedical technologies.

The Challenge

About seven million Germans suffer from diabetes – one of the biggest common national diseases. In order to better understand its causes, DZD scientists examine the disease from as many different angles as they can.

DZD's research network accumulates a huge amount of data distributed across various locations and is working to build a master database to consolidate this information. The master database will provide its 400-strong team of scientists with a holistic view of available information, enabling them to gain valuable insights into the causes and progression of diabetes.

The Strategy

A uniform dataset needed to be created to enable access to all DZD data. Metadata had to be standardized, normalized and integrated into a graph database model in order to compare measurement results and eliminate unhelpful gaps or redundancies.

"Our goal is to set up a centralized data and knowledge management system," explained DZD board member, Professor Dr. Martin Hrabě de Angelis. "We are enabling researchers to look for clues for effective prevention and treatment measures across all disciplines, species and data types that will stop the onset or progression of diabetes."

In search of a suitable data tool to achieve this aim, Dr. Alexander Jarasch, Head of Bioinformatics and Data Management at DZD, had worked with Neo4j's graph database in the past and understood its data processing power. He chose to employ Neo4j at DZD too.

Case Study

"Creating the first data models with Neo4j was very fast. In the first week I was able to connect metadata gained by our scientists in a data model, to test it and to show the added value of the graph database."

– Dr. Alexander Jarasch, Bioinformatics and Data Management, DZD

The Solution

With support from Neo4j, Dr. Jarasch set up a new internal tool dubbed DZDconnect. The graph database sits as a layer over the various relational databases linking different DZD systems and data silos. DZD scientists so far are particularly impressed by the graph visualization and easy querying that Neo4j has made possible.

"Creating the first data models with Neo4j was very fast," said Dr. Jarasch. "In the first week, I was already able to connect metadata gained by our scientists in a data model, to test it and to show the added value of the graph database." Thanks to the high scalability and performance of Neo4j, data integration possibilities are limitless. In the future, measurement data will also be automatically transmitted to the DZDconnect tool.

In turn, this means other scientists can use this measurement information for their own research without having to repeat previous measurements and waste valuable sample material. The result is an information context in which scientists conduct parallel research without asking the same questions or repeating tests.

For the researchers, data quickly gains relevance and the Neo4j graph database helps answer rich, complex questions. The more detailed the information, the easier it is to identify relationships and patterns, and Neo4j Bloom visualization tool will enable natural-language questions in the future.

The Result

"With the Neo4j graph database, we were able to combine and query data across various locations," said Jarasch. "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."

Jarasch also noted the next step is to see how human data from clinical research will be complemented with highly standardized data from animal models – like mice – to find communalities or other insights, such as: How do factors such as diet or physical inactivity affect the disease? Are there regional or age-specific differences? Which genes play a role in the type of diabetes?

Such discrepancies, patterns or relationships are easy to recognize in the graph database. Machine learning and natural language processing (NLP) will play a key role going forward. A particular area of interest is to build 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."