

## FALLSTUDIE



**DZD**  
Deutsches Zentrum  
für Diabetesforschung

## Deutsches Zentrum für Diabetesforschung e.V. (DZD)

# Zentrale Wissensdatenbank mit Neo4j

### BRANCHE

Bioinformatik / Forschung

### USE CASE

Datenmanagement /  
Knowledge-Management

### HERAUSFORDERUNG

Standortübergreifendes  
Zusammenführen von Daten  
in eine zentrale Datenbank,  
um Wissenschaftlern einen  
einfachen und schnellen Zugriff  
zu ermöglichen

### LÖSUNG

- DZDconnect Graphdatenbank zur zentralen Abfrage von Metadaten
- Effiziente Suche dank intuitiver Abfragesprache und anschaulicher Visualisierung

### ERGEBNISSE

- Schnelle Integration eines Teils der klinischen Studien des DZD
- Aufdecken von Zusammenhängen, die in neuen Forschungsprojekten weiter untersucht werden

Um Wissenschaftlern einen besseren Zugriff auf Forschungsdaten zu ermöglichen, baut das DZD ein standortübergreifendes Daten- und Knowledge Management auf. Das neue Datenbanksystem basiert auf der Graphdatenbank Neo4j und verknüpft Metadaten aus Grundlagenforschung und klinischen Studien.

### Das Unternehmen

Das 2009 gegründete Deutsche Zentrum für Diabetesforschung e.V. (DZD) wird vom Bundesministerium für Bildung und Forschung (BMBF) und den beteiligten Ländern finanziert und bündelt auf nationaler Ebene Experten, um disziplinübergreifend und mit Hilfe moderner biomedizinischer Technologien wirksame Präventions- und Behandlungsmaßnahmen für Diabetes zu entwickeln.

### Die Herausforderung

Diabetes gehört mit rund sieben Millionen Erkrankten zu den größten Volkskrankheiten in Deutschland. Um zur Aufklärung der Krankheitsentstehung beizutragen, untersucht das DZD die Krankheit aus unterschiedlichen Blickwinkeln. Wie viele Institute verfügt der Forschungsverbund über eine riesige Anzahl an Daten, die über die verschiedenen Standorte deutschlandweit verteilt sind. Eine übergeordnete Datenbanklösung soll die Informationen zusammenführen und den 400 beteiligten Wissenschaftlern eine holistische Ansicht der Daten ermöglichen, um so neue Erkenntnisse zur Entstehung und dem Fortschreiten von Diabetes zu gewinnen.

### Die Strategie

„Unser Ziel ist der Aufbau eines zentralen Daten- und Knowledge Managements“, erklärt Prof. Dr. Martin Hrabě de Angelis, DZD-Vorstand. „Wissenschaftler sollten standort-, disziplin-, spezie- und datentypübergreifend nach Informationen suchen und wirksame Präventions- und Behandlungsmaßnahmen entwickeln, die die Entstehung oder das Fortschreiten von Diabetes stoppen.“

Auf der Suche nach einer geeigneten Datenbanklösung stieß Dr. Alexander Jarasch, Leiter Bioinformatik und Datenmanagement beim DZD, auf ein Projekt am Helmholtz Zentrum München, das mit Hilfe von Neo4j DNA-Sequenzdaten mit Metabolomics-Daten, sprich Daten zur systematischen Erfassung von Stoffwechselprodukten, verbindet und Abfragen durchführt. Das gab den Anstoß, Graphtechnologie auch im DZD einzusetzen. „Die ersten Datenmodelle mit Neo4j zu erstellen ging sehr schnell. Schon in der ersten Woche konnte ich von unseren Wissenschaftlern gewonnene Metadaten in einem Datenmodell verbinden, testen und den Mehrwert der Graphdatenbank aufzeigen“, so Dr. Jarasch. Für einen zentralen Zugriff auf alle Informationen im DZD ist zunächst eine einheitliche Datengrundlage nötig. Die Metadaten müssen standardisiert, normalisiert und in ein Graphmodell integriert werden. Nur so ist es möglich, Messergebnisse zu vergleichen und Lücken oder Redundanzen zu beseitigen.

### Die Lösung

Mit Hilfe von Neo4j machte sich Dr. Jarasch an den Aufbau von DZDconnect. Die Graphdatenbank liegt als Layer über den relationalen Datenbanken und verknüpft unterschiedliche

FALLSTUDIE



„Mit der Neo4j Graphdatenbank konnten wir in wenigen Wochen Daten standortübergreifend verbinden und abfragen. Obwohl erst ein Teil integriert ist, haben die Abfragen bereits interessante Zusammenhänge aufgedeckt, die nun von unseren Wissenschaftlern weiter untersucht werden.“

– Dr. Alexander Jarasch,  
Bioinformatik und Datenmanagement,  
DZD

Systeme und Datensilos im DZD. Noch befindet sich das Projekt im Aufbau. Über einen Prototyp können Wissenschaftler jedoch bereits auf Metadaten der klinischen Studien zugreifen. Die Forscher überzeugt dabei insbesondere die Visualisierung und einfache Suche.

Dank der hohen Skalierbarkeit und Performance von Neo4j, sind der Datenintegration so gut wie keine Grenzen gesetzt. Zukünftig sollen deshalb auch alle Messdaten automatisch ins DZDconnect übermittelt werden. Wie das in der Praxis funktionieren könnte, lässt sich anhand von Blutproben erklären: Sie sind in der Menge limitiert und daher sehr wertvoll. Misst Forschungsgruppe X auf Blutprobe 0001 den Glukoselevel A, wird dieser Wert in der Datenbank gemeinsam mit Angaben zum Lagerort und Zeitpunkt der Messung hinterlegt. Andere Wissenschaftler können diese Information für ihre eigene Forschung heranziehen und die verbleibende Probe auf weitere Parameter untersuchen, ohne bereits erfolgte Messungen wiederholen zu müssen und hierbei wertvolles Probenmaterial zu verschwenden. Es entsteht ein Informationskontext, in dem Wissenschaftler gemeinsam forschen, ohne dabei dieselben Fragen zu stellen oder Tests wiederholen zu müssen.

Über das Visualisierungstool Neo4j Bloom sollen Fragen in Zukunft auch in natürlicher Sprache möglich sein: Wie viele Blutproben von männlichen Patienten unter 69 Jahren haben wir? Aus welchen Studien stammen die Proben? Und welche Parameter wurden gemessen? Je detaillierter die Angaben, desto leichter lassen sich Zusammenhänge und Muster aufspüren. Für die Forscher können dabei unterschiedliche Informationen schnell an Relevanz gewinnen: Über welchen Zeitraum wurde der Zuckerwert gemessen? Fand die Messung im nüchternen Zustand oder nach einer Glukosezufuhr statt? Wie hat sich der Wert langfristig verändert? Und kann die Veränderung auf eine gesündere Ernährung, ein Medikament oder einen bis dahin unbekanntem Faktor zurückgeführt werden?

**Die Ergebnisse**

Das erste Fazit ist positiv: „Mit der Neo4j Graphdatenbank konnten wir in wenigen Wochen Daten standortübergreifend verbinden und abfragen. Obwohl erst ein Teil der Daten der klinischen Studien integriert ist, haben die Abfragen bereits interessante Zusammenhänge aufgedeckt, die nun von unseren Wissenschaftlern weiter untersucht werden.“ Langfristig sollen möglichst viele DZD-Daten in die Graphdatenbank integriert werden. Die nächste Projektphase steht schon fest: Humane Daten aus der klinischen Forschung sollen mit hochstandardisierten Daten aus Tiermodellen, z. B. aus Mäusen komplementiert werden, um Gemeinsamkeiten zu erkennen und Rückschlüsse zu ziehen. Wie wirken sich externe Faktoren wie Ernährung oder Bewegungsmangel auf die Krankheit aus? Gibt es regionale oder altersspezifische Unterschiede? Welcher Diabetes-Typ lässt sich auf welche Gene zurückführen? Solche Auffälligkeiten, Muster oder Zusammenhänge lassen sich im Graphen gut erkennen.

Eine wesentliche Rolle sollen künftig auch Machine Learning und Natural Language Processing spielen. Denkbar ist ein System, das wissenschaftliche Texte aus Literaturdatenbanken einliest und relevante Informationen in die Datenbank integriert und analysiert. „Die Technologien erleichtern es, medizinische Fragen aus unterschiedlichen Blickwinkeln und indikationsübergreifend zu betrachten“, erklärt Dr. Jarasch. „So lassen sich auch Zusammenhänge zwischen verschiedenen Volkskrankheiten erkennen.“

Neo4j ist das Graph-Unternehmen hinter DER Plattform für vernetzte Daten. Die Neo4j Graph-Plattform unterstützt Unternehmen dabei, vernetzte Daten in vollem Umfang zu nutzen und die komplexen Verknüpfungen zwischen Personen, Prozessen und Systemen zu entschlüsseln. Der Fokus auf Datenbeziehungen ermöglicht die Entwicklung intelligenter Anwendungen, um Herausforderungen von Heute erfolgreich zu meistern – von künstlicher Intelligenz über Betrugserkennung und Echtzeit-Empfehlungen bis zu Stammdatenmanagement. Das Unternehmen weist ein gezieltes Investment in native Graph-Technologie und mehr als zehn Millionen Downloads auf und besitzt eine riesige Graph-Community, die weltweit graphbasierte Anwendungen entwickelt. Mehr als 250 kommerzielle Kunden nutzen Neo4j, um erfolgsentscheidenden Nutzen aus ihren Datenverbindungen zu ziehen – darunter globale Branchengrößen wie Walmart, Comcast, Cisco, eBay und UBS.

- Großbritannien** [uk@neo4j.com](mailto:uk@neo4j.com)
- Frankreich** [ventes@neo4j.com](mailto:ventes@neo4j.com)
- Skandinavien** [nordics@neo4j.com](mailto:nordics@neo4j.com)
- DACH** [vertrieb@neo4j.com](mailto:vertrieb@neo4j.com)
- Südeuropa** [southern-europe@neo4j.com](mailto:southern-europe@neo4j.com)