

## Case Study



## Novartis

# Novartis Captures the Latest Biological Knowledge for Drug Discovery

**INDUSTRY**

Healthcare / Life Sciences

**USE CASE**

Knowledge Graph

**GOAL**

Accelerating the process of developing life-saving medicines

**CHALLENGE**

Finding associations across vast experimental data and medical research

**SOLUTION**

Create a graph that captures biological knowledge, linked to research

**RESULTS**

- Empower researchers to correlate experimental data with medical research
- Identify promising compounds to accelerate drug development

*Novartis needed a way for researchers to ask important questions and get rapid answers, informed by medical research. The Novartis biomedical research team uses Neo4j to mine and organize huge volumes of biological data that support development of the next generation of medicines.*

**The Company**

Novartis is a global healthcare company based in Basel, Switzerland that provides solutions to address the evolving needs of patients. It is one of the largest pharmaceutical companies by both market capitalization and sales. The Novartis Institutes for BioMedical Research comprise the innovation arm of Novartis, with 6,000 researchers at six locations around the globe.

**The Challenge**

Novartis has amassed decades of data on how various compounds affect protein targets, such as enzymes, with about a billion data points in all. That historical data is critical but sparse compared with the amazing granularity of the data currently being collected.

Today, Novartis uses an automated process that captures high-content image data showing how a particular compound has affected an entire cell culture. This generates terabytes of phenotypic data.

Novartis faced the challenge of combining its historical data stores with this burgeoning phenotypic data. They also needed a way to place all this data within the larger context of ongoing medical research from around the world.

The Novartis team wanted to combine its data with medical information from NIH's PubMed. PubMed contains about 25 million abstracts from some 5,600 scientific journals.

The Novartis team sought a way to empower researchers to ask questions connecting the dots between all of this data in the context of the latest medical research.

As Stephan Reiling, Senior Scientist at Novartis, put it, "When we try to analyze this data, it becomes much more apparent that we need to have a way to store biological knowledge and then run queries against it."

**The Solution**

Ingesting and connecting data about diseases, genes and compounds – along with identifying the nature of the relationships between these elements – held the promise of accelerating drug discovery.

## Case Study



"In merging this data together, we're creating a giant graph to better understand the biology and how we can use this scientific knowledge to develop the next generation of medicines."

– *Stephan Reiling*  
Senior Scientist at Novartis

The Novartis team wanted to link genes, diseases and compounds in a triangular pattern. "For successful drug discovery, you need to be able to navigate this triangle," explained Reiling. The Novartis team decided to create a knowledge graph stored in Neo4j, and devised a processing pipeline for ingesting the latest medical research.

Text mining is used at the beginning of the pipeline to extract relevant text data from PubMed. That data is then fed into Neo4j, along with Novartis's own historical and image data. The data pipeline populates the 15 kinds of nodes that were devised to encode the data. The next phase fills in the relationship information that links the nodes together. The team identified more than 90 different relationships.

Novartis uses Neo4j graph algorithms to traverse the graph and identify a desired triangular node pattern linking the three classes of data together. Graph analytics not only find relevant nodes in the desired triangular relationship, but also employ a metric the team designed to gauge the associated strength between each node in each triangle. Using this capability, the team devised queries to find data linked by the desired node pattern, with a given association strength, and then sort the triangles according to this metric.

When researchers query the knowledge graph, results show the strength of the correlation between elements. If a researcher already knows about a highly associated correlation, they might choose to investigate others, which could take their work in new directions.

### The Results

Neo4j enables Novartis to flexibly navigate all of their data sources, which is a boon to research.

"In merging this data together, we're creating a giant graph to better understand the biology and how we can use this scientific knowledge to develop the next generation of medicines," said Reiling.

The Neo4j knowledge graph captures elements needed for an operational system of biological understanding that continues to grow with expanding medical science. There are currently **half a billion relationships in the knowledge graph, and the team expects to easily triple that number** as data is added.

Using Neo4j, researchers see which compounds and genes are most closely associated with diseases, and drill down into the medical literature to examine the evidence for the association. "There is a huge amount of biological data available, along with incredible data sources. By really bringing all this data together, for the first time we can say, 'I want to find compounds similar to this compound that have annotations about this disease.' The flexibility to navigate all of these data sources is really powerful," said Reiling.

Neo4j is the leader in graph database technology. As the world's most widely deployed graph database, we help global brands – including [Comcast](#), [NASA](#), [UBS](#), and [Volvo Cars](#) – to reveal and predict how people, processes and systems are interrelated.

Using this relationships-first approach, applications built with Neo4j tackle connected data challenges such as [analytics and artificial intelligence](#), [fraud detection](#), [real-time recommendations](#), and [knowledge graphs](#). Find out more at [neo4j.com](#).

Questions about Neo4j?

Contact us around the globe:  
[info@neo4j.com](mailto:info@neo4j.com)  
[neo4j.com/contact-us](https://neo4j.com/contact-us)