

CASE STUDY

Integrating GXP and public data: a case study on pharmaceutical innovation

Use cases in:

- Competitive Intelligence
- Target Identification
- Substances & Product Overview



DISCOVER



Industry
Pharmaceutical



Number of employees
+60,000



Location
Global

Our customer

Our customer is focusing on four therapeutic areas: neurology and immunology; oncology; fertility; and cardiology metabolism and endocrinology, with a clear ambition to become a global specialty innovator.

Challenges

In recent years, our customer has doubled down on data visibility, data access, and data democratization. Initially they invested in knowledge graph databases, the underlying technology for bringing together different internal data sources which in turn serve as use cases on specific applications.

They needed help to streamline their own knowledge graph of internal data and supplement it with a richer database of public knowledge. The challenge was to combine their strictly controlled GXP infrastructure with less strictly controlled public data sources to increase the context of the data and serve it up for end-users and machine users. They needed to find a way to achieve this that allows their researchers to easily query both, while keeping the resources separate but seamlessly connected.

Outcome

ONTOFORCE started working with them to expand and build out the semantic layer, combining their underlying graph database technology with the DISCOVER technology that brings in public data. This combines all the various dictionaries for capabilities, taxonomies, and ontologies, combined with both an easy-to-use user interface and application programming interface (API).

Initial data integration and harmonization

The first step was to define where to start. Where were they missing information? Where does information need to be cleaned up? This was defined through a series of workshops to allow an effective prioritization on how to approach such a large journey of data transformation.

ONTOFORCE began operating within the customer's enterprise ecosystem, considering the connection to both upstream and downstream systems and integrating with necessary security measures. A wide range of sources were integrated, including internal relational databases like Oracle, public databases and API connections from external data providers such as Clarivate and Informa, as well as graph databases and different types of graph plugins that allowed the customer to automate a variety of processes. This was all incorporated into a platform that was fully branded according to the customers own internal IT architecture.

Identification of medical products (IDMP)

The first use case in the regulatory domain was an IDMP case (identification of medical products) which is a new strategy driven by the European regulator EMA. As guidelines for IDMP evolved, ONTOFORCE combined the various data domains together in the graph database according to the five relevant ISO standards. Then the DISCOVER non-GXP layer was added on top, providing connections and mappings between all the different data points. This is not only a richer layer, but it is easier and faster to bring it together because it is not subject to the same controls and regulations as the GXP system.

Data democratization was enabled through a unique way of extracting the data from the underlying operational layer, adding in the data model, and applying semantics and ontologies in a way that allows the non-data experts to query and retrieve relevant and complete information. This means that the data can be navigated effectively across the entire pharmaceutical value chain in a much more automated and coherent way. In gaining this visibility of the data quality across internal systems and data, as well as visibility of licensed data, it became clear where further investment and effort was needed for data cleanup.



Transforming data into knowledge:
DISCOVER seamless integration of GXP and public databases
empowers researchers to drive innovation and excellence in
neurology, oncology, and beyond.

In silico target identification

A further beneficial outcome of this data integration was a project based around in silico target identification. Information on genes, proteins, and diseases was aligned and analyzed in downstream algorithms. The benefit of DISCOVER here was to harmonize and provide a clear overview of the data, while offering control and flexibility in data modeling that was specifically tailored to their needs, offering much more visibility and transparency into how the data were harmonized than would a standard service solution. That target identification is combined with a competitive landscaping to provide insights into what can be done from a biological and chemical perspective and also where it makes economic and scientific sense to focus, based on the company's specific expertise.

Benefits



DISCOVER seamlessly integrates disparate data systems

Combining a highly regulated GXP knowledge graph of data with less strictly controlled public data sources in a way that allowed researchers to query both easily, keeping the sources separate but seamlessly connected.

DISCOVER offers transparency and control over the data

Through close collaboration and development with ONTOFORCE, DISCOVER provides flexible data modeling and data ingestion that is fully transparent and under the customer's complete control.



DISCOVER paves a faster road to discovery

Using a lot of prebuilt and prepackaged public data allowed for fast integration, decreasing the time needed to build upon the business request. DISCOVER facilitates the use of human and machine interfaces combined with the application of the FAIR data principles of being findable, accessible, interoperable and reusable.

DISCOVER facilitates competitive intelligence

Integrating internal and external licensed data with public data allows competitive landscaping during target identification to pinpoint the best development opportunities.



Final statement

"DISCOVER's intuitive user interface makes it pretty easy to onboard users and for them to start exploring the internal and external data. ONTOFORCE does not just provide a framework, but a massive package of pre-linked external data that you can immediately plug your internal data into. It can change the mindsets of colleagues who are used to working in data silos and not connecting data," states our customer's data strategist.

About ONTOFORCE

ONTOFORCE helps life sciences organizations accelerate research and drug development for improved patient outcomes by unlocking hidden insights from data. Founded on semantic technology and an ontology-based knowledge graph, their flagship product, DISCOVER, is a data and knowledge platform developed specifically for the life sciences industry.

DISCOVER seamlessly connects an organization's internal, siloed data with licensed data and public data in one easy-to-use, customizable platform, enabling efficient data exploration and analysis.

+10,000

DISCOVER
users

13

Awards
won

12

Years of
experience

+120

Data sources
and integrations

Book a meeting via this link or email us at hello@ontoforce.com

Locations

Belgium office

108 Moutstraat street
9000 Gent
+32 93968007

USA office

955 Massachusetts Ave #312
Cambridge, MA 02139
+1 415 996 8988

Book a demo →

Transform data into knowledge