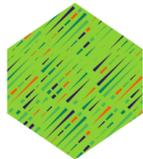


# Semantic Modelling & Knowledge Graphs

Many companies and research institutions **struggle with integration** of their research data; data lives in **silos** and vendor solutions, and is scattered over departments. In addition, there is often **no conformance** to a shared standard for experimental data.

This is a missed opportunity as annotating and integrating data enables scientists to answer broader research questions and therefore increases the value of their data.



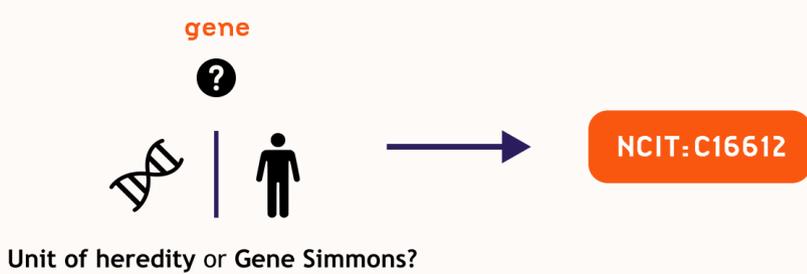
## SEMANTIC MODEL: BRINGING CONTEXT TO DATA

The semantic model is a formal specification of **domain specific entities** (e.g. cell line, assay) and their relations, with formal semantics providing **context and meaning**.

The semantic model can be applied as a sort of **glue between disparate data sources**, so they can be described as how they fit together.



By using **ontologies** – which can be seen as the data schema of the graph – to describe the entities and relationships in a domain, it enables the semantic model to be understandable by both people and machines.



*In contrast to humans, computers do not understand context. Ontologies enable a common understanding.*

## SEMANTIC MODEL: REQUIREMENTS

There is first a need for **data landscape exploration** to discover all data assets for integration. Relevant entities can then be created in the semantic model.

Other required resources include **data engineers with domain expertise**.



## BENEFITS OF USING A SEMANTIC MODEL



**Common understanding of structure and domain**

→ Serve as reference for **data conformance**



**Integration of siloed data**

→ Wide search and traceability

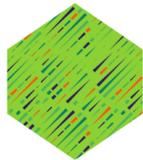


**Software independent, yet interoperable**

→ Broadly applicable



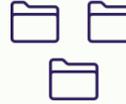
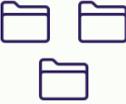
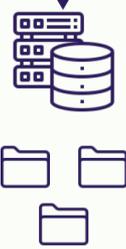
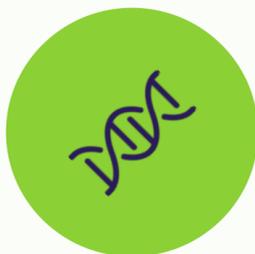
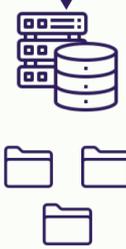
**Enterprise knowledge management**



## KNOWLEDGE GRAPH: INTEGRATING DISPARATE DATA

The semantic model is at the heart of a knowledge graph. A knowledge graph **organises data from multiple sources**, captures information about entities of interest in a given domain, and **forges connections** between them that are enriched with formal semantics provided by the semantic model. This way it establishes a **flexible framework for data integration, harmonization, analytics, and sharing**.

## LIFE SCIENCES DATA IS SEPARATED IN SILOES



*Research data is stored in various systems and applications, which are often disconnected and not integrated.*

## SOLVING A DATA INTEGRATION CHALLENGE

The Hyve has helped to solve a **data-integration challenge** for a global top-10 pharmaceutical company with drug discovery and development programs in several therapeutic areas.

We first created a **semantic model using public domain ontologies** that captures the IOCT research and business domain, which interoperates with other semantic models in R&D. This then served as a foundation for generating a **knowledge graph**, using **data from different systems** within the company. We demonstrated that **end-to-end use cases across research domains and departments** can be answered by querying this knowledge graph.

## OUTCOME

The finished semantic model revealed that certain information was missing in various vendor systems to fully understand and integrate their research data, thus the model now serves as a **reference for modeling newly generated research data** and is used to **integrate research data assets** into the company's enterprise knowledge graph.