

# Data Landscape Exploration with The Hyve

Systematic assessment of how an organization creates, receives, processes and disseminates data

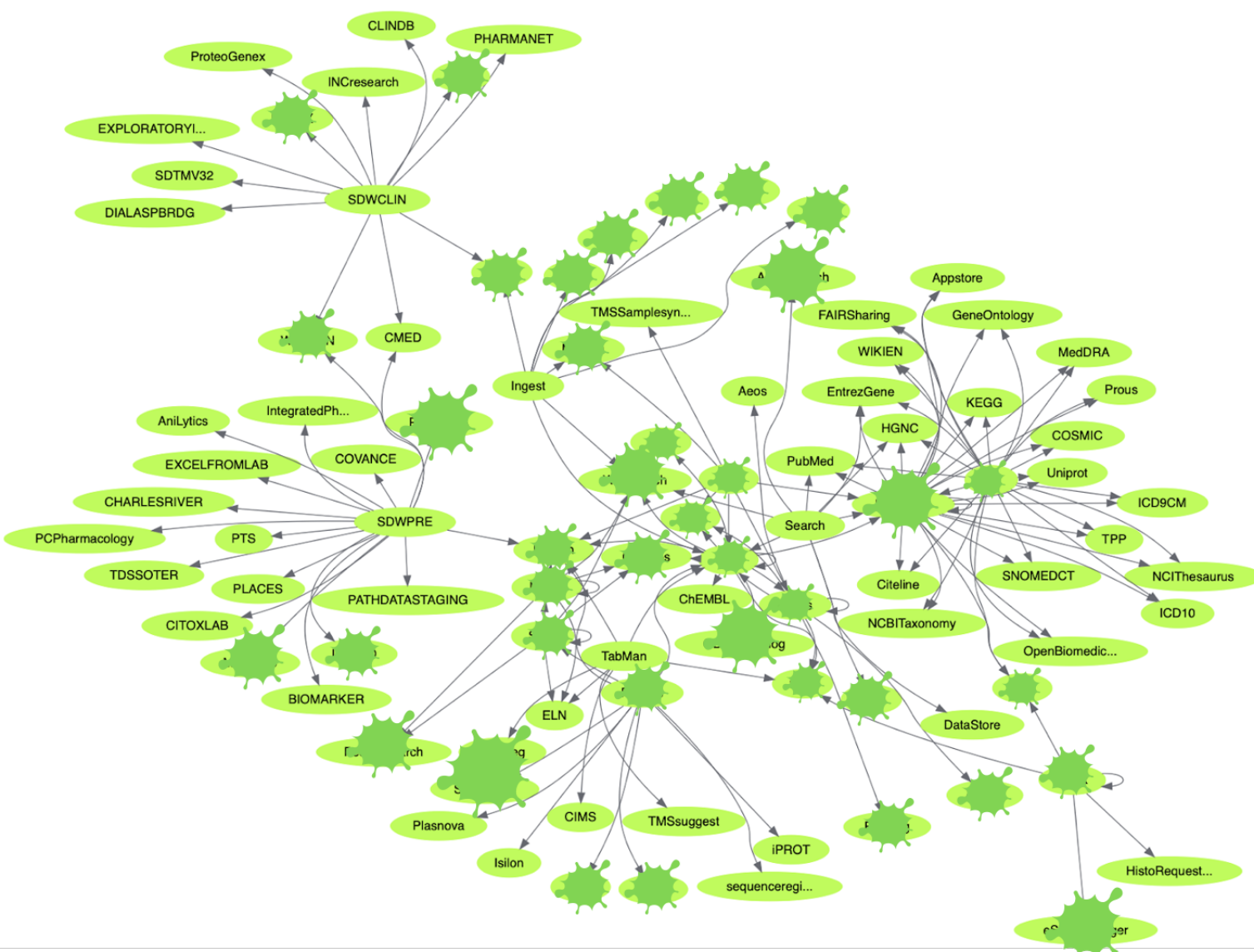
## The challenge

- How much data do we have and how is it connected? Which R&D data assets are available internally?
- Which datasets have the highest business impact?
- Which departments work on omics? How and where do they fit in the context of 4D Maps (Drug Discovery, Development and Deployment)?



These are the kinds of questions the Data Landscape Map answers which The Hyve built for a Top 10 pharmaceutical company. Our client had already exerted efforts to create data models and improve the information and data management of their Research & Development activities. We used this as input for a knowledge graph. This graph enables everyone in the company to perform a search based on (meta-) data spanning from early-stage research to late-stage clinical trials. The knowledge graph leveraged a custom semantic model of the data and the application dependencies. The Hyve also paved the way for adoption of the FAIR principles by this company, assisting to create a consensus on data usage and the company culture required for FAIR data management.

## The outcome



"Semantic representation of clinical trials and data provenance"

## Business impact

- Interviewed 50 stakeholders across several company divisions
- Inventorised data systems to a master list of 1000s from all available sources
- Worked across multiple geographical locations and departments
- Prioritized few dozen key systems cover highest impact data
- Developed a single conceptual model reflecting data landscape
- Established metadata elements in the model for top 50 systems

## Goals

- To understand quantitative and qualitative aspects of existing data assets
- Map relationships between the data assets

## Problems

- Data across stages of R&D is unlinked thus there is no clear overview of the data and it is not findable.
- Data accessibility is an issue because the associated metadata is either unclear, incomplete or entirely missing.

## Solutions

- Data landscape visualizing explicit and implicit connections between data sources, systems and owners/users.
- Semantic representation in RDF/OWL of data flows and applications
- Instantiated semantic model with research and clinical (meta)data to enable querying

## Results

- Queryable knowledge graph with ~14M triples

\* Read more about our Data Landscaping services at <https://thehyve.nl/services/data-landscape-exploration/>

## Project team



**Kees van Bochove**  
Senior Consultant



**Ilaria Maresi**  
Semantics Data Engineer



**Sukhi Singh**  
Knowledge Scientist/  
Consultant