ONTOFORCE

Leveraging knowledge graphs to overcome omics data challenges

Life sciences companies, especially pharmaceutical organizations, generate vast amounts of omics data as they research and develop new treatments. This data comes from a variety of sources, including genomics, proteomics, and metabolomics, and is critical for understanding disease mechanisms and identifying potential drug targets.

However, managing and analyzing omics data can be a daunting task, particularly when dealing with large amounts of complex datasets. In this blog post, we will explore some of the major challenges related to omics data and how life sciences companies can use knowledge graphs to overcome these key challenges associated with omics data.

IS DATA INTEGRATION ONE OF THE BIGGEST OMICS DATA CHALLENGES FOR THE LIFE SCIENCES INDUSTRY?

A survey was recently conducted by ONTOFORCE with 145 people within the life sciences industry. The survey asked individuals to identify their challenges when working with omics data. The results indicate that a variety of challenges exist across the different stages of data management and analysis.

The most commonly reported challenge is data integration, with 24% of respondents indicating that this is a significant obstacle. This is not surprising, given that omics data can come from multiple sources and may require extensive preprocessing and harmonization to be usable for analysis.

Data standardization is also a significant challenge: 14% of respondents reported that this is their most pressing omics data challenge. This underscores the importance of establishing common vocabularies and formats for omics data to facilitate data sharing and secondary analysis.

Other notable challenges included general data management (15%), data quality (4%), and data governance (4%). These challenges suggest that there is a need for effective data management practices and protocols to ensure that omics data is accurate, reliable, and accessible.

In addition, several respondents reported challenges related to the size (8%) and complexity of omics data (5%). These challenges may require the use of advanced computational methods and tools to effectively analyze and manage the data. Further, respondents also reported challenges with data analysis, data exploration, and data pipelines, suggesting that these areas may require additional attention and resources to effectively leverage the full potential of omics data.

These results highlight the diverse challenges that exist when working with omics data and emphasize the need for effective data management practices, standards, and protocols to enable the effective

integration, analysis, and interpretation of this complex data. A knowledge graph platform could be such a solution to support this need.

LEVERAGING KNOWLEDGE GRAPHS WHEN WORKING WITH OMICS DATA

A knowledge graph stores information in a structured format, making it easier for reasoning, understanding, and retrieving connections between data. It is a graph-based data model that represents knowledge as a set of entities (nodes) and their relationships (edges). Each entity is defined by its attributes, and each relationship is defined by its type and direction.

In the pharmaceutical industry, knowledge graphs are used to store and analyze large amounts of data across the drug development timeline, from target identification to clinical trial design. These graphs can include information on drugs, genes, proteins, diseases, and clinical trials, among other things.

Knowledge graphs can help researchers and data scientists identify relationships and patterns between different types of omics data, which can lead to new insights and discoveries. Additionally, knowledge graphs are especially helpful in mitigating the various challenges relative to dealing with the vast amounts of omics data within an organization. ONTOFORCE's data and knowledge management platform, DISQOVER, is built on knowledge graph technology and helps users to simplify the processes of managing and analyzing omics data.

Knowledge graphs, and especially the DISQOVER platform, are highly beneficial when it comes to addressing some of the main challenges related to omics data: integration, standardization, and data exploration.

KNOWLEDGE GRAPHS FOR OMICS DATA INTEGRATION

Omics data comes from a variety of sources and formats, making integration a significant challenge. A knowledge graph can be an effective tool for integrating such diverse data types from multiple sources. A knowledge graph can help to unify different omics data sets by providing a common vocabulary, metadata, and context that facilitates the identification of meaningful relationships between different entities.

DISQOVER provides powerful integration capabilities, allowing users to combine and search for data from multiple sources, including public databases, third-party licensed data, and in-house datasets. The platform supports various data formats so users can access all the relevant data they need in one place via one holistic overview, without having to switch between multiple tools or platforms.

UTILIZING KNOWLEDGE GRAPHS FOR DATA STANDARDIZATION AND DATA HARMONIZATION

Standardization is essential to ensure consistency and reproducibility of results. A knowledge graph can be used to define standard vocabularies and ontologies for omics data, ensuring that data is consistently represented and can be easily understood by different users and tools.

DISQOVER's comprehensive knowledge graph technology provides an ontology-based framework that standardizes and harmonizes data. The knowledge graph provides users with a consistent and unified view of the data, regardless of its original source, and enables them to make connections and identify new relationships between data points. DISQOVER also improves annotations by comparing them to other well-curated annotations from known datasets, in order to eliminate errors made by inexperienced annotators and to ensure that all relevant information is included.

EFFICIENT OMICS DATA EXPLORATION WITH A KNOWLEDGE GRAPH

Life sciences companies need to be able to explore omics data in order to extract value from it, such as through identifying potential drug targets and biomarkers, among other activities. Knowledge graphs provide a powerful interface for data exploration by delivering users with a flexible framework for querying and visualizing omics data so they can gain insights that might have otherwise remained hidden. Researchers and data scientists can easily explore data from multiple angles, navigate complex data relationships, identify novel connections between different data types, enabling them to develop new hypotheses and refine existing ones.

DISQOVER provides a user-friendly interface that enables researchers to easily explore and navigate complex data sets, regardless of their technical expertise. Because the platform integrates data from various sources, researchers can easily perform cross-domain analyses and identify new insights that may not be evident when examining data from a single source.

WHAT ELSE IS POSSIBLE WITH DISQOVER'S KNOWLEDGE GRAPH TECHNOLOGY?

Supporting data-driven decision-making
DISQOVER's knowledge graph can help life sciences companies make data-driven decisions by providing a framework for integrating and analysing data from different sources across an organization that would otherwise remain siloed. By combining omics data with other sources of organizational information, such as clinical trial data, electronic health records, or real world evidence, users can develop a more complete understanding of the data available to drive more informed decision making.

Enabling collaboration and knowledge sharing
DISQOVER allows users to share searches, templates, and dashboards with collaborators, making it easier to work together on complex omics projects and share knowledge. By facilitating collaboration and knowledge sharing, DISQOVER helps life sciences companies accelerate research and development processes, reduce duplication of effort, and improve the overall quality of research.

HOW DO KNOWLEDGE GRAPHS SUPPORT ALONG THE OMICS DATA LIFE CYCLE?

We discussed solutions for managing omics data, along with other topics related to multi-omics research approaches during our recent webinar with ZS Associates, a management consulting and technology firm focused on transforming global healthcare.

We detailed how you can optimize your multi-omics research and data management through the use of a knowledge graph and how a knowledge graph can support across the data life cycle via searching for data, selecting data of interest, and building insights.

Looking to get more value of your omics data while simultaneously simplifying your data management process? Check out the webinar recording and learn more about the power of knowledge graphs when working with omics data: Watch the recording >>>