

# Collaborative and Interactive Data Analytics in Biotechnology

CENTOGENE connects patients around the world to transform the science of clinical and genetic data to diagnose, understand, and treat rare diseases. Leveraging unparalleled levels of unique data from their Bio/Databank, CENTOGENE is a leading generator of rare disease insights, who harnesses the power of multiomics and advanced technology to resolve the mysteries of rare genetic diseases. By working together with patients, physicians, and partners to research, discover, and develop medical solutions, CENTOGENE can provide answers today for a patient's better tomorrow. For this project, the CENTOGENE team (Ionut Onila, Zhu Guanchen, Anne Schwerk) was assisted by KNIME Partner Discngine (Riccardo Martini).

## Access to Data and Algorithms for Non-AI Experts

The main business problem was to push forward the frontier of metabolomics: The biomarker department was enabled to use novel and highly advanced methods for metabolomic analysis and biomarker discovery, i.e., AI-based algorithms that allow discovery of new insights on big data sets. A software environment was required in which the biomarker experts had access to the data and algorithms without having to be IT or AI experts. The environment also had to be flexible enough to enable innovation by seamlessly integrating new or improved algorithms.

A collaborative and interactive infrastructure was crucial to enabling everyone to work seamlessly together. Given the nature of the work, and the speed at which CENTOGENE operates, the platform needed to be reliable, offer reproducible workflows, provide a way to standardize operations (including tracking and auditing), as well as automate certain steps.

## Strong Interdepartmental Collaboration with Guided Analytics

KNIME Analytics Platform provides the ideal environment for a team of data scientists to build several large, automated workflows and visualizations. These workflows, deployed to the KNIME WebPortal via KNIME Server, create web-based applications that users can access and interact with. The intuitive and easy-to-use nature of the solution enables strong interdepartmental collaboration. With Guided Analytics, users can interact with and explore the data through interactive web pages where a KNIME workflow is running under the hood. This provides the team with access to the data, which is pre-processed and offered in a more usable form. It also provides the ability to interact with it at pre-determined points and dive deeper when needed. The seamless integration with databases, processes, and other software means KNIME is used as a hub for the optimization and linkage of a bigger software architecture. This is made possible by features such as the native Python integration and integration with Jupyter notebooks.

Due to this central, pivotal role, KNIME is also used as an early detection point for variations in the infrastructure to which it is connected to. This was made possible as a side benefit of a feature that was implemented primarily to increase reliability during workflow development: automated workflow testing. Both the company and project goals intersect at the level of automation, reproducibility, and security related aspects.

### Company

CENTOGENE engages in diagnosis and research around rare diseases – transforming real-world clinical, genetic, and multi-omics data to diagnose, understand, and treat rare diseases.

Bioinformatics

Biomarker Development & Diagnostics

Guided Analytics

*"KNIME has enabled the automation of workflows, which has reduced the amount of manual work required by the data scientists, as well as enabled users to interact with intuitive visualizations to get even more out of the data than was previously possible"*

Anne Schwerk, IT Project Manager Artificial Intelligence, CENTOGENE

Security and control is addressed by native KNIME functions, such as versioning and extensive logging for auditing. However, what allows the services and implementations to exceed expectations in this case, is the flexibility with which newly developed components enable the sharing and reusing of KNIME workflow snippets – either in other workflows or by other teams – which ensures reproducibility and standardization of different processes.

### Results

KNIME has enabled CENTOGENE to provide their rare disease patients with better medical solutions more quickly and cost effectively. By optimizing automation, reproducibility, and security, KNIME has provided the CENTOGENE team with the tools needed to minimize time spent on non-essential training of the technology, and fewer processes with regards to how the domain expert interacts with the Bio/Databank.

Using KNIME has resulted in enhanced collaboration and interactive infrastructures, which are now the core elements in creating machine learning models for biomarkers. KNIME has also enabled the automation of workflows, which has reduced the amount of manual work required by data scientists, as well as enabled users to interact with intuitive visualizations to get even more out of the data than was previously possible. Being able to meaningfully integrate data coming from different resources and experiments is what gives users the edge to address complex scientific scenarios. The overall result: the ability to identify biomarkers and improve workflows for screening and diagnosis for patients more quickly.

### Why KNIME?

KNIME Analytics Platform is intuitive and makes creating data science workflows easy due to visual programming, the drag and drop workflow building method, and shallow learning curve. However, the power of KNIME lies in the self-documenting nature and reproducibility of these workflows. This ensures that knowledge and expertise is captured and saved automatically and enables others to understand what is going on. Parts of the workflow can be packaged up into components and shared among colleagues or added to other workflows. This guarantees standardization as well as compliance of certain steps or processes, such as data processing rules. Tracking and auditing of KNIME workflows is automatically captured via the workflow metadata.

KNIME Server plays a pivotal role in this solution because it offers workflow automation, enables collaboration among team members in remote locations, and provides users with access to the KNIME WebPortal. It also functions as hub for the bigger infrastructure in which it is embedded. Moreover, KNIME Server on AWS enables users to adapt resources once the data process becomes intense, without compromising existing infrastructure and without burdening other connected resources. From a business perspective, KNIME addresses all concerns around risk, security, and auditing.

Expertise from KNIME Partner Discngine was a strong contributing factor to the success of this project due to both their sound technical knowledge and understanding of KNIME, as well as their life science background.



Discngine, a KNIME Partner, is a Paris-based software company specializing in applications for the research departments of life sciences companies; specifically for work related to active molecule discovery. It was established in 2004 by two of Pfizer's research alumni in order to develop software that increases the effectiveness of life science research. It currently has 55 employees. Clients include world leaders in the pharmaceutical, cosmetics and agrochemical industries.