

Unlocking CMC process data for predictive analytics

CUSTOMER STORY

At a global biopharmaceutical company, data scientists analyze process data from chemistry, manufacturing, and controls (CMC) to improve operational efficiency and product quality. Their data architecture, however, cannot generate the datasets needed for advanced analytics and AI.

The company uses PharmaMV to monitor and control its manufacturing operations in real time. Measurements from in-line sensors and at-line analytical instruments, such as those from bioreactors and continuous chromatography systems, are collected and stored in a SQL database. With this data, PharmaMV provides real-time insights into the manufacturing process. But the software isn't designed to enable predictive analytics.

To build predictive models, data scientists must aggregate extensive historical datasets, a task that is prohibitively difficult in PharmaMV. The data also requires additional scientific context or feature engineering to train models effectively. As a result, data scientists are unable to develop predictive models that could dramatically improve drug production.

Replatform and engineer data at scale

The biopharma partnered with TetraScience to make its process data accessible to data scientists and ready for modeling. The new data workflow centers around the Tetra Scientific Data and AI Cloud™. Data from PharmaMV is automatically replatformed and engineered into Tetra Data, which is optimized for advanced analytics and AI.

To accomplish this, a SQL connector regularly pulls data from PharmaMV's database into the cloud. Automated pipelines then contextualize the data with relevant metadata and transform it into an open, vendor-agnostic (JSON) format with scientific taxonomies and ontologies. The resulting Tetra Data is compliant, liquid, and purpose engineered for Scientific AI. Data scientists can easily assemble large-scale datasets using a programmatic REST API or a SQL query interface.

Power predictive models with AI-native data

The Tetra Scientific Data and AI Cloud provides data scientists access to a trove of valuable process data—previously unusable for analytics and AI. It enables them to efficiently compile the datasets required for accurate modeling of manufacturing processes. Search is fast and intuitive thanks to centralized and richly contextualized data. With flexible downstream integration, data scientists can analyze the datasets in their preferred application or platform.

AI-native Tetra Data frees data scientists from the burden of extensive manual data processing. They can focus their time on developing dashboards and predictive models rather than preparing data. Plus, Tetra Data is future-proof, ensuring historical and current datasets can be fully leveraged for years to come.

Challenge:

Data scientists at a top 25 biopharma cannot readily use process data from CMC operations to build predictive models.

Solution:

The Tetra Scientific Data and AI Cloud automatically retrieves data from PharmaMV and engineers it into analytics- and AI-ready data.

Result:

- Search and assemble contextualized data rapidly
- Access AI-native data for advanced analytics and modeling of drug production

Learn more

Ready to see what TetraScience can do for your CMC workflow?

Contact us today