tetrascience

Case Story

Fueling drug discovery with AI-native data

A leading biotechnology company aims to make drug discovery and development more efficient and successful through artificial intelligence (AI). By analyzing high-throughput screening (HTS) data, its advanced AI models can predict drug properties, thereby accelerating the development of effective therapies.

The Challenge

The company relied on an unscalable, labor-intensive workflow to feed data into its AI models (Figure 1). Scientists had to move data manually between instruments and applications, while keeping track of critical information like plate numbers and the names and locations of files. This process was slow and mentally taxing, with frequent delays due to unexpected downtime. Data eventually made its way to IDBS ActivityBase and then was pushed once a day to Snowflake for AI applications.

There was no single source of truth for scientific data or a standard naming convention for files. When files were manually exported from the plate readers, most of the metadata was lost. This approach was inefficient, impaired search and collaboration, and compromised data integrity.

Challenge:

Scientists at a top biotech company relied on an unscalable workflow to prepare HTS data for AI-powered drug discovery.

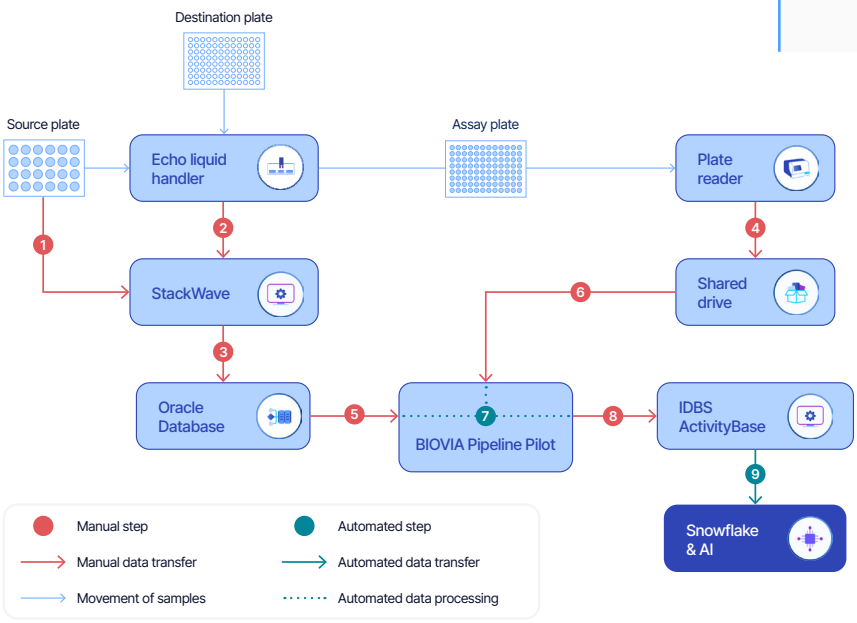
Solution:

The Tetra Scientific Data and AI Cloud™ optimizes HTS by automatically assembling raw scientific data and engineering it into AI-ready datasets.

Outcomes:

- Streamlined operations, freeing up 240 hours per year for scientists
- Improved data accessibility with centralized, contextualized data in the cloud
- Enhanced data integrity by minimizing manual, error-prone tasks
- Future-proofed integrations and pipelines
- Automatically prepared AI-native data

Figure 1. Initial workflow



1. The scientist registers the source plate in StackWave software.
2. After creating assay plates on the Beckman Coulter Echo liquid handler, the scientist transfers the Echo log files to StackWave.
3. StackWave extracts the volumes dispensed from source to destination plates. This data is manually transferred to the database.
4. Following measurements on the plate reader, the raw data is manually exported to a shared drive.
5. The scientist transfers sample data from the database to Pipeline Pilot.
6. They also import assay results from the shared drive.
7. Automated pipelines reformat the sample data and assay results.
8. The scientist uploads the reformatted data into ActivityBase for analysis and review.
9. Data from ActivityBase is transferred once a day into Snowflake, where it is subsequently used for AI applications.

The convoluted, manual-intensive workflow made errors difficult to detect and diagnose, with troubleshooting often taking 30 minutes to 2 hours. Scientists and data engineers relied on complicated diagrams of custom scripts in BIOVIA Pipeline Pilot to pinpoint issues. While suitable for prototyping, Pipeline Pilot demands significant maintenance over the long term. Continuing on this path would have required substantial internal resources for ongoing support, which was not sustainable.

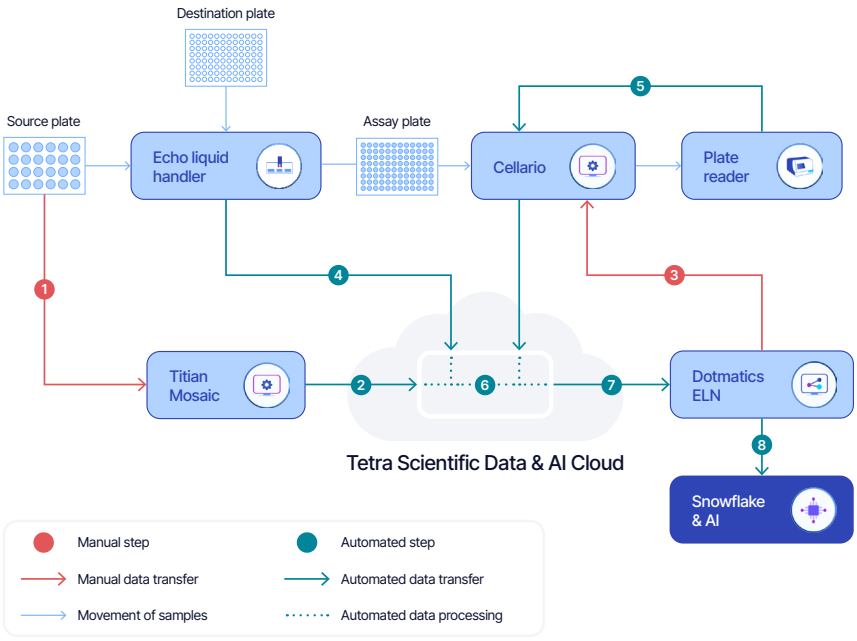
The Solution

The biotech company partnered with TetraScience to overhaul its HTS data workflow. At the heart of the solution is the Tetra Scientific Data and AI Cloud, which integrates seamlessly with instruments and newly introduced software—including Titian Mosaic, HighRes Biosolutions Cellario, and Dotmatics ELN—to streamline the flow of data (Figure 2).

The Scientific Data and AI Cloud not only enables the automatic collection of data from various sources but also engineers it into Tetra Data. This Tetra Data is contextualized with scientifically relevant metadata and stored in an open, vendor-agnostic format, optimizing it for analytics and AI.

The new workflow automatically maps liquid handling and compound information to plate results and sends the harmonized data to the scientists' electronic lab notebook, Dotmatics ELN. As a result, the data is now available in real time for downstream applications rather than waiting for daily pushes.

Figure 2. Tetra workflow



1. The scientist registers the source plate in Titian Mosaic sample management software.
2. The sample information is automatically ingested into the Tetra Scientific Data and AI Cloud.
3. The scientist creates an experiment in Dotmatics ELN, obtains an experiment ID, inputs this value into Cellario lab automation software, and initiates the run.
4. Following the creation of assay plates, the Echo log files are automatically migrated to the cloud.
5. The assay results from the plate reader are automatically collected in the cloud via Cellario.
6. The Tetra Scientific Data and AI Cloud transforms the raw data from various sources into an open, vendor-agnostic format. It also adds metadata to contextualize the data.
7. The data is then automatically sent to the ELN.
8. Data is automatically transferred to Snowflake, where it is subsequently used for AI applications.

The Outcomes

Powered by the Tetra Scientific Data and AI Cloud, the new workflow is substantially faster, more efficient, and more reliable than the original. It provides immediate and lasting benefits to scientific, IT, and data teams, including:

- **Streamlined operations:** Automation of the data flows saves scientists approximately 240 hours per year. They no longer need to configure or troubleshoot Pipeline Pilot or ActivityBase, and manual data transfer is drastically reduced. Scientists can instead focus on more valuable HTS activities.
- **Improved data accessibility:** Scientific data is automatically centralized in the cloud and contextualized with standardized metadata, making it much easier to search and access for analysis. Downtime is also minimized, as the shared drive, Pipeline Pilot, and ActivityBase would experience outages lasting several hours every few weeks.
- **Enhanced data integrity:** The new workflow provides full data traceability and integrity, significantly lowering the risk of human error by eliminating manual steps.
- **Future-proof integrations and pipelines:** With TetraScience's library of industrialized and validated integrations and pipelines, the biotech company can transition from a patchwork of generic tools, which are resource-intensive to maintain, to a scalable approach.
- **AI-native data:** The Tetra Scientific Data and AI Cloud transforms raw scientific data into liquid, large-scale, purpose-engineered, compliant datasets that are optimized for AI, supporting the biotech's approach to drug discovery and development.

Conclusion

In partnership with TetraScience, the biotech company has radically revamped its HTS data workflow to increase the speed and scalability of its research processes. The AI-ready data generated will fuel predictive modeling to accelerate downstream drug development.