

Unlocking the value of big data in a global biopharma for biologics discovery and development

Results: Empowering a global biopharma company with the Tetra Scientific Data and AI Cloud

- **17 million files from 107 pipelines across 13 departments** standardized and onboarded to the Tetra Scientific Data and AI Cloud after 9 months
- **Near real-time data** ingestion, harmonization, and migration across across multiple laboratories
- **~2 million samples** from contract research organizations (CROs) automatically uploaded, standardized, and migrated into the LIMS systems
- **76 hours per week** shifted away from manually preparing data for CROs
- Reduced time for data prep and analysis from **10 days to 10 minutes**
- **90% faster** pipetting error detection in HTS workflows

Challenge: Harnessing an enterprise-level data boom

This global biopharma company was determined to accelerate its discovery and development, improve time-to-insight, and bring therapies to market faster. As part of this goal, they leveraged four contract research organizations (CROs) and scaled operations in several data-focused specialties, including genetic research and development.

Background

A large, global biotech was searching for a data solution that would aid in their goals of:

- Accelerating biologics discovery and development
- Improving time to insight
- Bringing therapies to patients faster

Customer Profile

A leading biopharmaceutical company with a focus in genetic research and development.

- Best-in-class biologics research with numerous products in market
- Works with four CROs to power their research programs
- Employs over 10,000 people
- Earns an annual revenue of over \$16 billion

KPI/Results

- Time-to-insight
- Speed of discovery
- Capital and resource expenditure

However, their standard data protocols were insufficient and could not scale with the sharp increase in data production.

The team recognized that they needed a scalable computer infrastructure that could provide on-demand access to data being produced by CROs and numerous laboratories across their global enterprise. One possible solution was utilizing cloud-based technology.

However, integrating existing systems and applications with the cloud would require in-house IT resources. Additionally, the massive amount of incoming data necessitated a completely new data infrastructure. Building and maintaining a new data infrastructure would incur significant costs and require a team of dedicated engineers.

Furthermore, data produced by various instruments arrived in heterogeneous formats. This meant it had to be standardized

and inspected before it could be used in downstream analysis (data cleansing). This process was being performed manually through customized Microsoft Excel macros. Data analysts and QC then had to manually verify no errors were made during data processing. These manual data curation processes were time consuming and led to errors and poor data integrity.

Scientists also had no way to automatically attach relevant context or metadata, leaving much of the curated data unsearchable and difficult for advanced analytic programs, like AI/ML, to process.

Data cleansing is time-consuming and expensive. For this company, some scientists spent a majority of their time manually curating and transforming data.

In short:

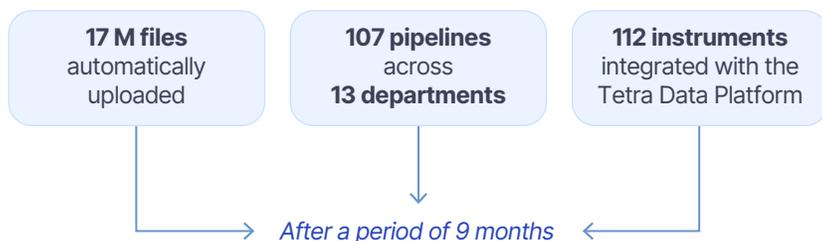
- Accelerating discovery requires leveraging data at scale
- In-house migration to the cloud posed significant resource and labor challenges
- “On-demand” access to data from LIMS, ELNs, and assay repositories was crucial
- Manually cleansing and transferring large volumes of heterogeneous data was time consuming, posed significant risks to data integrity, and was not scalable
- Without proper metadata and scientific context, data becomes unsearchable and unusable for analysis

Solution: The Tetra Scientific Data and AI Cloud

Partnering with TetraScience helped this global biotech develop a scalable and cost-effective cloud infrastructure to manage their ongoing data boom.

Access data immediately and automatically

The Tetra Scientific Data and AI Cloud enables the company to automatically gather, standardize, and migrate data from instruments (e.g., protein purification data from AKTA) and applications to LIMS, ELNs, and assay repositories. This provides near real-time, remote access to data from multiple departments via dozens of pipelines.



Scale with a flexible infrastructure

Automated, company-wide data ingestion provides a scalable data solution with no additional hardware (CapEx). Additionally, having a centralized, regularly updated data infrastructure will allow the company to forgo Isilon center data storage while eliminating the efforts and costs associated with hardware end-of-life migrations and physical asset management (e.g., GPUs).

Harmonize and enrich data

Heterogeneous data, both legacy and novel, across the company's instruments and applications is automatically converted into the standardized Tetra Data. Common data schemas enable data migration (e.g. to their Benchling ELN) with no manual data processing. This means data is automatically and easily accessible for scientists and data scientists within their chosen programs. Additionally, the data is automatically enriched with metadata within the pipelines. Combined, these processes create searchable data that is operable within advanced analytics and visualization tools as well as AI/ML programs such as Dataiku.

Accelerate discovery through data-powered insight

The Tetra Scientific Data and AI Cloud helps scientists reduce time spent on manual data processing and migration related to their automated liquid handlers within their high-throughput screening (HTS) workflow. This accelerates time-to-failure detection, increases pipeline throughput, and frees scientists to focus on higher value work. Meanwhile, data scientists have easily accessible and harmonized data to build their advanced analytical models. The result is an acceleration of the discovery process.

Use cases for TetraScience automation

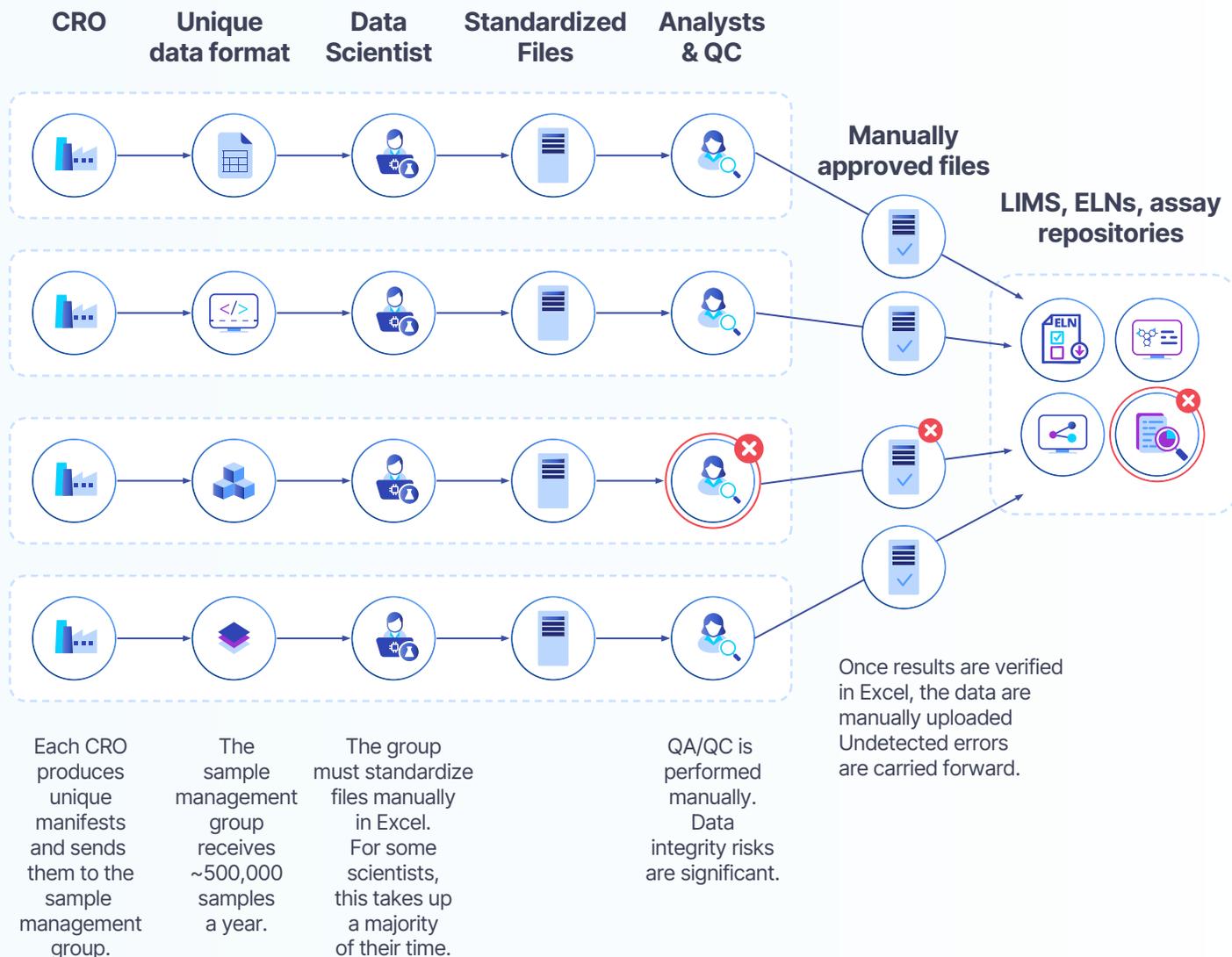
Use case 1: Streamlining and standardizing data packages for CROs

Before: The company works with four CROs. Data produced by these CROs was manually uploaded into LIMS. However, the data is received in multiple formats. This required time-consuming manual transformation and inspection before it could be uploaded to ELNs, LIMS, and assay repositories. To solve this

issue, the company built customized Excel macros to process the files. Then, data analysts and QC both had to check that no errors occurred while data was manually edited within Excel.

The company realized that they could not ensure data integrity with their Excel-based processes. There was no way to track version history, and any human error was carried forward, leading to downstream waste and inefficiency.

Before TetraScience: Manual data transfer and handling

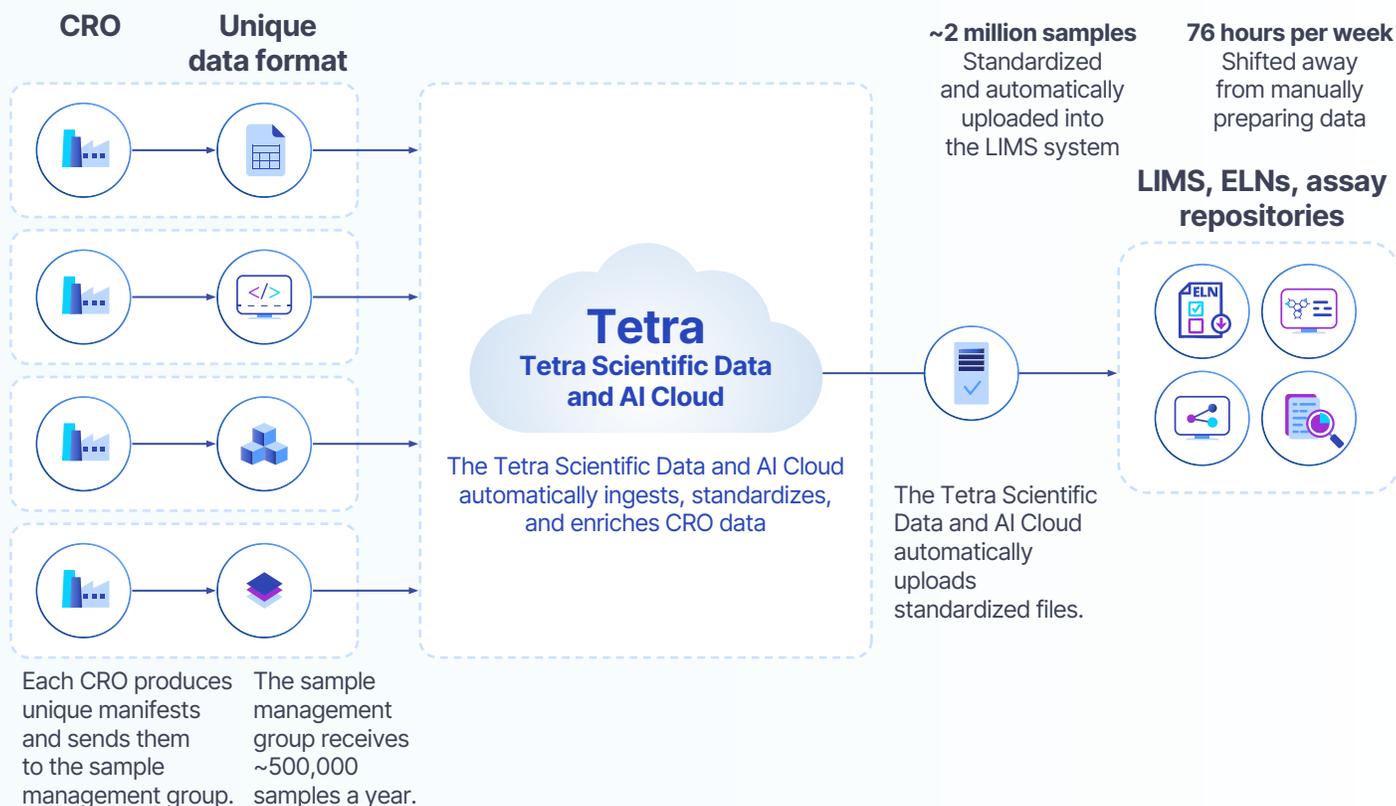


Use case 1: Streamlining and standardizing data packages for CROs

After: Tetra Scientific Data and AI Cloud automation eliminates the need for manual processing. It converts the heterogeneous data it receives into standardized Tetra Data, which is then immediately

available throughout the platform including LIMS, ELNs, and assay repositories.

After TetraScience: Automated digital transformation



Use case 2: High volume Zeiss .CZI image processing

Before: High volumes of .CZI imaging files were uploaded to AWS, which triggered Spark jobs to parse the image data into tiles for upload into Dataiku. This process was managed manually and executed upon request. It was exceedingly time consuming (each image produces 100,000 to 250,000 tiles), nonstandard, and expensive.

After: With the Tetra Scientific Data and AI Cloud, .CZI data is available without manual uploads. Simplifying .CZI data processing results in faster time to FAIR data. Image data is also given context and metadata, making it easier for the image modeling team to query and access data within Dataiku.

3-4 manual steps eliminated
Automation reduces time-consuming and costly manual data handling

Immediate data availability
Data are available in the platform at the point and time of generation

Discovery and tracking with Dataiku
Automatically generated metadata empowers querying and access

Use case 3: Supercharging error detection for high-throughput screening

Before: The company built a genetics research wing that is growing rapidly and is expected to double in size in the near future. With that expectation, scaling their high-throughput screening capabilities became a priority. However, manually checking for equipment output errors and other critical status changes were frequent bottlenecks hindering their progress.

After: The Tetra Scientific Data and AI Cloud has successfully integrated with the Hamilton Microlab STAR liquid handler and the parser, allowing pipetting errors to be detected much more quickly.

<2 minutes The time it takes research teams to search and identify hundreds of pipetting errors

Significance: Empowering global scientific discovery with the Tetra Scientific Data and AI Cloud

Outcome	Significance
<p>Eliminated manual data preparation, cleansing, and migration</p>	<ul style="list-style-type: none"> • The Tetra Scientific Data and AI Cloud eliminates the need for timeconsuming and error-prone manual interventions. No more manual data handling, Excel workbooks, or human quality control • 76 hours per week gained from eliminating manual preparation of data from CROs • Data preparation and analysis that once lasted 10 days with in-house solutions now takes 10 minutes
<p>Scalable data management</p>	<ul style="list-style-type: none"> • 17 million files from 107 pipelines across 13 departments standardized and uploaded to the Tetra Scientific Data and AI Cloud after 9 months
<p>Immediate access to CRO data</p>	<ul style="list-style-type: none"> • Data produced by CRO partners is automatically and immediately available in LIMS, ELN, and assay repositories.
<p>Improved workflow performance</p>	<ul style="list-style-type: none"> • Pipetting error detection is 90% faster. Pipetting errors in the company’s gene therapy research and development HTS workflow can be detected in under 2 minutes using the Tetra Scientific Data and AI Cloud near real-time data access
<p>Standardized and enriched data</p>	<ul style="list-style-type: none"> • The Tetra Scientific Data and AI Cloud standardizes data across multiple departments, pipelines, and CROs. These functions greatly increase usability for preclinical investigations, process improvements, and AI/ML capabilities

Summary: Empowering biopharma with the Tetra Scientific Data and AI Cloud

Partnering with TetraScience helped this global biopharma develop a scalable and cost-effective cloud data solution that improves workflow efficiency and sets a foundation for advanced analytics and visualization tools. By combining these capabilities with industry-leading data production, this global biopharma can expedite their goals of accelerating biologics discovery and development, improving time-to-insight, and bringing therapies to patients faster.

What's next?

After implementing the Tetra Scientific Data and AI Cloud, this company found huge increases to their workflow efficiency. By leveraging these learnings, they plan to further optimize processes and control their market.

Multiplying cost savings

Leveraging scalable cloud capabilities will enable this global pharmaceutical company to forgo long term storage in Isilon data centers, significantly driving down costs.

Furthermore, TetraScience will help the company eliminate the costs of:

- Physical asset management (e.g., GPUs)
- Manual data processing
- Hardware end-of-life migrations
- Home-grown data solution experts

Keeping (and attracting) top industry talent

Tetra Scientific Data and AI Cloud automation will allow this global biotech company to shift their focus away from time- and resource-intensive manual data transformations toward knowledge generation and life-changing discoveries. This will help them retain their industry-leading talent and attract new generations of scientific leaders.

Bringing therapies to patients faster

By standardizing and capturing near real-time system and application performance data, this global company can increase their research throughput to unparalleled levels, bringing better therapies to market faster.

To learn more about how TetraScience can optimize your scientific data, visit tetrascience.com