tetrascience

databricks

Partnering to unlock scientific data and accelerate the Scientific AI revolution

Databricks | TetraScience

Partnership

Scientific Data and AI

The biopharmaceutical industry is racing to harness AI to improve drug discovery, development, manufacturing, and quality control. The biggest hurdle in this endeavor is the current state of the data. Scientific data, one of the most valuable and fastest-growing datasets, faces a significant **data utility gap** (see figure below). While the industry possesses exabytes of scientific data, much remains inaccessible for developing AI/ML models, applications, and analytics. Closing this data utility gap could usher in an era of accelerated scientific advancement.

The challenge is that AI/ML algorithms require large-scale, liquid, and well-engineered scientific datasets, and producing these datasets in biopharmaceutical organizations is not easy. Why? Four reasons:

1. **Data Complexity and Contextualization:** Scientific data and its related use cases are extraordinarily complex, as are the workflows and methods that create that data. Creating and enriching data with metadata is key for contextualization, requiring deep subject matter expertise.

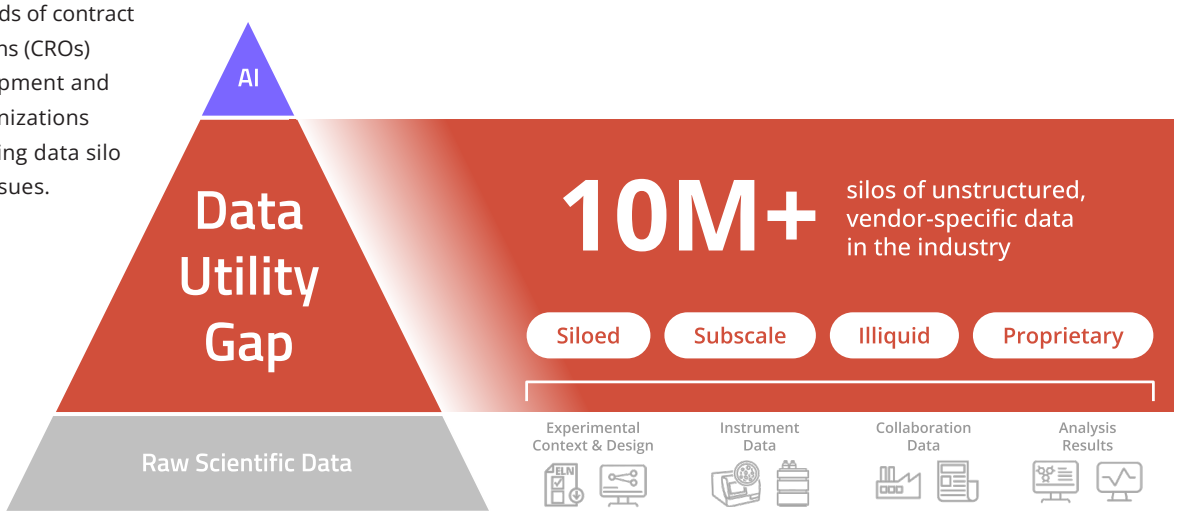
2. **Fragmented Ecosystem:** Most larger biopharma organizations are working with over 500 vendors in the scientific ecosystem, including those providing scientific instruments, lab informatics solutions, scientific software, robotics, and IoT sensors. This diversity leads to complex data ingress and egress challenges, as well as a myriad of incompatible vendor-specific formats.

3. **Globally Distributed Value Chain:** Scientific data is generated across a globally distributed value chain involving sponsors working with hundreds of contract research organizations (CROs) and contract development and manufacturing organizations (CDMOs), exacerbating data silo and data integrity issues.

Key Impacts Realized Across Global Biopharma Companies

- Saved up to **48 hours per screen** by automating data preparation and analysis for flow cytometry, increasing antibody screening throughput by **3x**.
- Reduced hit selection time from **2 weeks to 1–2 days** by powering interactive dashboards, speeding up drug discovery.
- Reduced sampling points by **50%** in preclinical ADME/Tox testing with data science and AI, while improving IC_{50} accuracy.
- Boosted high-throughput media screening by **3x** and cut wet lab experiments by **88%** using AI in cell culture media development.
- Decreased time to insight from **1 week to 1 day** for liquid chromatography workflows, enhancing bioprocessing purification efficiency.

4. **Outdated Approaches:** Often due to internal priorities, many labs still rely on 20th-century DIY IT solutions that address only isolated problems, like point-to-point connectivity, instead of a holistic, data-centric foundational platform.



Raw scientific data is unsuitable for advanced analytics and AI, creating a data utility gap.

Corporate Headquarters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115

tetrascience

tetrascience.com

TetraScience

TetraScience has developed a unique, science-led approach to transform raw scientific data into **large-scale, liquid, well-engineered, and AI-native scientific datasets.** We replatform and engineer data with sophisticated taxonomies and ontologies that are actionable by AI for real-world scientific use cases. The Tetra Scientific Data and AI Cloud helps half of the world's top 25 biopharmaceutical companies design and industrialize analytics- and AI-native scientific datasets, bringing them to life through a growing suite of next-generation scientific data management capabilities and scientific use cases, ultimately driving impactful business outcomes. A little more on TetraScience's unique model:

Purpose-Built Scientific Data Stack and Use Case–Based Approach

Design: Each layer of the Tetra Scientific Data and AI Cloud is explicitly designed to handle the complexity of scientific data.

Expertise: Our team combines expertise in scientific data and use cases with intimate knowledge of the modern data stack to develop data schemas, taxonomies, ontologies, etc.

Scientific Business Analysts (SBAs) and Scientific Data Architects (SDAs): Our forward-deployed SBAs and SDAs are embedded with customers, serving as the "connective tissue" between science, IT, and data science teams to achieve their scientific outcomes. We call them "Sciborgs" because these experts have advanced scientific training (e.g., a Ph.D. in a life science discipline with relevant biopharma industry experience) and data modeling skills. They are integral to our customers' success.

Deployment-Ready and Ongoing Support: TetraScience provides a library of deployment-ready scientific use cases, best practices, data models, supporting tools, and access to peer learnings and benchmarking from other customers. This framework accelerates implementation and ensures efficient, right-first-time deployment.

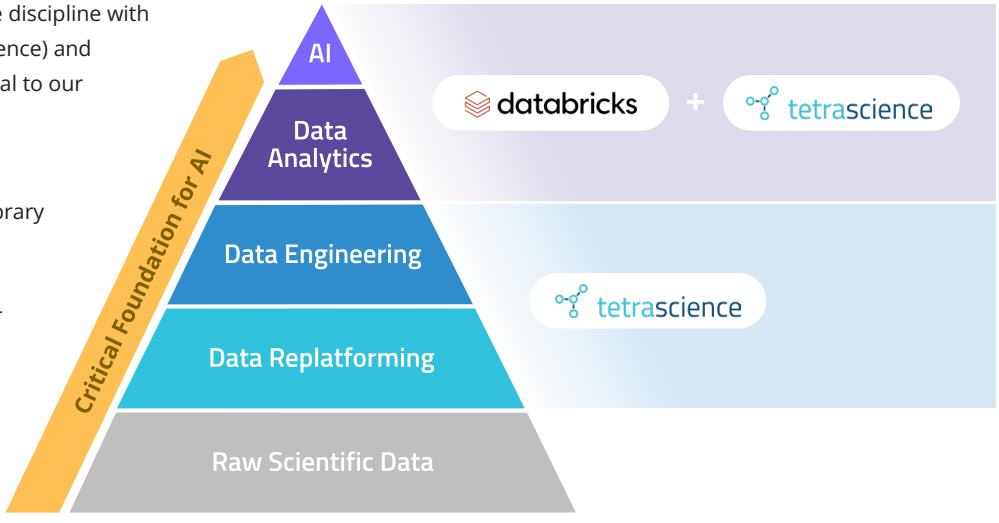
Vendor-Agnostic, Open Business Model

Our model is vendor-agnostic and collaborative. We focus on liberating, harnessing, and future-proofing customers' scientific data. Unlike scientific endpoint vendors who lock customers into proprietary walled gardens or have access to only a small subset of the data, we help customers build the kinds of supersets of scientific data required for analytics and Scientific AI.

This unprecedented scale of data volume and liquidity leads to richer taxonomies, more robust ontologies, and higher-fidelity AI-native data.

Databricks and TetraScience Partnership

The partnership between TetraScience and Databricks combines TetraScience's expertise in scientific data, AI, and scientific domain knowledge with Databricks' industry-leading Data Intelligence Platform. This collaboration aims to drive the Scientific AI revolution and tackle some of the biopharmaceutical industry's most complex challenges. By integrating advanced analytics and AI/ML capabilities into discovery, development, manufacturing, and quality control, the partnership enhances process efficiency, reduces risk and costs, and ultimately accelerates innovation.



Scientific data must undergo an immutable series of operations for AI, uniquely enabled by TetraScience and Databricks.

Corporate Headquarters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115

tetrascience

tetrascience.com

As previously described, TetraScience replatforms and engineers scientific data to create open, large-scale, liquid, AI-native datasets, known as Tetra Data. These datasets can be seamlessly exposed as Delta tables for analytics and AI/ML applications within the Tetra Scientific Data and AI Cloud. Additionally, Tetra Data is registered in the Unity Catalog and made available to customers' Databricks platform via Delta Sharing using zero-copy integration, enabling its use in an expanding suite of Scientific AI applications across the pharmaceutical value chain. This enables customers to conduct advanced analytics, build custom ML models, leverage generative AI, support data sharing, and establish unified governance for data and AI/ML models.

Delivering Tangible Scientific Outcomes

Through the combined strengths of TetraScience and Databricks, customers can achieve accelerated and enhanced scientific outcomes, demonstrated by the following representative examples:

Accelerate Time to Milestone: Utilize predictive modeling to significantly reduce the number of experiments required in clone selection and media formulation, cutting upstream bioprocess development by more than 80%.

Increase Efficiency: In bioprocess development, automate data workflows end to end, increase productivity by 50% without expanding the lab footprint, and free scientists to focus on more valuable work.

Reduce Risk: Triple QC efficiency and predict deviations early with performance monitoring dashboards to reduce out-of-specification events by up to 75% and ease compliance burden. Additionally, automate error-prone manual transcription across R&D to improve data integrity.

Use Cases Powering Scientific Outcomes

This partnership brings AI-native data to life through a rapidly expanding suite of scientific use cases across the pharmaceutical value chain. These use cases are designed to apply advanced analytics and AI models to specific scientific challenges, enabling organizations to extract data-driven insights, accelerate time to milestone, improve efficiency, lower costs, and drive innovation.

Some representative use cases include:

Research and Discovery

- High-throughput screening
- Cell profiling/sorting (flow cytometry)
- Asset utilization

Early and Late Development

- Bioprocess development (including lead clone selection, cell culture media formulation, purification, and biologics characterization)
- Analytical method development
- Preformulation/formulation screening
- Asset utilization

QC and Manufacturing

- Quality testing for release and stability (small molecules and biologics)
- Asset utilization

Accelerate Your Scientific AI Journey

Contact us today to discover how we can support your scientific and business goals.

Corporate Headquarters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115

tetrascience

tetrascience.com