

# The Tetra Scientific Data and AI Cloud

## SOLUTION BRIEF

With an average of over ten years and \$2 billion to develop a new therapeutic, life sciences organizations are seeking a new paradigm to achieve sustainable success. Given the increase in computational power, democratization of cloud computing, and emergence of more powerful neural networks, the industry expects artificial intelligence (AI) and machine learning (ML) to radically accelerate and improve scientific outcomes, reducing time and costs for therapeutics development.

But AI/ML algorithms require large-scale, liquid, and well-engineered datasets, and producing these datasets is difficult given the complexity of scientific data and the diversity of its sources. Across approximately ten thousand companies in life sciences, there are more than one million data silos. Typically, the datasets in these silos are static and locked in proprietary, vendor-specific formats. The data is often unstructured and lacks the scale required for AI applications. This data is unfit for Scientific AI.

## Scientific Data

Successfully preparing data for AI applications requires a thorough understanding of what scientific data is and what attributes must be preserved throughout the scientific data journey.

**Complexity:** Scientific data is complex—as are the workflows that create that data. Methods for characterization and testing, and for conducting scientific experiments, could not be more diverse. With a highly fragmented scientific ecosystem of more than 500 vendors across instruments, electronic lab notebooks, informatics apps, robotics, and IoT sensors, the processes and workflows that generate data include a variety of ingress and egress points with distinct format and content requirements. Any work to prepare scientific data for AI must retain information about instruments and processes.

**Contextualization:** Metadata is critical to providing the context for scientific data. Creating metadata requires deep subject-matter expertise in all dimensions. It requires an understanding of not only scientific data but also the life sciences technology vendor ecosystem of instruments and applications, as well as scientific workflows and use cases.

**Compliance:** Scientific data must be compliant. The life sciences industry is heavily regulated by authorities to ensure patient safety. Hence scientific data must be compliant with regulations and guidelines (e.g., US FDA 21 CFR Part 11), maintain high quality, and provide full traceability and integrity.

**Control:** All data generated by users belong to the users. Independently of the scientific application, instrument, or vendor—and independently of the format, status, and location of the data—data must always be available to its originator to be reused and leveraged, unconditionally.

**Collaboration:** Scientific data is generated across a globally distributed value chain comprising tens of thousands of scientific organizations and thousands of their contract research organizations (CROs) and contract development and manufacturing organizations (CDMOs). Collaboration on this data, ensuring its integrity and traceability, and extracting value from it are equally complex and critical.

## The Scientific AI Journey

Raw data as generated by scientists and lab instruments cannot be used for Scientific AI—its value is very limited. This data must be transformed to drive AI outcomes and generate insights for truly transformational scientific breakthroughs. There is an immutable order of operations for the journey of scientific data from raw, primary data to AI-enabled data.

Here is the typical scientific data journey:

**Raw data** typically resides in vendor-specific silos and exists in proprietary, vendor-specific, and often unstructured formats. Datasets are too small and static to be exploited by AI and ML. This data has little to no value for the organization: It is locked in vendor silos, hard to reuse, has little to no utility for predictive analytics and AI, and is costly to maintain.

**Data replatforming** is the first necessary step to improving datasets. Data must become compliant, liquid, and accessible. Collecting, centralizing, and contextualizing all previously siloed data in the cloud allows scientists to leverage a modern cloud infrastructure natively. They gain unprecedented elasticity for storing, processing, easily retrieving, and accessing their data.

Sophisticated **data engineering** requires deep scientific, data, and technology expertise to fully convert scientific data into liquid, large-scale, harmonized, and compliant data with industrialized scientific taxonomies and ontologies. Only these open, vendor-agnostic, AI-enabled datasets can become the atomic building blocks for Scientific AI.

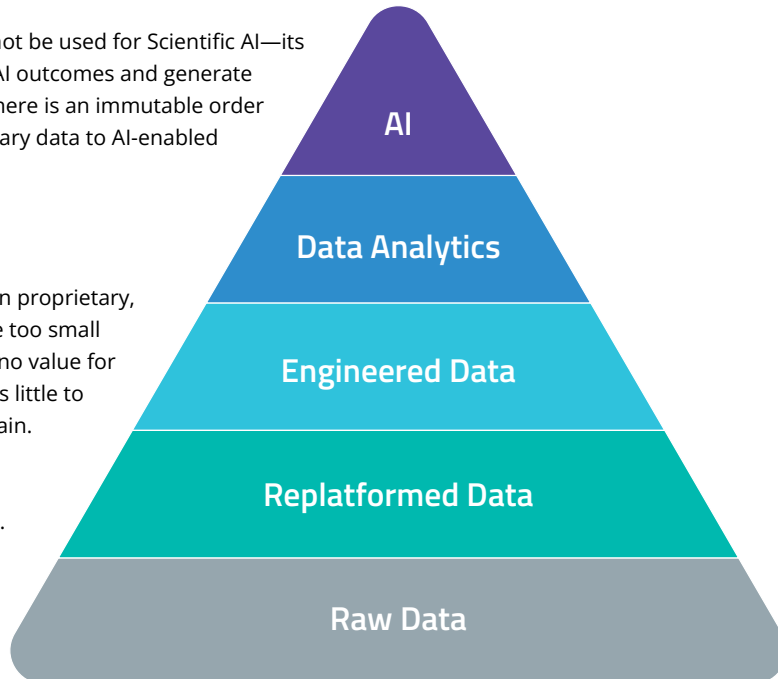
**Data analytics** is one avenue for gaining value from scientific data. While there are numerous vendors and tools that can help organizations leverage their data, fragmented data models will result in massively suboptimal outcomes. Organizations need to engineer their scientific data to optimize it for their favorite dashboards, visualization tools, and analytics applications.

**AI** can help answer the most pressing questions in science. But organizations often lack internal resources or skills to put the right data strategies into place to capitalize on cutting-edge AI applications and collaborative workflows. Only purpose-engineered, liquid data can enable groundbreaking Scientific AI-driven outcomes across the value chain.

The **Tetra Scientific Data and AI Cloud™** is a modern cloud data stack that has been purpose built to elevate the value of data through each layer of the data journey. The Tetra Scientific Data and AI Cloud can help close the gap between scientific raw data and transformational AI outcomes.

## The Tetra Scientific Data and AI Cloud

The Tetra Scientific Data and AI Cloud is the first and only industry data and AI cloud purpose built for science. It is a *sui generis* scientific data stack designed to produce large-scale, liquid, compliant, engineered, Scientific AI-ready data, which can be used by science-based organizations such as biopharma companies.



The Tetra Scientific Data and AI Cloud collects all data produced in each and every step of the **design-make-test-analyze (DMTA) cycle**. Users can link data related to experiment design, measurements, context, methods, compound creation, and analysis—generated internally or externally—to use it for analytics and AI.

## Purpose built for scientific data

Every aspect of the Tetra Scientific Data and AI Cloud is designed explicitly and optimally for delivering scientific AI-native data. It provides integration schemas as well as robust, rich scientific taxonomies and ontologies. It facilitates scientific data workflows and adheres to compliance standards. The Tetra Scientific Data and AI Cloud was designed by data-, technology-, laboratory-, and science-native experts at TetraScience. They have deep industry knowledge and know-how that can help customers design high-level scientific data use cases.

## Open and agnostic

As a data company, TetraScience aims to liberate, future-proof, and add value to our customers' scientific data so they can fully harness its power through advanced analytics and AI. To achieve these goals, the Tetra Scientific Data and AI Cloud automatically transforms data into an open, vendor-agnostic JSON format. Offering different types of integration, a programmatic REST API, and a SQL query interface, the Tetra Scientific Data and AI Cloud enables customers to gain unrestricted access to their engineered data. They can use all of their data with any value-generating best-of-breed application, regardless of the original scientific endpoints vendor's software.

## AI-native

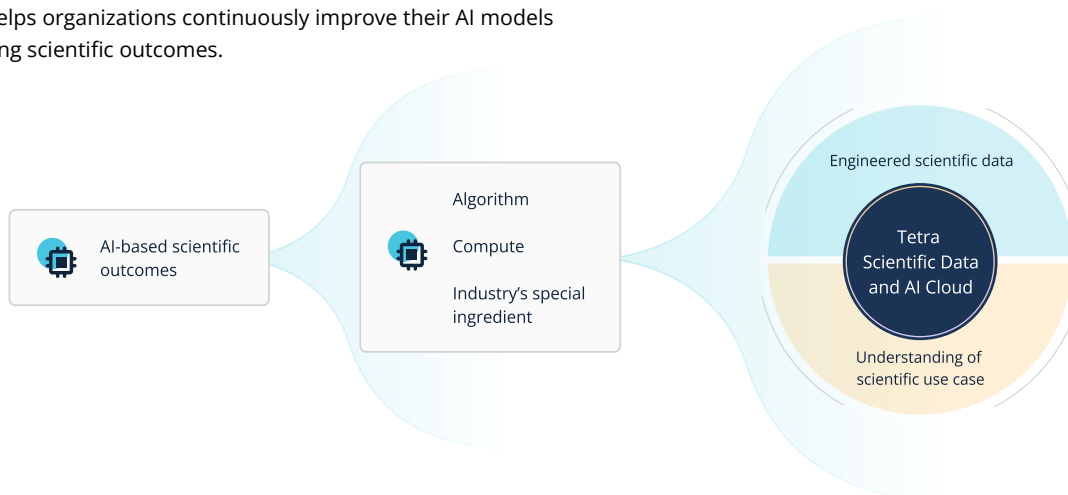
The Tetra Scientific Data and AI Cloud was purpose built to help life sciences companies achieve their AI goals. It produces the world's only AI-native scientific datasets, called Tetra Data, specifically designed and optimized for AI. Tetra Data offers large-scale, liquid, compliant, and purpose-engineered data with scientific use case-based taxonomies and ontologies that build the foundation for transformational Scientific AI outcomes. It also provides transparency for AI results and decisions, showing users layer by layer how a specific conclusion is reached.

## Collaborative

The Tetra Scientific Data and AI Cloud enables the flow of Tetra Data among biopharmas, contract organizations, and other critical partners, fueling unprecedented collaborative innovation through a collaboration substrate. Scientists can securely share and collaborate on complete sets of contextualized, high-value data in real time. This highly efficient and effective exchange of scientific data fuels cross-enterprise federated learning and helps organizations continuously improve their AI models and the resulting scientific outcomes.

The Tetra Scientific Data and AI Cloud is built on a **scientific lakehouse architecture** and natively supports Amazon Athena, Amazon Redshift, and Snowflake. Customers can leverage the Delta Lakehouse architecture, which provides a foundation for Databricks and other data analytics and processing solutions. It allows for multi-modal data consumption ranging from data discovery and analytics to high-performance computing and AI-model training.

The Tetra Scientific Data and AI Cloud powers the **"Scientific AI Factory,"** generating AI-based scientific outcomes with speed and scale. Organizations can rapidly prototype AI models through the collaboration of their internal AI and data science teams with TetraScience experts, who have a deep understanding of scientific use cases. Through an AI-based proof of concept using Streamlit/R Shiny, they can easily prototype, prioritize, and productize models for final Scientific AI-based outcomes.



## Tetra Data

The Tetra Scientific Data and AI Cloud enables the rapid assembly of all scientific data, transforming it into large-scale, compliant, liquid, engineered, and collaborative datasets. It enables the movement of that data to the top of the AI pyramid. At any stage, this data remains owned by the organization that produced it and accessible to its users. The Tetra Scientific Data and AI Cloud liberates the data from vendor-walled gardens, engineers it for the most demanding scientific use cases and AI initiatives, and future-proofs it to unlock maximum value. It also facilitates the sharing of that scientific data with other organizations and partners. Tetra Data is replatformed, engineered, collaborative, and AI ready.

**Replatformed data** is raw (primary) data that has been liberated from siloed data sources and repositories. The Tetra Scientific Data and AI Cloud collects and aggregates scientific data using industrialized, validated integrations. This robust and reliable process ensures data integrity and traceability, as well as compliance with regulations and guidelines like 21 CFR Part 11 and GxP. By contextualizing data for scientific use cases, it becomes searchable and reusable. The centralization of data into a single source of truth within the Tetra Scientific Data and AI Cloud eliminates data silos, enables the large-scale collection of data, and enhances data liquidity. Compliant, liquid data can now flow seamlessly among instruments, applications, departments, and partners.

**Engineered Data** is data transformed for science and Scientific AI. The Tetra Scientific Data and AI Cloud harmonizes raw data from proprietary formats into an open, vendor-agnostic JSON format. It augments data with sophisticated and industrialized taxonomies and ontologies relevant to specific scientific applications and workflows. TetraScience has the deep subject-matter expertise in data, technology, and science—from R&D to manufacturing and QA/QC—required to purpose engineer large scale, AI-native datasets for Scientific AI.

**Tetra X Data** is AI-native, shareable, collaborative, and extensible data. It supports further scale, liquidity, and context. It can drive collaboration across critical partners such as CROs and CDMOs, and scientific endpoint systems, including laboratory information management system (LIMS) and electronic lab notebook (ELN) providers. The unprecedented data liquidity fuels even larger-scale data, leading to richer taxonomies, more robust ontologies, and higher-fidelity AI-native data, which in turn can result in more transformational Scientific AI outcomes.

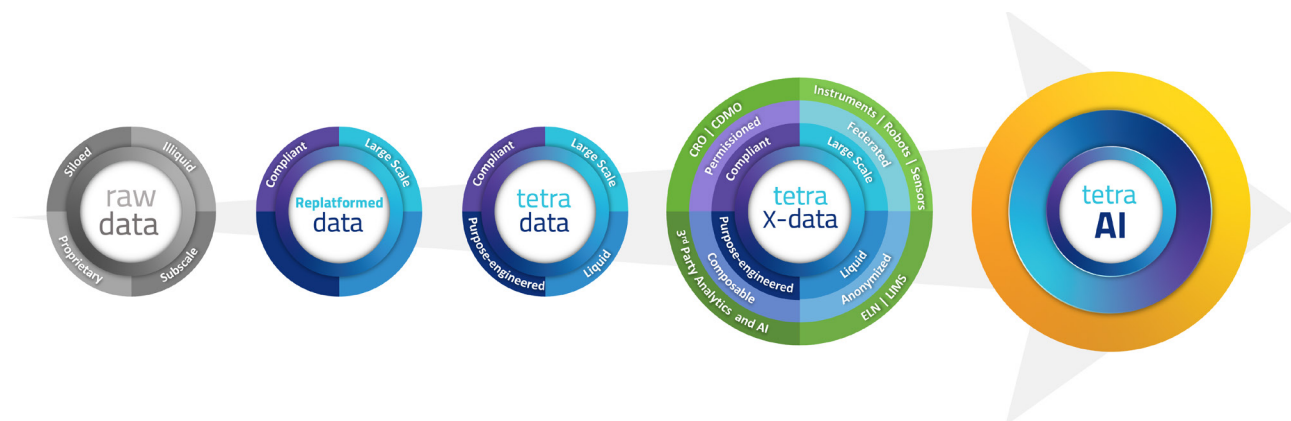
**Tetra AI** is the highest-fidelity AI-native data. It enables organizations to use AI applications to unlock the greatest possible value from their data and deliver tangible scientific outcomes. The world's only open, collaborative AI-native scientific data format, Tetra AI draws from deep, long-term expertise in scientific data, workflows, and use cases, as well as modern data and AI stacks. With Tetra AI, TetraScience enables the continuous engineering of liquid, large-scale, and contextual datasets required for a 21st century Scientific AI Factory.

The Tetra Scientific Data and AI Cloud includes a **Data App Workspace** that enables customers to directly access their Tetra Data through a tool of their choice. These can be existing analytics applications, such as Spotfire, Tableau, PoolParty, Databricks, and more; our Data App Workspace using Skyline, JMP, FlowJo, ACD/Labs, Streamlit, and R Shiny; and the TetraScience chromatography dashboards for stability, instrument and method performance, and lab optimization (SIMPLO). This allows for removing manual data movement, avoiding latency and data duplication of traditional workflows, and reducing IT overhead.

Tetra Data is **FAIR** (findable, accessible, interoperable, reusable) through replatforming and engineering.

**Findable and Accessible:** Tetra Data is contextualized with scientifically relevant metadata, making it easily searchable by humans and machines. Centralized in the cloud, it can be readily accessed by users.

**Interoperable and Reusable:** Tetra Data is harmonized in an open, vendor-agnostic format (i.e., JSON) that can be leveraged by third-party applications, including analytics and AI/ML. The use of industrialized, scientific taxonomies and ontologies ensures the data is fully understandable and reusable in different settings.



## Conclusion

Biopharmaceutical organizations that want to speed therapeutics development and reduce costs through Scientific AI must close the gap between their scientific raw data and value-generating AI applications. The Tetra Scientific Data and AI Cloud offers the only path to closing that gap. It is the factory for producing large-scale, liquid, compliant Tetra Data, which is purpose engineered for science and AI. With the Tetra Scientific Data and AI Cloud, organizations can fully harness their data for Scientific AI, accelerating and improving scientific outcomes.