tetrascience | The Scientific Data Cloud Company

Key takeaways: Scalable data strategies for scientific AI-enabled breakthroughs

FACT SHEET

Artificial intelligence (AI) stands on the brink of revolutionizing the biopharmaceutical industry. It promises to slash time to market, lower costs, mitigate risks, and help create novel therapies. However, the main barrier to realizing these benefits is not the development of AI models but rather the quality and accessibility of the scientific data that feeds them.

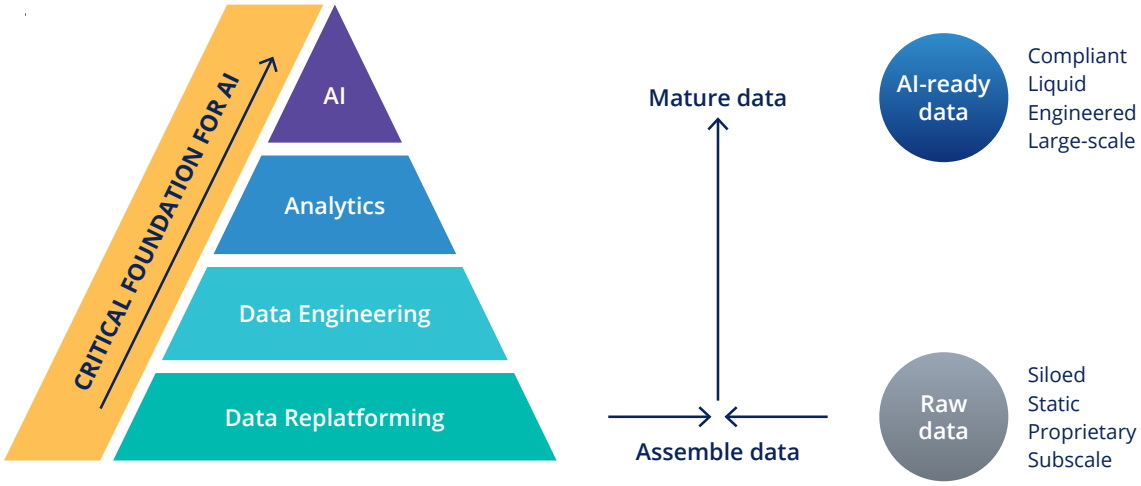
The recent webinar "Scalable data strategies for scientific AI-enabled breakthroughs" featured insights from industry experts Naveen Kondapalli, Senior Vice President of Product and Engineering at TetraScience, and Himanshu Jain, Healthcare and Life Sciences Strategy Leader at AWS. They delved into the data challenges underlying scientific AI and discussed strategies to overcome them. Below, we summarize four key takeaways from the conversation.

1. Lay the data foundation for AI

Many AI initiatives have fallen short of expectations because of data issues. Most scientific data within biopharma organizations is trapped in silos and proprietary data formats. Raw data can't power AI models at scale to drive scientific outcomes. The data must be transformed, or maturated, into a form that AI can readily consume and analyze for deep insights.

This journey from raw data to AI-ready data can be represented as a pyramid with four layers or stages (see below). At the bottom is data replatforming: raw data must first be assembled from scientific instruments and applications and contextualized (e.g., tagged with metadata). Then, it moves upward to the data engineering layer, where it becomes harmonized into a common, open, vendor-agnostic format with consistent schemas and scientific taxonomies and ontologies. At this point, the data is properly structured and enriched for analytics and AI (the two top layers).

Building AI models directly on top of raw data is usually a futile exercise. The process is not scalable. The costs can spiral out of control. And the outcomes, if any, are marginal. The best large language models can't compensate for poorly engineered data.



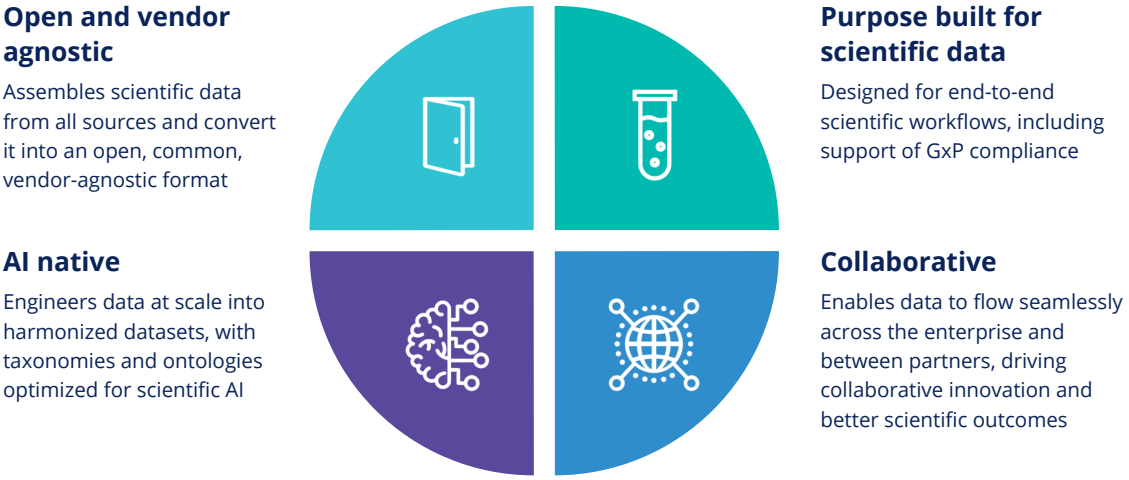
The scientific data journey

To sum up, follow these best practices when preparing scientific data for AI:

- Centralize raw scientific data in the cloud to liberate it from silos.
- Contextualize the data for scientific use cases, including meaningful metadata.
- Engineer raw data into an open, harmonized format (e.g., JSON) with scientifically relevant taxonomies and ontologies.
- Use engineered data sets—not raw data—to train, feed, and improve scientific AI models.

2. Build a data architecture optimized for scientific AI

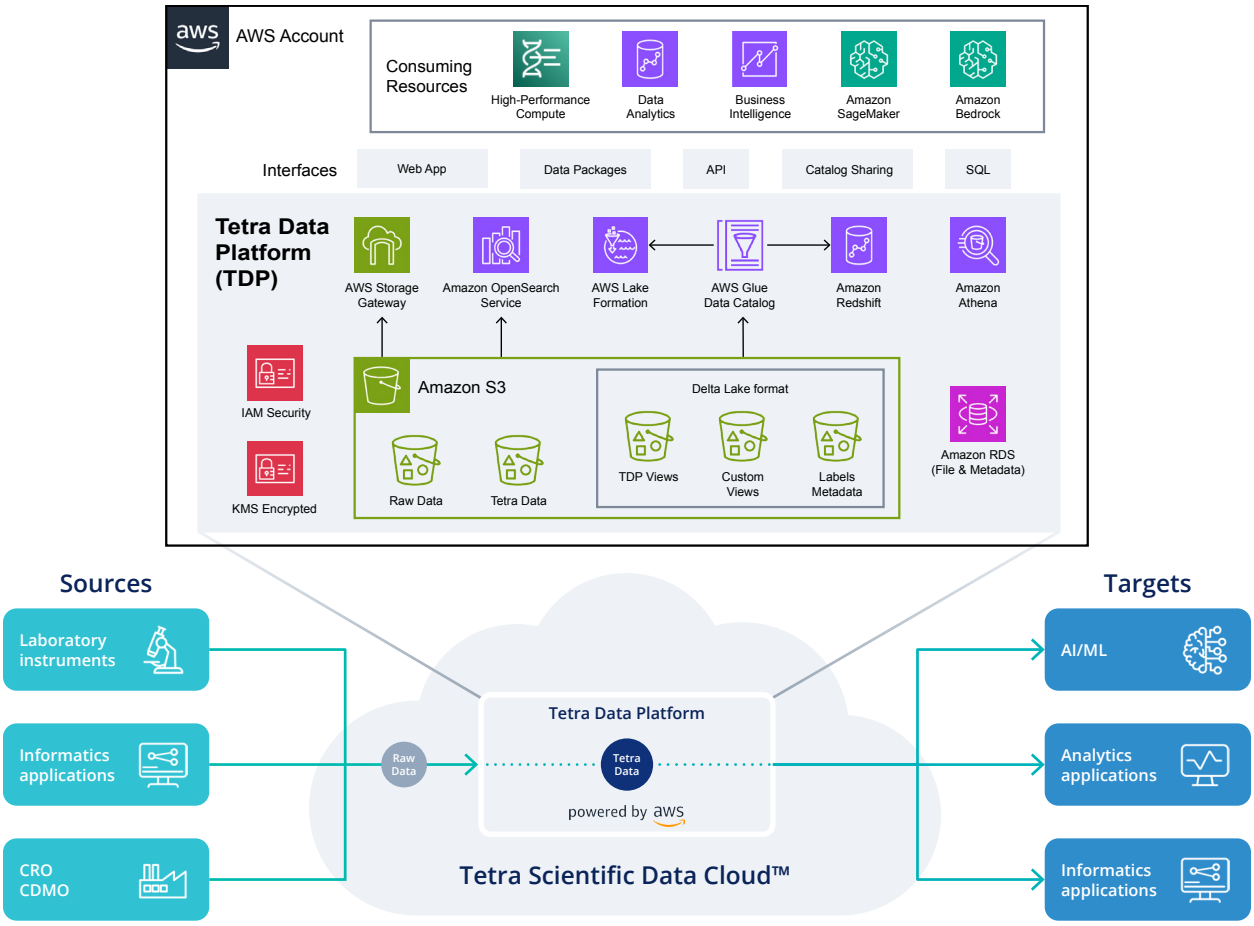
Scientific AI requires liquid, large-scale, purpose-engineered, and compliant data. To readily generate these datasets and maximize their value, you need a data architecture with the following properties.



Key elements of a data architecture for industrializing AI-ready data

The combined AWS-TetraScience reference architecture, illustrated below, meets these criteria and allows you to:

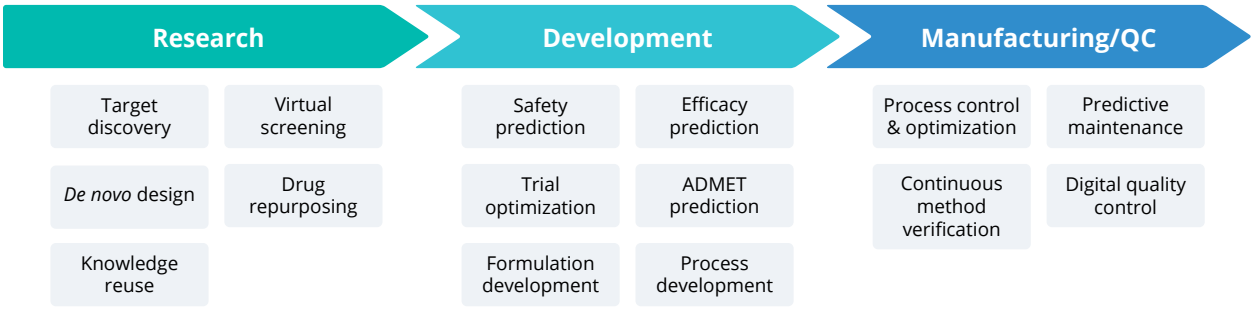
- Automatically collect raw data from sources (e.g., laboratory instruments, informatics applications, and CxOs) and engineer it into AI-ready data (Tetra Data) via automated pipelines.
- Make the data accessible through multiple interfaces, such as web applications, REST API, and SQL.
- Enable a single source of truth for the data using an integrated architecture. When pushed to downstream applications, data isn't duplicated, simplifying data retention and security.
- Ensure data integrity and traceability for regulated environments like GxP with features such as audit trails, checksum assessments, and diagnostic pipelines.



AWS-TetraScience reference architecture

3. Focus on scientific and business outcomes

Customers who have clearly defined the use cases and objectives for AI have achieved the greatest success. Many biopharma companies are moving data to the cloud, but replatforming alone is not enough to fuel AI and deliver tangible results. Data has to be engineered for the specific scientific use case, of which there are many different types across the biopharma value chain (see below).



At the outset of an AI project, consider the scientific and business outcomes. This exercise will help determine which AI projects are worth pursuing and where to allocate your resources. Steps include:

- Evaluate your current workflows and identify the most acute and costly pain points.
- Estimate the value (i.e., return on investment) of an AI solution. Outside experts like TetraScience and AWS can assist here.
- Prioritize your scientific use cases based on ROI.

The use case will guide the development of AI models and, importantly, the underlying data architecture. We advise customers to adopt a "crawl, walk, run" approach to AI:

1. Start with one or a few use cases. Design, build, and test a solution.
2. Scale the solution to other parts of the business once you've achieved the desired outcomes.
3. Take on additional use cases, leveraging lessons from the initial projects.

A biopharma customer who implemented this strategy for digital quality control has reduced deviations by 80 to 90 percent, leading to significant cost savings.

4. Use AI services to scale fast and control costs

A service-oriented approach to AI can help you achieve and expand AI-driven outcomes in several ways:

Get access to expertise and latest technologies	AI as a service provides you with access to a pool of experts and the latest AI tools without the need for extensive in-house development.
Scale easily	Cloud-based AI services offer scalable solutions that adapt to the company's needs. You can start with smaller, more cost-effective implementations and scale up as requirements grow, without upfront investment in hardware and infrastructure.
Reduce development time	Using pre-built AI services can significantly shorten the development cycle. You can quickly deploy AI solutions, and respond rapidly to market changes and new opportunities.
Achieve cost efficiency	With AI as a service, you pay for what you use. This eliminates large upfront investments in AI infrastructure and reduces the ongoing costs of maintenance and upgrades.
Mitigate risks	Using established AI services can help you reduce the risks associated with implementing new technologies. Service providers often have a proven track record and provide support, ensuring more reliable and secure AI solutions.

Next steps

- Watch the webinar on demand to get all the insights.
- Ready to build your data architecture for scientific AI? Connect with TetraScience to get started.