tetrascience

Case Study

Driving efficiency with a future-proof platform and self-service pipelines

Shape Therapeutics (Shape^{TX}) is focused on developing programmable RNA-based medicines to address unmet medical needs. Their mission centers on creating safe, effective, and accessible therapies, particularly in areas such as gene therapy, by harnessing breakthroughs in AI, RNA technology, and synthetic biology. Shape^{TX} recognized the critical need for a future-proof data platform to serve as a central repository for all scientific data, enabling advanced data sharing across teams and locations, ultimately scaling their AI initiatives and accelerating RNA therapy development.

The Challenge

In its earliest days, Shape^{TX}'s scientific progress was slowed by data trapped in silos across local machines, USB drives, and even Dropbox. This fragmented and onerous data flow made it difficult to drive scientific advancements and hindered data sharing and access among teams.

The IT and computational science teams faced significant hurdles handling diverse data formats, along with a lack of reliable backups, clear data traceability, and version control.

They identified a clear business need for a cloud-native, managed platform and a central, searchable repository for immutable raw scientific files, with automated data flow from lab instruments to Benchling Notebook. The team prioritized efficiency and self-sufficiency in implementing the necessary improvements.

Challenge:

Shape^{TX} had scientific data of different formats locked in silos, hindering data-driven scientific advancement. Meanwhile, IT teams needed to become more efficient and self-sufficient.

Solution:

They implemented the Tetra Scientific Data and AI Cloud, combined with TetraScience's technical enablement and self-service pipelines, to streamline data management and empower their teams.

Outcomes:

- Share data on a future-proof data platform
- Create, manage, and deploy code in their own CI/CD framework
- Leverage self-service pipelines for smoother and faster pipeline creation
- Increase IT efficiency with responsive support from the TetraScience enablement team

The Solution

After a rigorous evaluation of several solution providers, Shape^{TX} selected the **Tetra Scientific Data and AI Cloud™** due to its high level of productization, deep industry expertise, and proven best practices in scientific data management.

With TetraScience, Shape^{TX} now ingests raw instrument data directly into the platform, which serves as its central repository. This data, along with links to raw and analyzed files, is then pushed seamlessly into Benchling.

Significant efficiency gains come from TetraScience's **self-service pipelines (SSPs)** and the related technical enablement, which includes comprehensive documentation, onboarding guides, tutorials, and virtual office hours. TetraScience offers a robust set of pre-built pipelines that generate queryable, harmonized Tetra Data and automatically enrich and deliver it to downstream systems. SSPs build on this foundation by enabling users to create custom pipelines tailored to their specific needs. The SSPs include task scripts with code for the business logic to process data, as well as protocols that define the execution order of these scripts. A unique feature enables teams to work in their own development environment with their own dependencies.

SSPs have greatly improved how the team processes data from instrument files through centralization. They create, structure, deliver, and access results in a consistent format, bypassing the need for temporary storage in an S3 bucket.

Initially, the engineering team at Shape^{TX} relied on TetraScience's pre-built pipelines to tag and process instrument files. After evaluating TetraScience's low-code tool—Python exec() for custom code—they recognized the added value of managing and deploying pipeline code with even greater customization. They migrated to TetraScience's full developer solution with SSPs, which provides greater control and smoother integration with their existing CI/CD workflows. Creating a pipeline—from writing a task script and protocol to deploying the Benchling Notebook inbox integration—typically takes only one to two days. Notably, they can deploy directly from their own GitHub organizations and repositories. The team now uses its own CI/CD framework to manage and deploy code, including automated unit testing before deployment.

> The TetraScience self-service pipeline feature provides the unique ability to manage our own pipeline code, define the architecture of the code to match our internal standards, include our own dependencies, and deploy pipelines from our own CI/CD framework. These capabilities allow us to fully customize our pipeline logic to our own data and analysis workflows.
>
> *– Lauren Swanson, Staff Software Engineering Lead, Shape Therapeutics Inc.*

The Outcomes

Scientific data has become a **first-class business asset** for Shape^{TX}. Teams can now seamlessly collaborate, access, and share data via streamlined lab data workflows. This ensures data traceability and provenance, providing scientists with a seamless experience as their data is directly available within Benchling. Improving data flow efficiency for scientists also reduces the burden on IT staff, who now have centralized control over data location and user access.

The technical enablement had an immense impact, making the team **highly self-sufficient.** SSPs empower them to work more effectively with the platform and data, deploy code within their own CI/CD framework, access data faster, and receive data in the most convenient format for the computational team. IT has achieved greater efficiency and reduced bottlenecks.

This solution is also future-proof, as the AI-ready scientific data will allow Shape^{TX} to scale, undertake more complex analytics, and unlock the full value of its scientific data.