



DATA INTEGRATION PLATFORM

# Data Lake for Management

**Experimental data is distributed across silos.** On-premise on the instrument itself. In the cloud in different proprietary software systems. Different sites. Different databases or CRO/CDMO file folders. It is impossible to create a comprehensive view or enable advanced analytics without first centralizing your data. Break down the silos. Activate the power of your data. Manage your experimental data in one central location with a data lake.

---

## Centralize experimental data management

### Flexible

include different types of data sets in a single repository

### Productized

pre-defined data schema, with options to customize



### Scalable

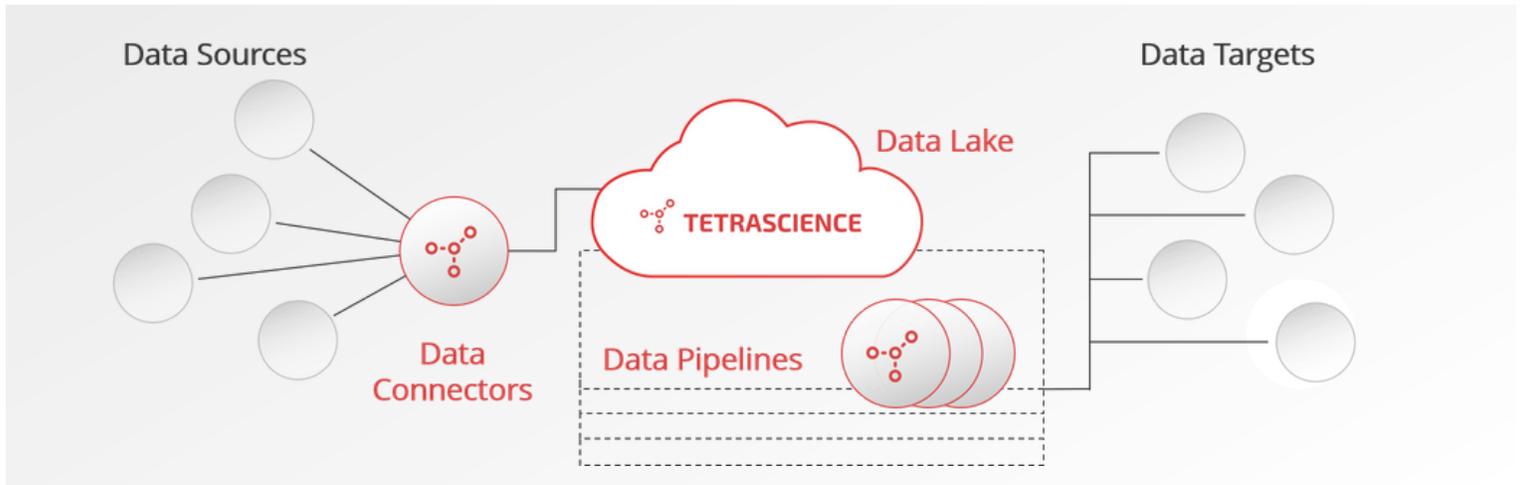
built on AWS S3 to affordably add an infinite amount of data

### Secure

can be deployed as single- or multi-tenant on AWS S3

## DATA LAKE

Manage, search, and store your experimental data in a central, cloud native repository. Built on AWS S3 for industry-leading security and scalability.



### Cloud platform

Built on Amazon (AWS) S3 - an object storage service that offers industry-leading scalability, data availability, security, and performance. Scale your data lake infinitely at an affordable price point. Know that your data is secured by the best in the business

### Deployment

Our multi-tenant SaaS platform can be configured for use by a single tenant and deployed into any AWS account that meets our deployment requirements. This AWS account can be controlled and managed by TetraScience, or it can be controlled and managed by the customer

### Versioning

The data lake versions files, providing complete information about a file's update history. Removing files from the data lake is not a true delete, ensuring you never lose your data, and providing a complete audit trail

### Batch actions

Forget the hours spent manually editing data row by row or screen by screen. Perform batch actions by selecting multiple files within the TetraScience user dashboard by utilizing the built-in actions

# THE POWER OF A CLOUD DATA LAKE

Here are some features that illustrate why so many biopharma companies are rapidly transitioning to the cloud.

## Search + filter

ElasticSearch allows users to find raw and processed data in an intuitive way. Filter and search on just about any data field to find data quickly and easily. Or use SQL + built-in filtering from directly within the UI

## Raw + processed data

The Data Lake is a repository of all files - both raw and processed. Anyone with permission will have access to all their data at the click of a button, whether in the office, at home, or anywhere with internet access

## Batch + stream data

Whether you're dealing with large, discrete data sets or need analysis in real-time, the Data Lake can handle both by supporting batch and stream data

## Custom metadata + tags

Enhance default and schema metadata with a flexible, user-defined process that accommodates organization-specific information. Use metadata + tags for search, filtering, and pipeline triggers

## Data schema

A predefined data schema is built in. The TetraScience Intermediate Data Schema, or IDS, covers the most common experimental data types in the lab. Use as-is or customize

## Ontology

We work with life science companies and industry consortia that are defining common ontologies and data models. The IDS is compatible with the [Allotrope Foundation Ontology](#) and [QUDT](#)

## FREQUENTLY ASKED QUESTIONS

### Is the TetraScience platform compatible with Google Cloud (GCP) or Microsoft Azure?

The TetraScience Platform is built natively on Amazon Web Services (AWS) to leverage the state-of-the-art AWS services. We work with cloud data sources and data targets in GCP and Azure, but the platform itself is supported only on AWS.

### How does the TetraScience platform work with my enterprise data lake?

Life sciences companies use TetraScience as one component of their enterprise data lake - specifically to handle data from experiments, preparing it for data science and AI. Since our platform is built on AWS S3, it is readily compatible with common data lake solutions.

### Is the TetraScience platform a Scientific Data Management System (SDMS)?

As a complete, end-to-end data engineering solution, the TetraScience Data Integration Platform fulfills key functions of a SDMS, plus a whole lot more. Some additional features include:

- automated data acquisition from lab data sources using built-in connectors
- open system design, optimized for data science and analytics
- both batch and stream data handling
- flexible and sophisticated data pipelines to orchestrate the flow of your experimental data, transforming and/or moving the data based on your needs



TetraScience is the leader in transforming the Digital Lab. We provide an advanced data engineering solution that makes life sciences R&D data truly accessible and actionable. More than 80 leading pharmaceutical and biotech companies rely on our cloud-native Data Integration Platform and applications to automatically centralize and harmonize their experimental data, preparing it so you can focus on value-add activities like data science and AI, trend identification, and anomaly detection and response. Activate the flow of your data.

[www.tetrascience.com](http://www.tetrascience.com)  
[hello@tetrascience.com](mailto:hello@tetrascience.com)  
[@TetraScience](#)  

294 Washington St.  
11th Floor  
Boston, MA 02108