

Unlock the Full Potential of Your Scientific Data

Findable, accessible, interoperable, and reusable (FAIR) data
is the key to accelerating breakthrough innovations



EBOOK



Unlock the Full Potential of Your Scientific Data

Findable, accessible, interoperable, and reusable (FAIR) data is the key to accelerating breakthrough innovations

TABLE OF CONTENTS

- 2 Introduction
- 2 Legacy data management: inflexible and unsustainable
- 4 The scientific data journey
- 6 Tetra Data: A solution for all steps of the journey
- 6 Improve productivity and accelerate innovation
- 7 What is Tetra Data?
- 10 Tetra Data in action: accelerating scientific research
- 10 Enabling analytics, visualization, and AI/ML
- 10 Tetra Data simplifies management of a growing volume of scientific data
- 11 What enables production of Tetra Data?
- 12 Conclusion
- 13 About TetraScience



Introduction

For years, life sciences organizations have struggled to gain insights efficiently from fragmented, siloed scientific data. The current state of scientific data diminishes productivity – forcing research and data scientists to spend countless hours searching for data, moving it to centralized systems of record, and preparing it for analysis. This slows discovery, development, and delivery of groundbreaking new therapeutics and adds innumerable risks.

The Covid-19 pandemic has led to a long-overdue examination of legacy data inefficiencies. Life sciences organizations are now prioritizing initiatives to replatform to the cloud and optimize data for use with technologies that accelerate innovation, including analytics, artificial intelligence (AI), and machine learning (ML).

Core to this optimization is making data FAIR (Findable, Accessible, Interoperable, and Reusable). This process involves transforming raw data to standardize it and make it searchable, ready for analytics, and able to flow freely among instruments and applications.

Legacy data management: inflexible and unsustainable

Costs and complexity associated with drug delivery and manufacturing continue to be high:

- Time to bring a new therapeutic to market: **10+ years**
- Total average capitalized pre-launch R&D costs: **\$161 M to \$4.54 B¹**
- New therapeutics ultimately approved after clinical testing: **fewer than 12%²**



For every dollar spent on R&D, less than a dollar of value is returned.³



These challenges waste precious time that might otherwise be spent discovering, developing, and delivering new therapeutics.

¹ Michael Schlander, Systematic Review, 09 August 2021

² DiMasi et al., Journal of Health Economics 47 (2016)

³ NRDD 2019, "Drug Repurposing"



The problem is legacy scientific data:

- Hidden file shares, emailed reports, paper records, databases, and records moved around on thumb drives frustrate efforts to find data
- Hundreds of proprietary scientific data formats and instrument control methods make data hard to extract, parse, validate, and use
- Scientific data are frequently stored with incomplete history and context, making it hard to validate, analyze, or prove compliant

The time spent on these tasks impedes their ability to do higher-value AI/ML and advanced analytics to help bring new life-saving therapeutics to market.

As scientific data volumes increase, so do delays and risks.



Delays in R&D and Manufacturing

- Data is harder to find, extract, parse, validate, transform, and use everywhere
- Time is wasted building one-off software integrations for new instruments and software



Increased Risk of Errors

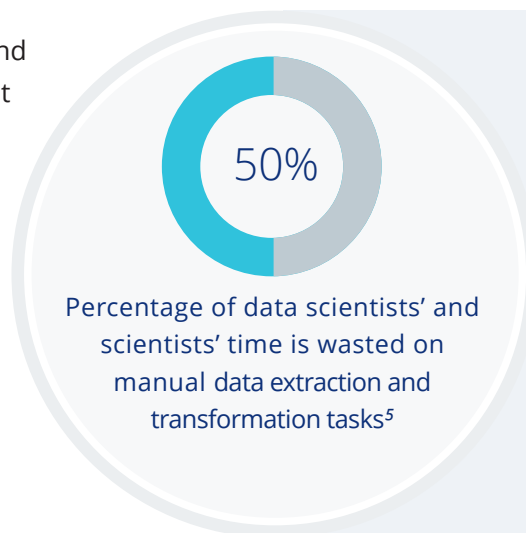
- Manual data entry and spreadsheet data manipulation introduce mistakes
- Errors may make results irreproducible, compel repetition of experiments, and complicate compliance

Opportunity costs also rise.



Legacy data-handling methods can:

- Reduce productivity of scientists, data scientists, and IT specialists, preventing them from focusing on highest-value work
- Frustrate initiatives to extract new value from data, make new discoveries, optimize processes, and operate more efficiently
- Inhibit recruiting of top scientific, data science, and IT talent



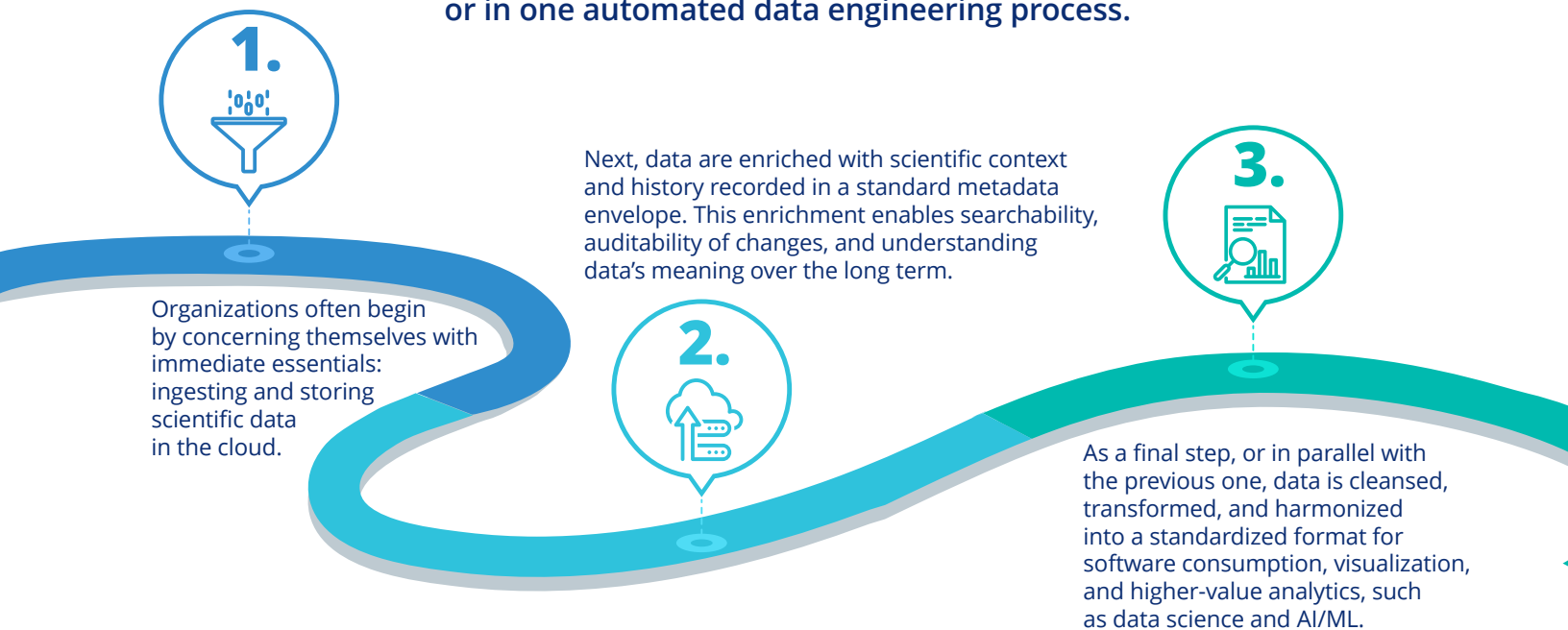
4. David Shaywitz, "Novartis CEO Who Wanted To Bring Tech into Pharma Now Explains Why It's So Hard", Forbes, Jan 16, 2019
5. Source: 2022 State of Digital Lab Transformation Industry Survey: Why Biopharma Is Replatforming To The Cloud



The scientific data journey

To drive innovation, scientific data must be compliant, FAIR, universally available, and prepared for analytics. However, getting there requires multiple, incremental steps.

Making data FAIR is typically approached in several steps, or in one automated data engineering process.



Organizations need a future-proof data model that will meet their needs at all stages of their journey. This need is only met by an open, vendor-agnostic standard for digitalized, scientific data – in a platform that engineers the data to be compliant, harmonized, liquid, and actionable. A solution that simply integrates endpoints without providing scientific context, history, harmonization into a common format, and access to centralized, FAIR data substantially limits the ability to derive value from the data to impact outcomes.

“

The first thing we've learned is the importance of having outstanding data to actually base your ML on. In our own shop, we've been working on a few big projects, and we've had to spend most of the time just cleaning the data sets before you can even run the algorithm. That's taken us years just to clean the datasets. I think people underestimate how little clean data there is out there, and how hard it is to clean and link the data.”

Dr. Sachin Jain, former CMIO at Merck⁴



Step 1: Ingesting raw or primary scientific data

The first problem many look to solve is that of fragmented, siloed, disconnected data across instruments and applications. These data sources are in different formats ranging from files, data collected via APIs, data obtained from SaaS software, or data produced by instruments.

Organizations need to integrate instruments and applications to collect and store data from these sources in the cloud. This first step breaks down silos, reduces the chance of losing data, and provides greater access.

However, stopping at Step 1 would be a costly mistake.



Step 2: Enriching context and ensuring traceable provenance

In their raw form, extracted scientific data may provide very little context. Additional information must be supplied to preserve data's full meaning, to make them searchable and comparable, and to give them history – enabling tracking and auditing of changes.

The second level of scientific data enhancement is to extract contextual metadata – from raw data, systems of record, environmental monitoring and other sources – and use this metadata to enrich raw data with tags and labels (e.g. sample ID, study name, experimenter name, etc.). This process makes data easier to find, preserves the meaning and provenance, and helps maintain an audit trail that provides visibility into changes made to data and configurations, aiding regulatory compliance.

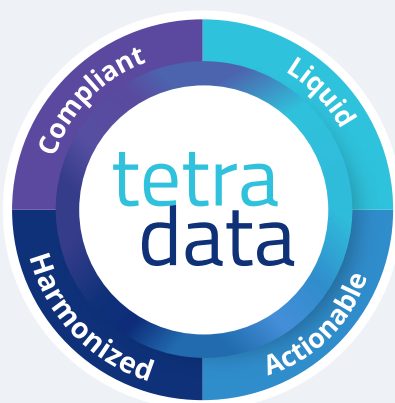


Step 3: Harmonizing across data sources and formats

Contextualized (but raw) data are still stored in divergent, vendorproprietary formats, complicating use, reuse, and interoperability. New instruments and applications add to the number of data formats in play, increasing data complexity when scaling. Engineering data to be FAIR is a desired outcome for most biopharmas today.

Thus, the third step of the scientific data journey is harmonizing data from different vendor-proprietary formats into a unified, open, vendor-agnostic format. Having a known, common format simplifies the process of ingesting data from sources and driving data to targets, easing the process of incorporating new technologies. Additionally, and importantly, a large, unified data set enables analytics, visualizations, and the full power and value from AI/ML to deliver more complete insights to the organization.





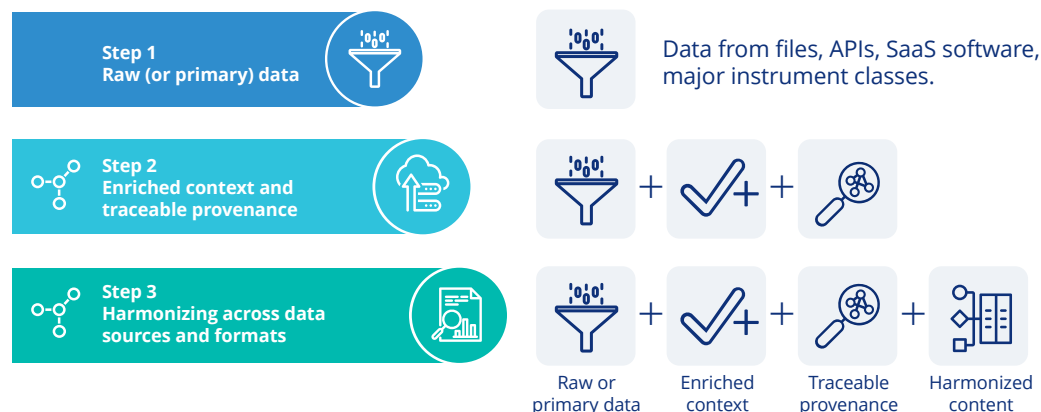
A better model for scientific data

Tetra Data is built around the data integrity best practices, life sciences workflows, and regulatory requirements most critical for biopharma and life sciences. Tetra Data encompasses the characteristics of instruments and their control software, informatics systems, and analytics applications used in research, development, and manufacturing.

Tetra Data is engineered to be compliant, harmonized, liquid, and actionable.

A solution for all steps of the journey

Although an organization's needs may change, it is important to have a data model and strategy that spans the entire scientific data journey. A solution only capable of ingesting and storing data will be unable to provide context, auditability, nor harmonization and will not easily integrate with downstream analytics. Without considering this entire journey, organizations can encounter challenges when scaling, waste time manually moving and processing data, and be unable to generate the insights needed to accelerate innovation. What's needed is a vendor- neutral data model engineered to serve the scientific data journey, end to end.



Improve productivity and accelerate innovation

Tetra Data is created via a sophisticated data engineering process by the Tetra Data Platform (TDP) across the end-to-end data lifecycle. This process extracts and archives raw or primary scientific data from instruments and informatics applications, incorporates metadata providing scientific context, harmonizes the data into an open format, verifies and enriches the data, and then publishes vendor-agnostic Tetra Data to all data targets. Tetra Data becomes instantly available within data lakes and warehouses, visualization and analytics packages, informatics applications such as ELNs and LIMS, and other instruments and control software.

Tetra Data speeds the work of scientists, data scientists, and data engineers. It enhances informatics, enables automation, and unblocks analytics, visualization, artificial intelligence, and machine learning technologies.

In the following pages, we discuss four key characteristics of Tetra Data – all critical for improving productivity and accelerating innovation, enabling Tetra Data to play a central role in scientific data management for life sciences R&D and Manufacturing organizations.

What is Tetra Data?

Tetra Data provides essential data properties organizations need to effectively manage and use their scientific data:



Compliant with GxP and other regulatory requirements



Harmonized from disparate data models into an open, vendor-agnostic format that is searchable and accessible



Liquid to seamlessly flow among instruments and software



Actionable to speed automation and enable analytics, visualization, and AI/ML without manual processing

With FAIR Tetra Data, your data is far easier to locate, join, compare, consume with software, and easily transform for any application you can imagine.





Compliant

The data regulatory landscape across life sciences is complex, and ensuring compliance becomes more difficult as data volume increases, uses of data expand, and new laws and regulations are introduced. Compliance is critical, and violations include fines, business disruption, loss of productivity, and damage to revenue and reputation.

A platform engineered to produce Tetra Data also enables data processing integrity, reusable integrations, pipelines, and pipeline availability monitoring.

Tetra Data enables GxP compliance and data security by supporting:

- SOC 2 Type 2 validation
- 21 CFR part 11 / Annex 11
- Data integrity protections, including end-to-end data audit trails and role-based data access

Compliant Tetra Data provides history, transparency, and security to:

- Provide greater visibility into scientific data provenance through context enrichment
- Identify changes to data or system configurations through a complete audit trail



Harmonized

Harmonization is a cornerstone of Tetra Data, making scientific data findable, accessible, interoperable and reusable (FAIR). It is the process where data of different types and formats are evaluated, transformed, and aligned to fit and work within a single, open, common vendor-neutral format. Harmonization breaks down silos, standardizes formats, and stores data to a centralized location. You can easily locate relevant Tetra Data, make queries, assemble large datasets, and leverage software (e.g., AI/ML) to uncover insights. For example, chromatogram records for an AKTA or Shimadzu HPLC device are very different. By harmonizing into a common format and enriching data with scientific context, you increase your ability to find all chromatogram records on a particular date, see all experiment records for a particular scientist or any number of complex queries.

Harmonized Tetra Data connects and cleanses all data into a common format with consistent semantics, enabling:

- ✓ Rapid search on data contents
- ✓ Easier aggregation of large datasets
- ✓ Simpler, more robust, and more reusable integrations with instruments, bioinformatics software, and analytics applications
- ✓ Easier data-sharing across the organization with FAIR (findable, accessible, interoperable, reusable) data enriched with scientific context





Liquid

As life sciences organizations begin to realize the benefits of FAIR, enriched, and harmonized Tetra Data, a ratcheting effect kicks in. New users benefit from robust, functional, automated dataflows that save time and eliminate manual effort. Adding a new instrument or other data source is simplified. As more data sources and targets are connected, automating a new workflow becomes an incremental challenge, rather than a painful lift. Inefficient, manual steps are identified and automated away. The pace of work increases, with scientists and IT experts refocusing their efforts on high-value tasks.

Liquid Tetra Data seamlessly flows across laboratory instruments and informatics applications

- 1 **Expand access to data** in scientists' preferred tools, hastening insight and enabling focus on highest-value work
- 2 **Flow data between sources and targets** painlessly and effortlessly, eliminating manual processes, and making it faster to incorporate new technologies
- 3 **Empower collaboration** by making data shareable across a global workforce, partners, and CROs and CDMOs



Actionable

Tetra Data allows scientific data to become a critical asset that drives business decisions and improves the direction of scientific research. Data is no longer simply collected and managed, but becomes a key driver of value. Organizations use Tetra Data to drive data analytics in order to identify trends and market opportunities for new drugs and treatments. Applying data analytics no longer needs to be front loaded with a time-consuming, manual preparation process.

Actionable Tetra Data eliminates the manual, time-consuming processing needed for:



Advanced analytics/visualizations

Gain insights from data and optimize R&D and Manufacturing



Automation

Leverage robotic instruments to speed work and enhance repeatability



AI/ML

Use artificial intelligence and machine learning to tackle big problems

Tetra Data in action: accelerating scientific research

Scientists rely on ELN software to iterate on experimental designs and record results. However, much of this work is manual with data imported and manipulated in Microsoft Excel and moved to the ELN before scientists can work with them. This process is time consuming with the potential for introducing errors.

With harmonized and centralized Tetra Data, experimental data are delivered from unified storage directly to the ELN in a standardized format without the need for custom integrations or data transformations. This same Tetra Data is also available to data analytics applications for easy visualizations and insights.

RESULTS:

Data on all interfaces are consistent and up to date, scientists no longer need to manually move and process data, and the risk of errors is greatly reduced.

Enabling analytics, visualization, and AI/ML

Visualization tools, such as Tableau or TIBCO Spotfire®, and AI/ML systems can provide powerful insights to drive business decisions and produce improved outcomes. However, the data must first be prepared, including curating, cleaning, and formatting. Data sets must also be large enough to create analytical models, which often requires assembling a uniform data set from multiple, siloed sources. Formed by collecting data across all sources, harmonizing them into a common format, and enriching them with additional metadata, Tetra Data is ideally suited for analytics, such as AI/ML, and greatly reduces the time and cost to prepare data for these applications.

RESULTS:

By collecting data from all sources, harmonizing them into a common format, and enriching with metadata, Tetra Data reduces the time and cost to prepare data for AI/ML applications.

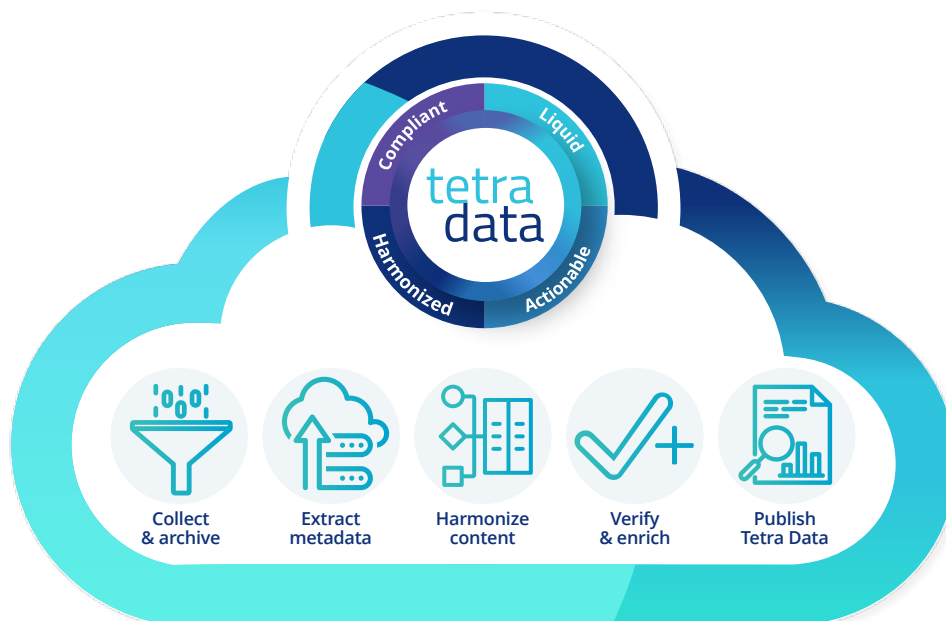
Tetra Data simplifies management of a growing volume of scientific data

With Tetra Data, entire life science organizations have access to high-quality data where they need it, when they need it, eliminating painful, manual preparation. These organizations can spend less time collecting, searching for, and preparing scientific data and more time using it to guide the discovery and manufacturing of ground-breaking treatments to improve human life.



What enables production of Tetra Data?

Tetra Data is an open, vendor-neutral scientific data format, usable by anyone. In the context of biopharma and biotech operations, Tetra Data is generated by a disciplined data engineering process that implies several prerequisites:



- **An end-to-end, data-centric platform**

Tetra Data Platform, from TetraScience, is a scalable, cloud-native solution, purpose built for biopharma requirements and use cases, designed to ingest and store scientific data, enrich them with context, then parse, transform, and harmonize them into Tetra Data for general use.

- **Productized integrations**

Produced by TetraScience in collaboration with members of Tetra Partner Network, Tetra Integrations are robust and customizable, enabling data to flow efficiently between sources and targets.

- **A broad ecosystem of partners**

Members of Tetra Partner Network collaborate with TetraScience, using Tetra Data and Tetra Data Platform to rapidly integrate vendor solutions and maintain them.

- **Tetra Scientific Data Cloud™**

The above combine to form the Tetra Scientific Data Cloud, accelerating and improving scientific outcomes. Data across sources are enriched with metadata and engineered into an open, vendor-agnostic format for deeper analysis, data science, and AI/ML.

Conclusion

Life sciences organizations typically take the following steps along their scientific data journey:

1. Organizations first integrate instruments and informatics applications to connect and centralize their fragmented, siloed data
2. By enriching the data with scientific content, organizations gain meaning, information about how the data were generated, and a historical record
3. Finally, organizations harmonize data from different vendor formats into a unified, open, vendor-agnostic format, enabling analytics, visualizations, and AI/ ML

Tetra Data helps organizations fuel the innovation needed to bring game-changing drugs and treatments to market faster through compliant, liquid, harmonized, and actionable scientific data.

A new standard is emerging – Tetra Data – FAIR data, engineered to support the scientific data journey, end to end.



About TetraScience

TetraScience is the Scientific Data Cloud company with a mission to accelerate scientific discovery and improve and extend human life.

The Tetra Scientific Data Cloud™ is the only open, cloud-native platform built for scientific data that connects lab instruments, informatics software, and data applications across the biopharma value chain and delivers the foundation of harmonized, actionable scientific data necessary to transform raw data into accelerated and improved scientific outcomes.

The Tetra Partner Network is the largest ecosystem of partners dedicated to unlocking the power of scientific data for pharmaceutical and biopharmaceutical customers. This network of lab instrument, informatics applications, CRO/CDMOs, analytics, system integrator, and data science partners creates seamless interoperability and an innovation feedback loop that will drive the future of life sciences and the delivery of life-saving therapeutics.

TetraScience currently counts 15 of the top 25 global biopharma companies as customers as well as leading biotechs and CROs/CDMOs.

Learn more, get a demo

To learn more about how Tetra Data can unlock the potential of your scientific data, please visit tetrascience.com



© 2022 TetraScience, Inc.

