tetrascience

Scientific Data Imperative: Establishing a New Benchmark After Two Decades

Why life sciences firms need to rethink their data platforms

RESEARCH REPORT

aws

Scientific Data Imperative: Establishing a New Benchmark After Two Decades

TABLE OF CONTENTS

Introduction

Biopharma companies are on a mission: deliver better therapeutics to patients faster to improve human health. Fulfilling this mission today still requires upwards of **10 years**[1], an average cost of **$1.1 billion USD**[2], and total cost as high as **$4.54 billion**[3] for each newly approved drug.

How can we accelerate drug discovery, development, and delivery? To achieve these ends, biopharmas have driven digital transformations in earnest. They've turned to technologies like artificial intelligence (AI) and machine learning (ML) to build predictive models from existing data sets. Unfortunately, AI/ML outcomes heavily rely on the quality of this underlying data.

The **TetraScience 2022 Biopharma Executive Survey** of 500 biopharma executives reveals an imperative for success in biopharma digital transformation lies in maximizing the value of scientific data across the value chain [Discovery --> Development --> Manufacturing] to impact three key industry drivers:

1. **Speed**: Learning from experiments faster, reducing time to market for final molecules
2. **Cost**: Improving productivity, increasing operational efficiency, increasing return on investment (ROI)
3. **Risk**: Improving safety, reducing errors, improving compliance with regulatory requirements

The Problem

For decades, biopharma leaders have attempted to unlock the full value of their scientific data. However, the rules of the game keep changing. Coping with the 5 "Vs" of Big Data (velocity, volume, value, variety, and veracity), incorporating new modalities, and making all data findable, accessible, interoperable, and reusable (FAIR) continually complicate the scientific data landscape. Pursuing breakthroughs creates a Catch-22, in that heroic advances remain hampered by fragmented data trapped in thousands of proprietary formats. Furthermore, informatics professionals can't easily utilize scientific data from R&D through QA/QC and Manufacturing.

1 Joseph A. DiMasi, Henry G. Grabowski, Ronald W. Hansen, Innovation in the pharmaceutical industry: New estimates of R&D costs, Journal of Health Economics, Volume 47, 2016, Pages 20-33, https://doi.org/10.1016/j.jhealeco.2016.01.012.

2 Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA. 2020;323(9):844–853. https://doi:10.1001/jama.2020.1166

3 Schlander, M., Hernandez-Villafuerte, K., Cheng, CY. et al. How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment. PharmacoEconomics 39, 1243–1269 (2021). https://doi.org/10.1007/s40273-021-01065-y

A New Benchmark

$100M in lost productivity for mid-size biopharmas due to fragmented, siloed data, and **1 million hours** that could be redirected toward scientific research

$1.26 million in cost for each one-day delay in getting a drug or therapeutic to market

7X more likely to have to repeat experiments due to data issues for organizations without scientific data in the cloud

3.7X more spent as proportion of revenue by cloud "laggards" for legacy scientific data management infrastructure compared to those advanced in their journey to the cloud

The Solution

The problem of "siloed data everywhere" can be effectively addressed by replatforming scientific data to an open cloud platform while transforming the data to FAIR principles. When data are compliant, harmonized, liquid, and actionable, benefits accrue even faster. The TetraScience 2022 Executive Survey reveals both the consequences of today's status quo for scientific data "laggards" and a path forward -- offering near and long-term operational benefits that can make FAIR scientific data in the cloud an economic and operational game-changer.

Below, we'll expand on these brief ideas and show you how things change with the Cloud.

Why Change from the Status Quo?

Survey respondents made the difference very clear between scientific data in the cloud "laggards" vs cloud leaders.

Today, only 23% of organizations have fully re-platformed their scientific data to the cloud or are cloud-native. We call these cloud leaders. 41% are partially replatformed; 33% are laggards who face competitive disadvantages because they have either just started their digital transformation journeys or haven't even planned to start. Respondents said that only 50% of scientific data in their organization (excluding inherently private or confidential data) is sufficiently prepared and available for analysis and data science applications. And just slightly more than 50% of data is considered FAIR.

Speed

On average, respondents said they wanted to reduce their time to market for new drugs by one year vs. current time to market. Biopharma executives bear substantial economic penalties for delivery delays.

> Each **one-day delay** in getting a drug or therapeutic to market costs their company an average of **$1.26 million in lost revenue** and/or loss of first-mover advantage.
>
> Each one-month delay increases the risk of not getting a new product to market by 30%.

Strategic partnerships, M&A, AI-driven development, and novel clinical trial approaches all play key roles in accelerating time to market. However, underlying the success of all of these is the key area of foundational data driving drug discovery and development. This spans improving the value and usability of scientific data as well as increasing IT staffing so more resources can make that data available to scientists and data scientists.

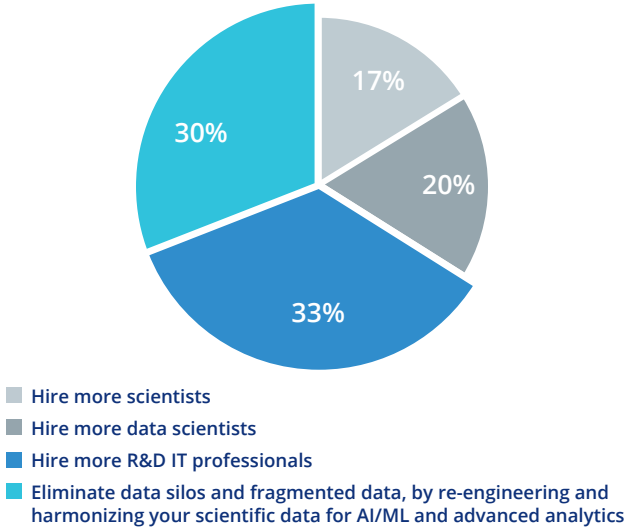
Just Hire More Smart People?

However, just hiring more scientists or data scientists doesn't solve our data woes. When asked, "Which would help your firm the most to accelerate time to market for new therapeutics?", the top ranked responses were "eliminate data silos and fragmented data by re-engineering your scientific data for AI/ML and advanced analytics" and "hire more R&D IT professionals." Nearly 70% of all C-Level respondents chose one of these two responses. That dwarfs by a factor of 3x the number of respondents who thought hiring either more scientists (17%) or data scientists (20%) would be the highest leverage to accelerate time to market.

No one doubts the value of scientists and data scientists to advance development and delivery of new drugs. We interpret the survey responses to mean that a solid foundation of engineered data supported and backed by IT is the best leverage point to enable higher-value work by skilled scientists and data scientists.

When asked by what percentage time to market could accelerate if data scientists, scientists, quality assurance, manufacturing, and partners had FAIR scientific data in the cloud, all respondents—including scientists, data scientists, R&D IT executives, and C-Suite executives—on average saw a 65% speed-up. "In my view, it can accelerate innovation and reduce the time needed to bring a drug to market," said a Senior Director of R&D IT at a mid-size Benelux biotech firm.

Which Would Help Your Firm the Most to Accelerate Time to Market for New Therapeutics?

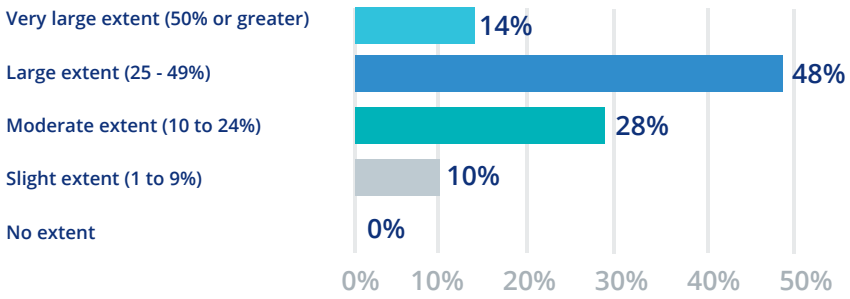


> "FAIR data automation frees up a lot of time for scientists and data scientists and allows them to focus on science rather than manual data collection."
>
> - CEO, U.S. mid-sized biotech

65% acceleration in time to market expected if data scientists, scientists, QA, manufacturing and partners had FAIR scientific data in the cloud.

FAIR data: key driver of faster time to market. 76% of respondents said FAIR harmonized data in the cloud would drive faster time to market at levels ranging from 10-24% up to more than 50% improvement. 2/3 of the C-Suite at companies advanced in their scientific data journey to the cloud reported gains in time to market from FAIR harmonized data in the cloud of greater than 25%.

2/3 of Those Advanced in Journey to Cloud See Large Gains in Time to Market



Even if actual percentage gains were to be less than the above findings, if time to market is accelerated by several months to a year, the impact of avoiding the $1M+ cost of each one-day delay or risk of a drug never coming to market can be worth billions to a large pharma. Furthermore, it can make the difference between a smaller biotech successfully hitting IND application milestones and eventually getting to market with their company's only one or two critical drugs or failing to gain market traction or partnerships key to their survival.

> "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."
>
> - Vas Narasimhan, CEO of Novartis

Source: "Novartis CEO Who Wanted To Bring Tech into Pharma Now Explains Why It's So Hard", Forbes, Jan 16, 2019

Cost

Informal surveys and hundreds of customer discussions indicate that 50-80% of scientists' and data scientists' time is spent on low-level data extraction, cleansing and manipulation tasks in order to get raw or primary data ready for higher value analysis. The results of the TetraScience/PharmaIQ market research report: **"2022 State of Digital Lab Transformation in Biopharma"** validated that scientists and data

Percentage of Time Scientists and Data Scientists Spend on Manual Data Extraction and Transformation

50%

scientists spend ~50% of their time on low-level data wrangling. For a biopharma with 1,000 scientists and data scientists at a fully-loaded cost of $200,000 per FTE, **this represents $100M in lost productivity and roughly 1 million hours a year of time that could more productively be invested in scientific development to advance delivery of new therapeutics.** At an individual division level, the R&D IT director for a U.S.-based biopharma expects for just one division that at least $20 million of scientist and data scientist time currently spent on low-value wrestling with data will be repurposed to higher-value data analysis.

Consider, too, the cost and burden of legacy point-to-point integrations. As a proportion of revenue, cloud "laggards" spend 3.7X more on inefficient infrastructure compared to cloud leaders. Furthermore, brittle, custom-built integrations inhibit the ability to adopt and integrate new data sources and technologies that could help accelerate time to market.

Cloud "laggards" **spend 3.7X more** on inefficient infrastructure compared to cloud leaders

Most C-Suite executives believe that dramatically higher quality scientific data that's FAIR, compliant, harmonized, liquid, and actionable in the cloud could help to drive fast, step-change improvements in costs over today's levels.

Bring Down Fixed Costs

Eight in ten executives say that FAIR data could reduce costs by greater than 10%. Executives at firms that are advanced in re-platforming FAIR data to the cloud are even more convinced, with nearly three-quarters seeing a large to very large impact on costs (saving in the range of 25% to 50% or more).

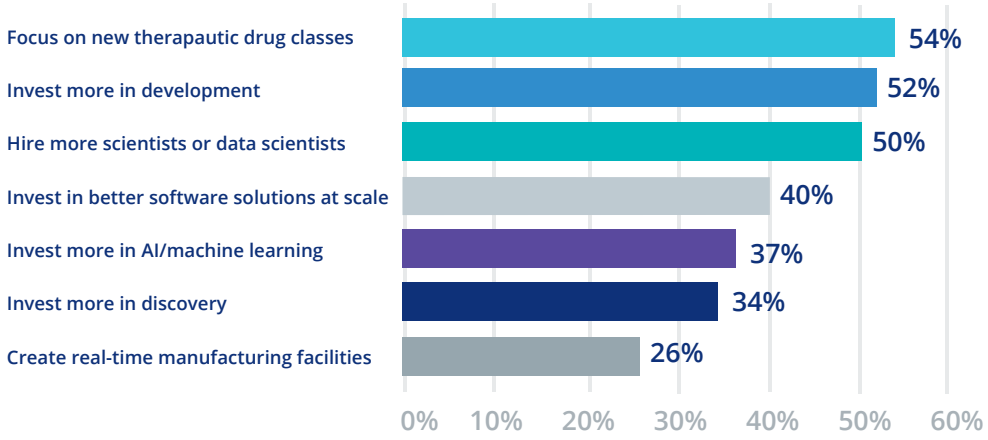
% Citing Potential Impact of Much Higher Quality Data on Reducing Costs

Impact	Advanced	Other	All
Very large extent (50% or greater)	21%	10%	12%
Large extent (25 - 49%)	52%	23%	28%
Moderate extent (10 to 24%)	28%	50%	46%
Slight extent (1 to 9%)	0%	16%	13%
No extent	0%	1%	1%

The magnitude of the above cost reduction findings represents hundreds of millions of dollars. Just one part of that cost reduction is seen for those who retire legacy infrastructures and replatform to the cloud. For biotechs, starting out cloud-native dramatically reduces costs. Companies that become advanced in using FAIR data in the cloud reported IT infrastructure **savings for Point-to-Point integrations of 73% compared to those who are very early in their cloud journey.**

> “It enables accurate data exchange with other companies, reducing the risk and cost of drug research while also allowing for speedier innovation.”
>
> - Chief Science Officer, U.S. biotech company

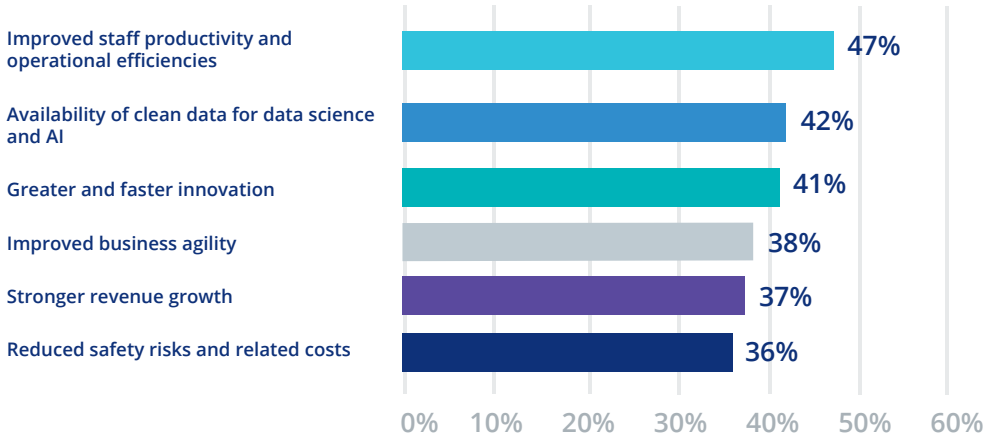
Top Use of Funds Saved from Avoidance of Custom Point-to-Point Integrations



Reducing IT infrastructure costs would free up companies to put their money to much better use. According to our survey, organizations that did not need to spend money on linking up siloed data systems would put those funds back into more productive areas, such as exploring new therapeutic drug classes, investing more in development, hiring more scientists or data scientists, and investing in better software solutions at scale.

FAIR harmonized cloud data boosts productivity for 47% of surveyed firms. For biopharma firms, the best benefit results from higher staff productivity, since researchers can get more research and analysis done in less time, which leads to greater and faster innovation, improved business agility, and stronger revenue growth.

Today’s Benefits of FAIR Data in the Cloud



Risk Reduction

As stated previously, risk reduction – by means of improving drug safety, reducing errors, and improving regulatory compliance – is a key driver in the life sciences industry. Organizations like the Food & Drug Administration (FDA) and the European Medicines Agency (EMA), along with professional organizations like the International Society for Pharmaceutical Engineering (ISPE) expect high standards of the resultant diagnostics, therapies, and treatments put into patients.

Missing and incomplete data across the value chain contribute substantially to risk. Our survey reveals a considerable amount of wasted time by researchers who are forced to repeat experiments in order to avoid errors due to missing or incomplete data, and how cloud "laggards" are impacted relative to their more advanced peers.

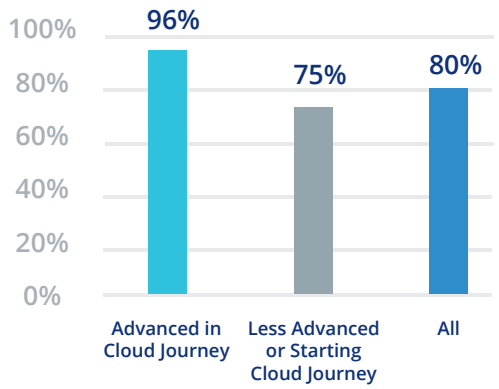
Organizations that have not yet begun putting their scientific data in the cloud or are early in their journey are 7X more likely to have to repeat experiments due to missing or incomplete data. Only 8% of those that are cloud-native said they have to repeat experiments due to such issues, whereas, 88% of those with no plans today and 60% of those in the planning stages reported they were repeating experiments for this reason.

> "FAIR adoption can help unlock the long-term potential of data for future study, increase collaboration, speed up the response time to scientific queries, and satisfy compliance needs."
>
> - Chief Science Officer, U.S. biotech company

80% of all respondents said FAIR compliant, harmonized, liquid, actionable data in the cloud would have a "moderate" to "very large" qualitative impact on reducing safety risks. That percentage goes up to 96% for respondents who are already advanced in their cloud journey and are fully replatformed or cloud-native.

"FAIR data will boost operational efficiency and make pharma practices safer, resulting in better patient care," said the CEO of a small Canadian pharmaceutical

% C-Level Execs Citing Moderate to Very Large Impact of Superior Data Management on Reducing Safety Risks



firm. Agreed the CIO of a small US biotech: “Adoption of FAIR data in the cloud will aid in the provision of detailed information regarding the safety and efficacy of pharmaceutical goods, as well as indicating who is eligible to use and benefit from the drug and highlighting major adverse effects, interactions, and potential dangers.”

Conclusion

Today’s economic and competitive landscape in biopharma strongly depends on accelerated time to market for new drugs, cost minimization, and reduced risk. Reliable, high-integrity, available scientific data underpins both patient health and pharmaceutical company strategy.

Until recently, no commercially available technology platforms were powerful enough to solve the problems we’ve presented. Today an off-the-shelf cloud-native platform can automatically engineer FAIR data in the cloud that is also compliant, harmonized, liquid and actionable. While the cost of the status quo is high, the near-term and long-term benefits of this new approach are staggeringly substantial.

Having all scientific data in a form that’s FAIR, compliant, harmonized, and actionable in the cloud for even one function, such as R&D, yields substantial benefits. Beyond that, bringing scientific data together from R&D through downstream manufacturing and QC will yield long sought-after breakthroughs including new process development efficiencies, analytical control strategy, technology transfer acceleration, and reduction of manufacturing risk.

Failure to act means risking the mission of improving patient health. The upside impact of the new world of harmonized actionable data in the cloud could accelerate the path to getting new life-saving therapeutics to patients and also unlock trillions of dollars in potential value to biopharmas.

Case Study

A multi-billion dollar U.S-based biotechnology company faced problems of fragmented, siloed scientific data across their organization and wanted to boost productivity of their R&D and IT staff as well as **cut a year off their average time to market of 10 years for new therapeutics.** By integrating and engineering data from hundreds of different instruments using TetraScience’s Scientific Data Cloud, they achieved some of the following accelerated and improved scientific outcomes in a matter of months:

SHIFTED 76 HOURS/WEEK: from manually preparing CRO data to assessing treatment effectiveness in clinical trials. Increases likelihood of achieving study’s endpoint faster. This saves two FTEs and redirects that time to higher value efforts.

90% FASTER DETECTION: of pipetting errors occurring in one of every ten robotic liquid handler jobs, resulting in increased genotyping throughput of lab mice.

The Tetra Scientific Data Cloud™ also frees up time for scientists to be greater thought leaders, supports work-life balance for scientists, and drives greater talent retention.

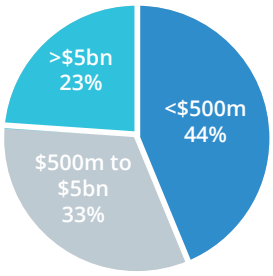
> “FAIR data in the cloud has boosted our efficiency, prevented work from being repeated unnecessarily, and allowed us to ask questions we could never have asked before.”
>
> - Head of R&D IT at a mid-size U.S. biotech/biopharma

> “With TetraScience, the five scientists we have working on a project are like having 50.”
>
> - Large global biopharma

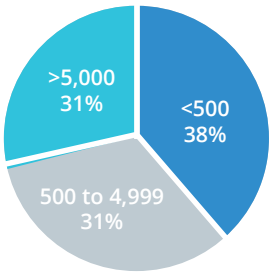
About the Survey

This report is based on research commissioned by TetraScience and conducted by ThoughtLab Group, a specialized technology thought leadership research firm. 500 C-level executives, IT R&D leaders, and leaders of research science and data science were surveyed in April and May of 2022 across biotech and biopharma organizations in North America and Europe.

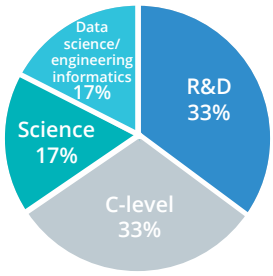
% of Firms by Revenue



% of Firms by Employees



% of Respondents by Function



Regions:

United States	Ireland	Germany
Canada	France	Nordics (Finland, Norway, Sweden, Denmark)
United Kingdom	Switzerland	Benelux (Belgium, Netherlands, Luxembourg)

Get Started

To learn how the **Tetra Scientific Data Cloud** accelerates drug discovery, development, QA, manufacturing, and delivery, schedule a demo at **tetrascience.com** or learn more about the **Tetra Scientific Data Cloud in AWS Marketplace.**

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 how the Tetra Scientific Data Cloud accelerates drug discovery, development, QA, manufacturing, and delivery, schedule a demo at tetrascience.com.

Get started with TetraScience Scientific Data Cloud on AWS Marketplace.

Corporate Headquaters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115
Atlanta | 3280 Peachtree Road NE, 7th Floor, Atlanta, GA 30305