Pharma iQ

2022 State of Digital Lab Transformation in Biopharma

Biopharma cloud adoption accelerates, but obstacles remain

Learn how improved, cloud-based data management can enable life sciences organizations to speed lab digitization, provide FAIR data to accelerate research and scientific outcomes, and adapt to changing business needs

Inside:

- Bench scientists and data scientists report still spending about 50 percent of their time unifying, cleansing and enriching data. What's the fix?
- Less than 18 percent of scientific data has been replatformed to the cloud, and an even smaller percentage is considered FAIR (Findable, Accessible, Reusable, and Interoperable). What's the holdup? And what is it costing biopharma?
- More than 80 percent of survey respondents claim maintaining point-to-point, DIY integrations is highly burdensome.

Brought to you in association with:

tetrascience

Introduction

Pharma has long lagged behind other industries in adopting cloud and other modern IT technologies. But the pace of change is now accelerating, driven by business challenges and black swan events such as the global pandemic. Cloud platforms, in tandem with breakout technologies like machine learning, have enabled organizations like Moderna to **meet pandemic challenges** and transform aspects of their business.

To find out where biopharma organizations sit with respect to their approaches to cloud technologies, *Pharma IQ* teamed up with R&D data cloud provider TetraScience to survey industry insiders and ask about their organizations' strategies for cloud adoption, digital labs, FAIR data and machine learning (ML) and artificial intelligence (AI) solutions.

The results of the *2022 State of Digital Lab Transformation in Biopharma Survey* highlight that the case for cloud adoption in biopharma has never been stronger, and while some organizations are still behind, many are demonstrating that technological progress is well underway.

Contents

"There are a large number of 'frustrated' scientists and engineers who cannot gain access to the core data they need to answer a query in a timely fashion."

Mike Tarselli

Chief Scientific Officer at TetraScience

Executive summary

The *2022 State of Digital Lab Transformation in Biopharma Report* is the industry's latest assessment of biopharma R&D data cloud adoption.

Biopharma faces massive challenges today, as leadership teams cope with increasing pressures to be more efficient, shorten drug-delivery timelines, and increase the number of therapeutic candidates in pipeline to maintain profitability in the post-blockbuster-drug era. Pharma IT and operations, meanwhile, cope with increasing data proliferation, legacy data silos and management methods, and the looming threat of new and ever-larger volumes of data to be processed. Scientists and data scientists – the "point of the spear" for discovery – spend ever-larger fractions of productive time in data cleansing, engineering, management, and transformation – activities, arguably costing millions to tens of millions of dollars in productivity that could otherwise be used to proceed with higher-value fundamental research and extracting new value from scientific data.

Definition:

"R&D data cloud" is an emerging biopharma IT market category of technology solutions that respond to biopharma IT and strategic challenges by enabling digital lab transformation and managing scientific data through its whole lifecycle. An R&D data cloud strategy is key to leveraging automation, analytics, artificial intelligence (AI) and machine learning (ML), and other game-changing technologies. It reduces drag and manual data management, speeds up drug discovery, and advances new therapeutics to market more swiftly, efficiently and predictably.

R&D data cloud proponents identify two related processes as critical to this undertaking:

Replatforming scientific data to the cloud: a process that, per our survey, is now most often first undertaken to enable scale and resilience, and to control costs.

Engineering solutions to make scientific data universally Findable, Accessible, Interoperable, and Reusable (FAIR) within an organization: essential to eliminating the large overheads associated with manual data-handling, and clearing the way for efficient automation, use of analytics and AI/ML.

The *2022 State of Digital Lab Transformation in Biopharma Report* takes the pulse of progress towards R&D data cloud and digital lab enablement on both these critical fronts: cloud replatforming and FAIR data provisioning. Respondents included Senior Data Scientists, IT Managers, Lab Heads and Enterprise Architects from the likes of Johnson & Johnson, AstraZeneca, Boehringer Ingelheim and Sanofi.

Key takeaways

- *Basic data-to-cloud replatforming and FAIR data provisioning is now accelerating*, with 74 percent of respondents reporting that their organizations are planning or implementing a replatforming/FAIR data strategy, and 18 percent already reporting completion. (Figure 1)

- *But blocks still exist.* Biopharma organizations who are hesitating to move data to the cloud still cite concerns about cloud security for sensitive data, still exists around cloud security (50 percent report discomfort with moving sensitive data to the cloud), uncertainty around costs, and lack of critical cloud skills on staff. (Figure 2). Additionally, 49 percent of users (Figure 3) report that "less than half" or "none" of their data can now be considered FAIR.

Meanwhile, lack of cloud-resident FAIR data is impacting scientists and biopharma bottom lines:

- *Major time is still being wasted by scientists (49 percent of working hours) and data scientists (52 percent of working hours) on manual data retrieval, and transformation.* Additionally, 66 percent (Figure 10) have trouble accessing data, reporting delays from 1 hour to >24 hours in getting hold of data for analysis. This can lengthen R&D timelines, -- wasting millions of dollars across the delivery-cycle for a non-blockbuster drug -- and can severely delay ROI for AI/ML and other acceleration and automation technologies.

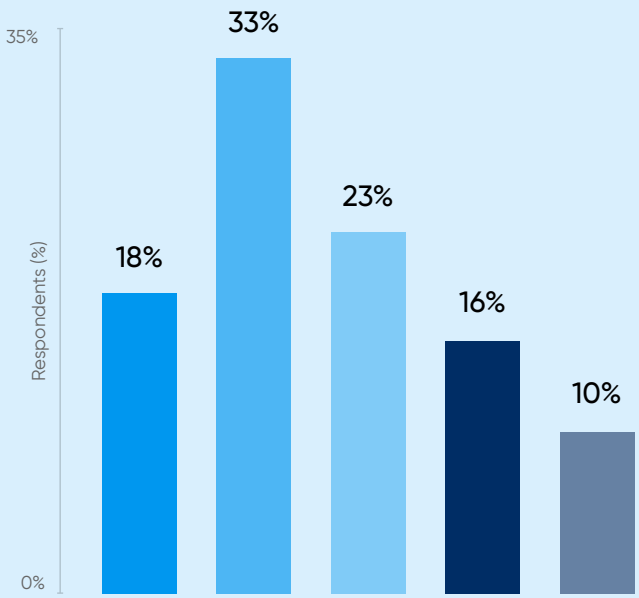
Adhering to FAIR principles while moving to the cloud

As revealed by the respondents to the *2022 State of Digital Lab Transformation in Biopharma Survey*, 74 percent or more of biopharma/life sciences organizations are now planning, implementing, or have completed the process of moving scientific data to the cloud, including 18 percent saying they had fully replatformed all of their scientific data on the cloud (see Figure 1). Some 33 percent of respondents said that they believed that their organizations were at the beginning-to-middle stages of implementation, while a further 16 percent had not yet started the move or simply had no plans to.

FIGURE 1

Does your organization have an initiative to replatform scientific data on the cloud, so that it is considered FAIR (Findable, Accessible, Interoperable and Reusable)?

- Yes - we have fully replatformed all our scientific data to the cloud
- Somewhat – my organization is in the beginning-to-middle stages of implementation
- Not yet – my organization is planning to replatform scientific data on the cloud
- No – we have yet to start replatforming scientific data on the cloud, or it is not part of our business strategy
- Not sure – someone else in my organization is responsible for this



Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

Of those whose organizations have yet to replatform scientific data to the cloud (see Figure 2), 50 percent stated that they were "uncomfortable with moving sensitive data to the cloud", with a further 25 percent either satisfied with their current management of scientific data or anticipating that the associated costs would be too high.

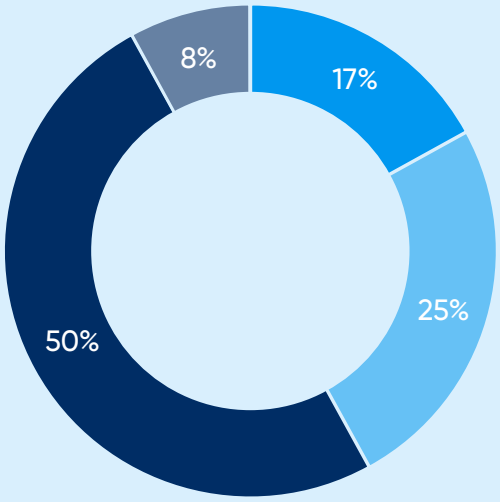
These results reflect a prevalent attitude that considers the huge volume of data life sciences organizations generate, acquire and use daily, which Peter Henstock, ML and AI Technical Lead at Pfizer, describes as "a tremendous challenge" to collect, organize and make available.

Henstock continues: "This includes all legacy data going back decades under many different technologies. It's more challenging than scanning all your old photos of your household and adding captions and names in them so they are searchable and available."

Tarselli congratulates the 18 percent of organizations that have fully replatformed to the cloud, however, declaring: "These companies are the future and are what we all hope to direct the rest of the industry toward. We're glad to see somebody showing the way."

FIGURE 2

Why is your organization not replatforming scientific data to the cloud?

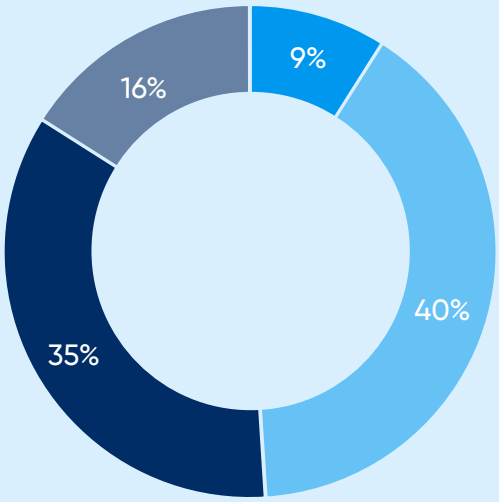


- We see the benefits but are not able to get started
- We are satisfied with our current management of scientific data
- We are uncomfortable with moving sensitive data to the cloud
- We anticipate the associated costs will be too high

Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

FIGURE 3

How much of the scientific data in your organization are considered FAIR (Findable, Accessible, Interoperable and Reusable)?



- None
- Less than half
- More than half
- All of the data currently used are FAIR data

Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

Surprisingly, just 16 percent considered that all of their data was FAIR, with those who say that more than half of their data was FAIR accounting for 35 percent (see Figure 3). A further 40 percent said that either less than half of their data or none at all was available for analysis and data science applications. These findings suggest that a majority are actually on the right path when it comes to adopting FAIR standards.

"The industry is making strides in lots of areas, especially in terms of single application or single tier of users through applications such as an ELN or a LIMS system that might be cloud-enabled," notes Tarselli. "They are making strides and because of that work they have better visibility than they had before. That said, there remains a large number of what I would call 'frustrated scientists and engineers' who cannot gain access to the core data they need to answer a query in a timely fashion."

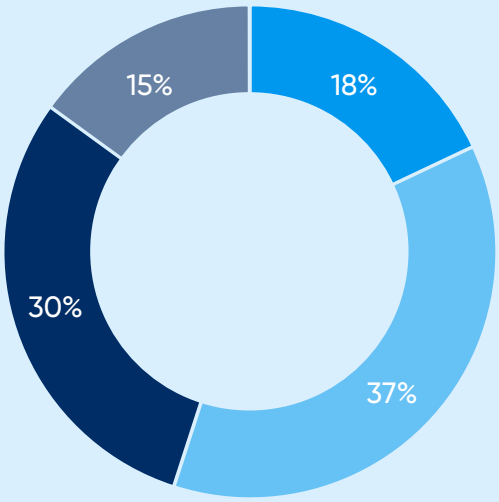
Albrecht says that he expected the percentage to be even lower. "There is a huge amount of 'dark data matter' and no one knows how much there is or where it is," he remarks. "No one knows how much data and 'knowledge' is hidden in PowerPoint presentations, Excel files or images, for example."

His advice is to first acknowledge that much of this information is hidden, but also to accept how valuable it can be. Once the strategy for each kind of data and the scope of the task has been defined, he says companies should "provide resources for 'data FAIR-ification' projects."

Do not keep the FAIR data only in presentations and talks, but implement the principles," he asserts.

FIGURE 4

What is the top driver behind your initiative to replatform scientific data on the cloud?

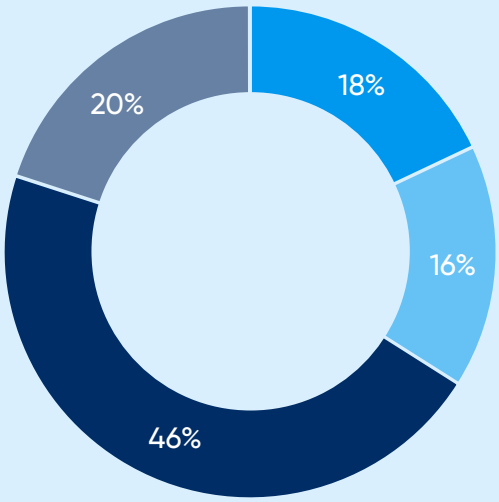


- Strategic importance of AI/ML applications
- Cloud benefits such as reduced costs, flexibility, elasticity, reliability, security, etc.
- Improved operational efficiency to accelerate innovation
- Breaking down technical and organizational silos

Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

FIGURE 5

Which functional group is the primary driver behind your initiative to replatform scientific data on the cloud?



- IT
- Science
- Data science and/or informatics
- Operations

Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

While accepting that biopharma lags behind other industries in its adoption of cloud-based solutions, TetraScience's Tarselli says that he believes the entire industry is moving in a progressive direction.

"We believe it's going to probably be a five-to-10-year paradigm shift to get it there," he predicts. "But the fact that more than 80 percent of those surveyed are either in the middle of their journeys or haven't started does mean we have to help our customers along."

Among those who have completed or are on the journey to replatforming their scientific data on the cloud, the survey revealed a mix of top drivers behind taking the digital initiative (see Figure 4), with more than a third of respondents opting for the benefits that come with cloud platforms such as costs, flexibility, elasticity, reliability, security (37 percent). This was followed by 30 percent who saw the key value being improving operational efficiency to accelerate innovation, with 18 percent opting for the strategic importance of AI/ML applications and the remaining 15 percent selecting breaking down technical and organizational silos.

"Automated analyzes are the perfect solution when you need real-time analytics to recommend a new product based on a click," states Pfizer's Henstock. "They are also ideal when the data streams are standardized and repeatable. However, the science often does not lend itself to these methods since new experimental approaches and custom analysis techniques often require new strategies."

Implementing digital solutions

The biggest challenges in implementing digital data initiatives in pharma are varied, according to our survey respondents, with the leading challenge reported to be change management and user adoption (32 percent). An unclear path to successful implementation and a lack of support - executive support, strategic priority, staff or budget resources - were both cited by a quarter of respondents as key challenges.

This close result between all of this question's options suggests that pharma organizations and their leaders struggle to change mind-sets in respect to digital adoption.

"Changing to cloud computing means reorganizing IT departments, systems development and IT systems to use the best from cloud services and the new ways of providing and using IT resources," says Roche's Albrecht. This is no small task and Albrecht suggests first using cloud computing on smaller projects that would significantly benefit from using cloud computing to prove the point.

TetraScience's Tarselli notes that it is important to keep in mind that there is not one 'magical' pharma customer that represents all use cases. "Pharma is a diverse landscape," he explains. "A screening chemist does not have the same information, security and access, or change journey as a clinical biologist or pilot plant manufacturer. They all may work on the same floor of a pharma, but they have very different data systems with which they interact, and very different values of each piece of data they produce."

"Digital labs offer a chance to reinvent and redesign platforms to work together with the latest technologies, standardized metadata and ontologies."

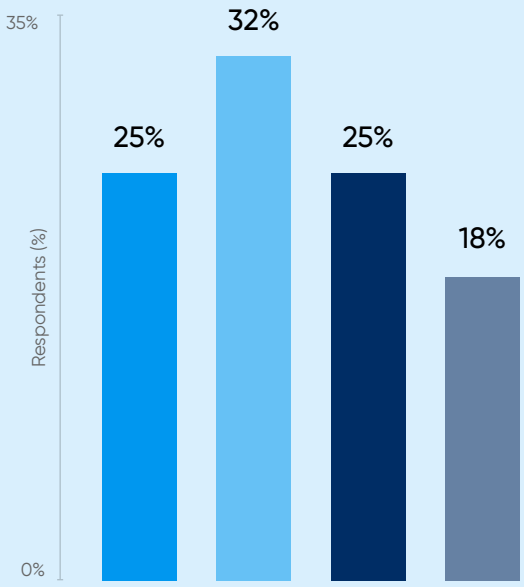
Peter Henstock

ML and AI Technical Lead at Pfizer

FIGURE 6

What is the biggest challenge in implementing digital data initiatives at your organization?

- Lack of executive support, strategic priority, staff or budget resources
- Change management and user adoption
- Unclear path to successful implementation
- Limitations in skillsets within the organization

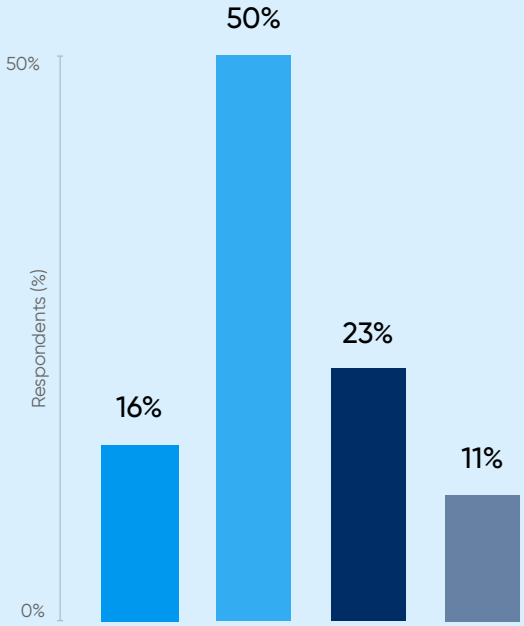


Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

FIGURE 7

Has your organization implemented a fully digital lab, where all appropriate scientific data is captured and harmonized in the cloud, and manual lab operations are eliminated for routine workflows and experiments?

- Yes. Our digital lab is implemented and working efficiently.
- Somewhat. My organization is in the beginning-to-middle stages of implementation.
- Not yet. My organization is planning for our digital lab.
- No. We have yet to start a digital lab initiative, or it is not part of our business strategy.



Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

While half of the respondents say their organizations are in the beginning-to-middle stages of replatforming their scientific data on the cloud, just 16 percent have completed the journey (see Figure 7). Surprisingly, more than a third of respondents (34 percent) have not yet begun implementing a move to the cloud, as they are either still in the planning stages or have no active digital lab initiative.

TetraScience's Tarselli adds that the 50 percent of organizations that are in the beginning-to-middle stages of implementation should be celebrated, noting that those already on the journey to having a fully digital lab, where all data will be captured and harmonized in the cloud, places biopharma in "a positive place."

"There are more instruments that are sending their data directly to the cloud now," says Pfizer's Henstock. "In addition, digital labs offer a chance to reinvent and redesign platforms to work together with the latest technologies, standardized metadata and ontologies."

Ending manual data extraction and transformation with automation

While the future of cloud looks bright, many digital lab projects are continuing to encounter significant friction on the ground. Building and maintaining 'point-to-point' integrations to extract raw data from instruments, control software, file shares and other silos, parse it out of dozens of different file formats, transform and enrich it, and then make it accessible for analytics or delivery to electronic lab notebooks, lab information management systems and other data targets.

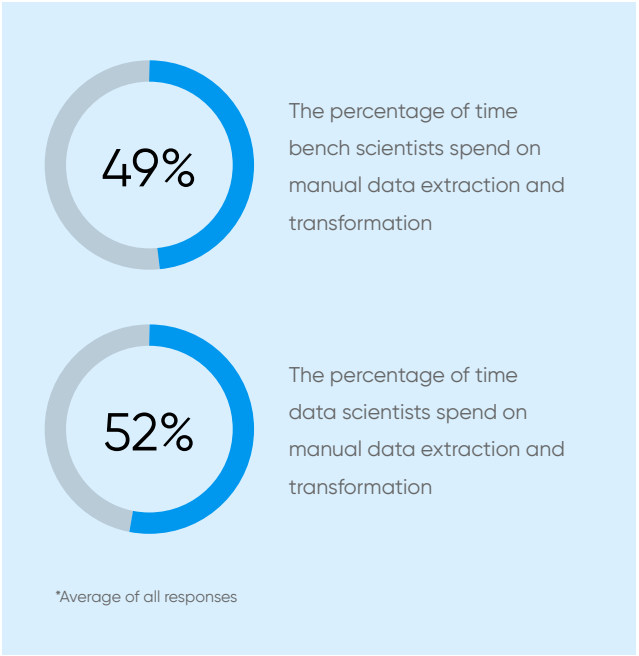
Organizations engineering their own data management solutions often find they cannot count on much guidance from instrument and software makers. Extracting data and building parsers for complex and proprietary files requires daunting forensics and experimentation. Engineering dataflows to serve complex scientific use-cases requires an intimate understanding of the science in context. Building reliable, secure, scalable and compliant pipelines requires a wealth of specialized knowledge, and the cost of operating and maintaining such one-off software toolchains is high.

It is, therefore, little wonder that more than 87 percent of survey respondents stated that the burden of building and maintaining point-to-point scientific data integrations ranged from 'significant' to 'extreme' (see Figure 8).

In addition to variable time (see Figure 10) spent locating and retrieving data, an average of 49 percent of bench scientists' time is spent on manual data wrangling, increasing slightly to 52 percent for data scientists. For purposes of the survey, 'manual data wrangling' denotes low-value, labor-intensive toil expended finding, extracting, transforming and validating data, often performed by expert individuals (e.g., scientists) whose time is extremely valuable. This data toil is systematic and repetitive enough to justify eliminating with an R&D data cloud: an organization-wide, centrally-maintained, cloud-native automation framework. Manual data wrangling does not include the normal and expected labor required, for example, to curate datasets for ML – but does include many preliminary steps essential to preparing raw data for such curation.

With that in mind, these are startling figures; particularly when taken alongside (see Figure 11) the fact that 45 percent of survey respondents say that "less than half of their organization's data is sufficiently prepared and available for analysis."

This all suggests two things: first, organizations may be losing staggering sums to low-value manual data manipulation (losses calculated in terms of fully-burdened scientist and data scientist time, opportunity costs to more profitable labor, risks to accuracy and compliance imposed when data is manually manipulated, and delayed arrival of therapeutics to market). Second, the large volume of time lost also suggests that "low-hanging fruit" likely exists for quick remediation with technology.

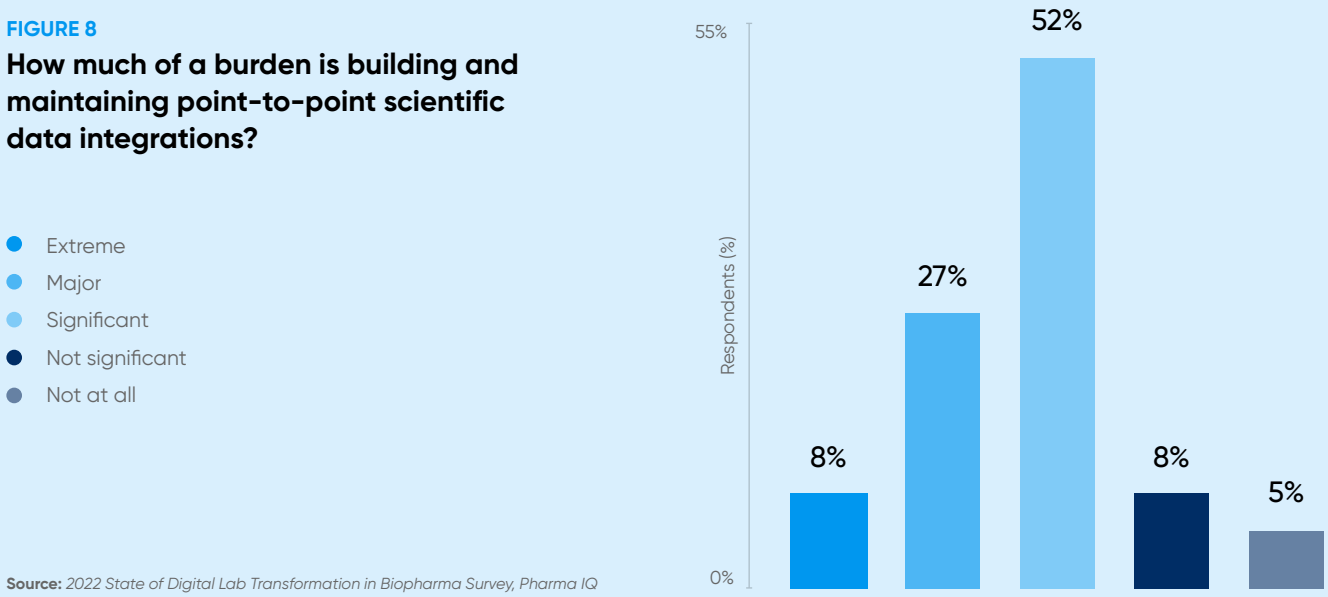


Certainly, organizations able to automate or better control their data will stand to benefit over those who are stuck working manually. As Roche's Albrecht remarks: "These companies (that automate data management) have higher throughput in their analysis and also can obtain more value from the generated data."

Ending manual data extraction and transformation with automation

FIGURE 8

How much of a burden is building and maintaining point-to-point scientific data integrations?



Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

"It is up to the IT, data science, data engineering and lab automation teams to show executives real examples of the benefits of implementing ML/AI."

Felipe Albrecht

Senior Scientist at Discovery Informatics at Roche

TetraScience's Tarselli notes that most bench scientists do not work 40-hour weeks. "Instead, they work 50–60-hour weeks because they're committed to their craft and they're passionate about it," he says. "This means they're spending 20–30 hours a week going between spreadsheets and PowerPoints, trying to get their data firmed up for the following week's meeting."

"The risk faced here is that these folks, with their core expertise in deep interrogation of scientific problems, are being asked instead to move data around," he adds. "That's something a machine has been doing and will continue doing better than a human."

Pfizer's Henstock remarks: "Biopharma companies have been innovating for decades to deliver new treatments to patients and new scientific technologies are explored daily to improve outcomes. But pharma has been conservative in its investment for a few reasons.

"Unlike Netflix or Amazon where AI/ML has a unifying billion-dollar focus on one problem, biopharma have hundreds or thousands of niche problems."

"As an expansive and evolving suite of solutions, the technical AI/ML strategy must tie closely with the business strategy to efficiently solve related classes of problems," he advises. "Each solution requires custom training and tuning for the specific problems it tackles."

"AI/ML techniques are up for the challenge and continue to broadly gain support across the industry from all levels," adds Henstock, who coordinates a bimonthly internal talk series at Pfizer highlighting how AI/ML is successfully solving industry problems.

Ending manual data extraction and transformation with automation

The biggest inhibitors to using AI/ML for scientific research (see Figure 9), however, are skill or knowledge gaps in the organization (32 percent) followed by lack of funding, executive support or strategic priority (24 percent).

"AI and ML are not new disciplines," Tarselli notes. "There is a 40-to-50-year rich history of teaching computers through enhanced vision and text recognition, and many ML data curation and indexing tools are now coming into some level of maturity. But in pharma we just aren't quite there yet in a meaningful way, because we don't have organized well-curated, well-harmonized data with consistent metadata."

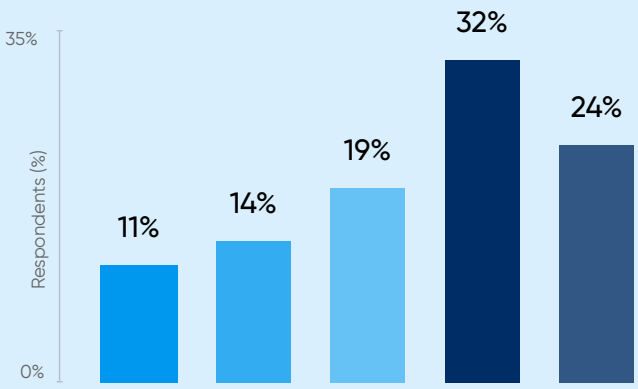
The benefits of adopting these processes will put you ahead of many of your competitors, with just 14 percent of organizations surveyed revealing they can access experimental results for subsequent analysis, visualization and data science applications within one hour (see Figure 10).

Roche's Albrecht advises that AI/ML should not be the goal of pharma companies but instead seen as a tool with which to develop better medicines.

"It is up to the IT, data science, data engineering and lab automation teams to show executives real examples of the benefits of implementing ML/AI, focusing on the return of the investment, rather than on buzzwords or fear of missing out," he remarks.

FIGURE 9

What is the biggest inhibitor to using AI/ML for scientific research?

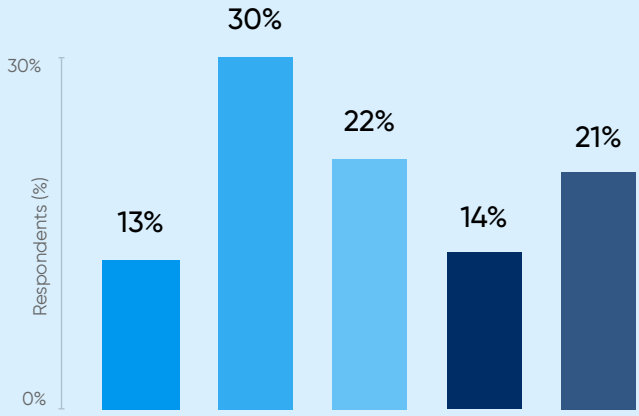


- Insufficient volume of data
- Poor suitability of data for AI/ML
- Lack of use cases for AI/ML applications in our business
- Skill or knowledge gaps in the organization
- Lack of funding, executive support or strategic priority

Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

FIGURE 10

How quickly can your organization access experimental results that are prepared for subsequent analysis, visualization and data science applications?



- <1 Hour
- 1–12 Hours
- 12–24 hours
- >24 hours
- Not sure

Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

Conclusion

The results of the *2022 State of Digital Lab Transformation in Biopharma Survey* demonstrate that pharma is on the right path toward cloud adoption but, as Figure 10 clearly demonstrates, it continues to be far too slow in adopting data-driven solutions that increase efficiency and embolden decisions.

Pharma may have demonstrated to the world its ability to react and work together during the Covid-19 global pandemic, but in terms of technology, it remains a laggard. The *2022 State of Digital Lab Transformation in Biopharma Survey* demonstrates a strong appetite within pharma for cloud adoption and an opportunity to do more than simply pay lip service to FAIR principles.

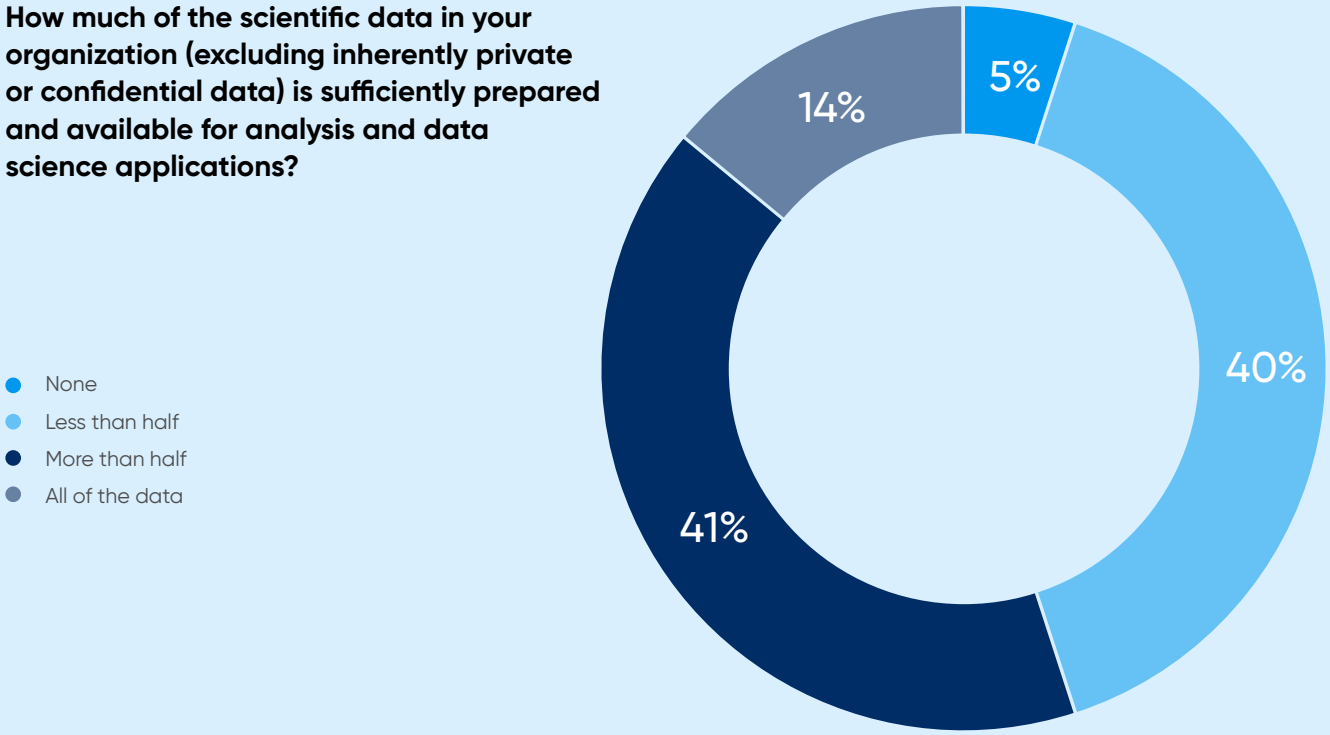
Executive support and strategic prioritization toward replatforming to the cloud offers scientists more time to do what they were employed to do, while adopting progressive approaches to AI/ML solutions enables organizations to maintain forward-looking approaches to business decisions via relevant data.

By investing in cloud computing to resolve some of their enormous data challenges, life sciences firms will take a leap of faith that should see them remain competitive, relevant and able to research and develop the next generation of medical discoveries.

Pharma's unique challenges does not mean it should look within itself, but instead look out at other industries to witness the success stories emanating from cloud-fueled solutions. Ultimately, in pharma, patients will benefit, while executives will see reductions in expenditure and increases in profitability.

FIGURE 11

How much of the scientific data in your organization (excluding inherently private or confidential data) is sufficiently prepared and available for analysis and data science applications?



Source: *2022 State of Digital Lab Transformation in Biopharma Survey, Pharma IQ*

Please note: All comments made by the contributors of this report are solely the views of the individuals without any relation to their employers, institutions or business partners.

About TetraScience

TetraScience is the R&D Data Cloud company with a mission to transform life sciences R&D, accelerate discovery, and improve and extend human life. While the company was founded in 2014, we began our R&D Data Cloud journey in 2019 with the origin of the Tetra Data Platform (TDP).

The Tetra R&D Data Cloud provides life sciences companies with the flexibility, scalability, and data-centric capabilities to enable easy access to centralized, harmonized, and actionable scientific data and is actively deployed across enterprise pharma and biotech organizations.

tetrascience

visit www.tetrascience.com

About Pharma IQ

Pharma IQ is an online resource for life sciences and biopharma supply chain professionals providing value-rich content and insight through industry reports, customer trends, best practice, latest industry news, interviews with pharma thought-leaders, online events and so much more! Our focus is on the content that matters, providing our members with a primary resource to turn to and giving our partners a clear voice within a crowded spectrum.

By joining our growing network you will be able to connect, learn and interactive with other likeminded industry professionals and keep up-to-date with industry inside advice to help you achieve your business goals.

Pharma iQ

If you are interested in becoming a *Pharma IQ* contributor, please get in touch with ***Pharma IQ***

Become a member today to receive exclusive access to our content and regular industry updates.