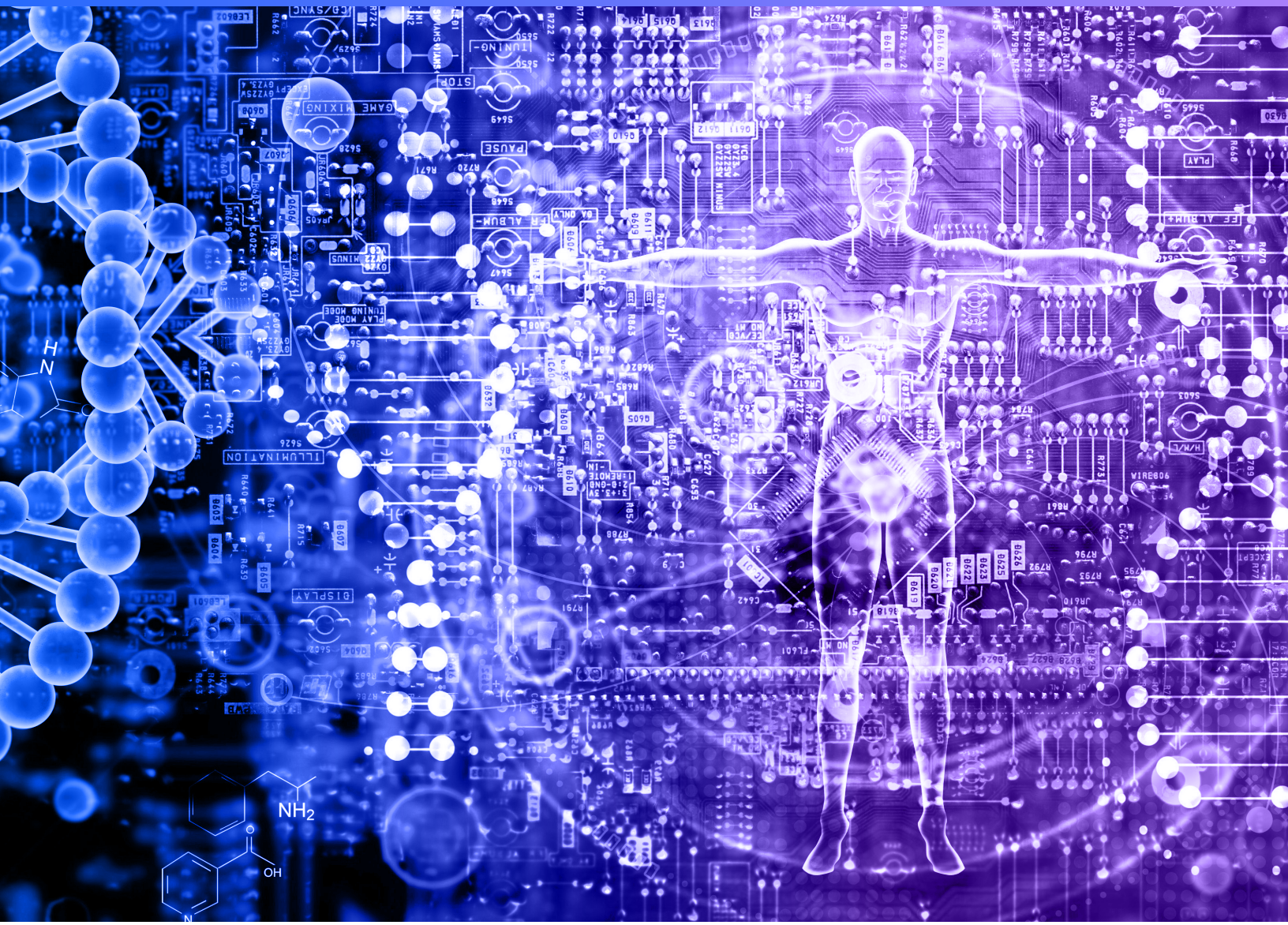
WHITEPAPER

Reimagining the Future of Biomedical Research

Enabling Access to Real-World Data with Secure, Cloud-Based Collaboration



PRESENTED BY: Flywheel

PUBLISHED BY: FIERCE Biotech

© 2022 Questex Content Marketing. All rights reserved. All registered trademarks are property of their respective owners.

It's no secret that drug development is an expensive and time-consuming process. One paper from the Tufts Center for the Study of Drug Development has pegged the cost of bringing a drug to market (including post-approval research and development) at $2.87 billion. A second study from the Tufts Center noted the timeline for new drug development ranges from 12.8 years for the average drug to 17.2 years for ultra-orphan drugs that only affect several hundred patients.

Data acquisition and curation are some of the biggest obstacles that life science organizations (LSOs) face in their efforts to accelerate the drug discovery process. Much of this data – whether it's the results of previous clinical trials, imaging data, or non-clinical data sets – sits outside the organization's digital walls, unable to be accessed by on-premises applications. External collaborators who are researching the same rare diseases or breakthrough treatments are similarly difficult to locate.

Due to these challenges, LSOs spend considerable time and money finding data and preparing it for analysis, which contributes to the long timeline and high cost of drug development. However, emerging platforms that automate data curation and improve access to data are poised to minimize the manual effort associated with data acquisition and curation – while at the same time enabling connections with stakeholders no matter where they are based.

Why Accelerating Drug Discovery Is an Uphill Battle

The early stages of drug discovery and development focus heavily on research and testing. Researchers may need to determine which therapeutic compounds out of thousands of potential candidates should be studied further, which indications of a specific cancer or chronic condition may respond best to a particular treatment, or how different segments of the population may respond to or be affected by a given therapy.

Data plays a key role in the research process. The ability to access other research into a therapy or condition, and to collaborate with the researchers who have done that work, can accelerate the drug discovery process. Instead of running an experiment or trial to obtain data for further analysis, an organization can leverage data that others have previously gathered or are even currently working with. In addition, being able to reuse the same data that

Data acquisition and curation are some of the biggest obstacles that life science organizations (LSOs) face in their efforts to accelerate the drug discovery process.

an organization collected for a previous research project can save time and money, while also offering a layer of historical insight for ongoing research. Increasing the volume, velocity, and variety of data paints a more complete picture of the therapy, condition, or population being studied, and it reduces the likelihood of bias within the data set.

Unfortunately, LSOs typically face an uphill battle in their efforts to accelerate drug discovery. In some instances, they may have their own data that they'd like to reuse, but are still faced with a huge task to capture and curate it. In other cases, they do not own the data they need; instead, it is sitting in academic and clinical institutions that have done this research or have access to actual patient data. This poses two important challenges to LSOs: One is trying to access the right data, and the other is collaborating with the research teams who produced that data.

Working With Data—Especially Imaging Data—Is Difficult

As discussed, data acquisition and curation are expensive and time-consuming processes. Data and associated metadata from different sources are often formatted very differently. MRI scans, for example, are stored in complex file formats that include the raw images as well as expansive metadata that detail how the images were acquired. Working with complex data sources like these requires format-specific mechanisms for exchange, processing, and annotation.

Most LSOs perform data curation and analysis manually. This process risks creating inconsistencies across the data set, as each person who views the data may spot different patterns. It is also error-prone, as even the most thorough examiners will miss important information, occasionally skip steps, and may not look at every data set the exact same way. When looking at a set of MRI scans for a single patient, a tumor may look different depending on which type of machine conducted the scan, or if the patient was positioned at a different angle as the scan was taken. Finally, the output of an analysis becomes outdated as soon as new data becomes available – and requires the entire process to start all over again.

There are five other key reasons that working with data can be difficult for drug discovery.

This poses two important challenges to LSOs: One is trying to access the right data, and the other is collaborating with the research teams who produced that data.

Size of data sets. Biomedical data sets are notoriously large. For example, the Human Connectome Project that maps the brain is approximately 1 Petabyte of data. (This is the equivalent of 11,000 high-definition movies.) This poses hurdles for on-premises data storage and computing.

Diversity of data sets. For an unbiased analysis, data sets must be diverse in addition to large. Artificial intelligence (AI) applications are increasingly used in biomedical research and are particularly susceptible to insufficient data diversity. AI applications trained on small single-source data sets generalize poorly and risk producing misleading results when presented with edge cases and other anomalies not present in the original training data.

Deidentification of data sets. Under HIPAA, data sets must be stripped of 18 specific identifiers, ranging from names and geographic data to Social Security numbers and account numbers, before they can be used for analysis and shared with third parties. This necessary step can be costly and time-consuming.

Availability of data sets. Because curating and analyzing data requires a significant investment in resources, much of this work tends to be done at large academic medical centers (AMCs). This has led to the creation of so-called "data deserts" – 71% of the algorithms used in pathways approved by the U.S. Food and Drug Administration (FDA) come from just three states (California, Massachusetts, and New York), according to Stanford University researchers.

Complexity of computing. Perhaps the single biggest issue is that image processing is complex, multi-step, and computationally intensive. Interlinking algorithms for medical imaging analysis create "pipelines" where the output of one element is the input for the next, requiring a computation infrastructure that can support this large-scale analysis for hundreds and thousands of data sets. It also needs to be repeatable and reproducible.

Challenges in Fostering Collaboration

Access to data is one significant challenge for life science organizations – and efficiently collaborating with other researchers is another. Researchers need to cast a wide net to find collaborators who might be based in any number of research settings, whether it's a hospital, an academic research center, a clinical trial partner, a nonprofit organization, or a corporate sponsor. In some cases, finding collaborators can take years.

Researchers need to cast a wide net to find collaborators who might be based in any number of research settings, whether it's a hospital, an academic research center, a clinical trial partner, a nonprofit organization, or a corporate sponsor.

Once collaborators have been identified, relationships have been established, and data sets have been acquired, curation and analysis can also pose problems. With collaborators based in different locations, using different analytics systems and working in different time zones, it can be difficult to audit who has been using data at what time – and, in a step that's critical for quality control, audit who has updated or changed the data set at what time.

The processes of identifying research collaborators and acquiring and curating data sets means that much of the research work associated with drug discovery and development has very little to do with research. Instead, researchers spend most of their time finding data and preparing it for analysis. These efforts can significantly lengthen the drug discovery timeline.

How a Research Data Platform Streamlines the Process

Fortunately, time-consuming and error-prone manual processes no longer need to be the status quo for LSOs. A research data platform offers the potential to automate and streamline data analysis and collaboration for drug discovery and development.

Flywheel is a cloud-scale research data platform designed to complete four primary tasks: Ingest multiple types of medical data, organize and annotate data to common standards, make data available for scientific analysis, and enable collaboration with external stakeholders through access-controlled projects. This benefits LSOs in important ways.

Availability. Institutions need the ability to host data from previous research studies, clinical trials, and external collaborators on one single platform. This allows the data to contribute more value to the research process than it would by otherwise remaining in an archive or internal silo – but it also allows institutions to retain control over who can access and work with their data.

Automation. Platforms like Flywheel curate and normalize data automatically, streamlining the most time-consuming part of the process of preparing raw data for analysis. Platforms also offer options such as organizing data by groups or projects; making it easily searchable; conducting quality checks that ensure consistency across the data set; pulling data from sources that are not always considered, such as the header and footer of an MRI scan; and flagging anomalies within data sets. Automation takes the

The processes of identifying research collaborators and acquiring and curating data sets means that much of the research work associated with drug discovery and development has very little to do with research.

heavy lifting off the shoulders of researchers so they can focus on the complex problems they are trying to solve.

Security. Keeping patient data secure is a paramount concern for researchers. Data platforms can address this need with a range of de-identification capabilities that are applied when data is uploaded or captured, which ensures compliance with regulations such as HIPAA and GDPR. These features target typical sources of protected health information as well as hard-to-find sources of PHI such as patient names included in radiology imaging metadata. Flywheel also leverages role-based permissions that give entities hosting data control over which collaborators can access and modify data, minimizing the chances that data will fall into the wrong hands.

Collaboration. A research data platform enables LSOs to reach beyond the usual suspects to collaborate with others who are working to solve the same problem – whether they are part of a team at a large AMC or are working as a single principal investigator at a small clinical site, and whether they are in the same city or halfway across the globe. With Flywheel, organizations can search for data sets of interest based on a variety of criteria. This gives LSOs a place to search for available data and potential collaborators, allowing them to build more robust algorithms and more relationships in less time.

Containerization. As more collaborators work on a project, the likelihood increases that data will be imported to machines that run different operating systems or different types of analytics software. This could pose problems for compatibility and require end users to download new software or set up new infrastructure – adding cost, time, and complexity to a project. A sophisticated data platform addresses this by leveraging containerization, which allows the analytics application to run in its own isolated environment, or container. Along with eliminating the need for new hardware or software, this has the added benefit of allowing multiple researchers to run queries on the same data set at the same time.

Tracking and reporting. Opening a data set to dozens (or hundreds) of eager collaborators can create an almost-immediate version control headache. Working within a data platform like Flywheel makes it possible to track who makes changes when – both to the data sets and to the algorithms that are used to analyze them. This has benefits for regulatory compliance, as it provides an audit trail of changes that can be matched against user permissions

Cloud-based platforms enable LSOs to reach beyond the usual suspects to find others who are working to solve the same problem – whether they are part of a team at a large AMC or are working as a single principal investigator at a small clinical site, and whether they are in the same city or halfway across the globe.

to demonstrate that only authorized users have accessed and changed the data. This also benefits the research process itself, as an audit trail ensures that changes made by one user can be reproduced by another – a critical step in advancing clinical trials or obtaining FDA approval.

Savings. LSOs that make the transition from on-premises to cloud-based hosting of research data can realize significant cost savings. Hosting data in the cloud allows for the flexibility of purchasing storage as needed, instead of investing in (and setting up and maintaining) on-premises hardware such as mainframes. Meanwhile, the availability of elastic computing lets organizations spin up (and pay for) nodes only when they are needed for data analysis, which allows for more flexible scaling of resources. LSOs also realize indirect savings because research projects are not put on hold when labs are forced to close, as was the case during COVID-19.

Why Data-Driven Discovery Takes Research to the Next Level

Traditional research methods force life science organizations to spend considerable resources searching for external data sources, acquiring data, and curating data before it can be analyzed. Not only does this contribute to the long timeline and significant expense associated with drug discovery and development; it also makes it difficult for researchers to work with external collaborators, as they don't have access to the same data and analytics resources.

Flywheel makes high-quality data ready at scale for machine learning. Data sets are longitudinal, incorporating data from a range of provider, pharmaceutical, and non-clinical sources that LSOs have traditionally not been able to access without paying significant acquisition costs. These larger and more comprehensive data sets are also less likely to be biased, giving researchers the peace of mind to know that the algorithms they develop to analyze the data can effectively be used for subsequent analysis on different data sets.

Data that is ready for machine learning enables a data-driven approach that allows LSOs to push innovation further and focus on increasingly complex problems. With less time devoted to preparing data for analysis, research teams can hit the ground running to develop new therapies faster, to identify

Opening a data set to dozens (or hundreds) of eager collaborators can create an almost-immediate version control headache. Working within a data platform like Flywheel makes it possible to track who makes changes when – both to the data sets and to the algorithms that are used to analyze them.

new indications to target, and to weed out options with harmful side effects or that do not produce the expected results for certain population segments. Researchers can create powerful algorithms to develop predictive models that can aid in treatment decisions or even diagnoses.

Working with a research data platform also significantly expands the universe of potential research collaborators. With more diverse data, skill sets, and points of view, teams can uncover insights and deliver more effective therapies to more patients in less time.

Making the transition to a cloud-scale research data platform gives LSOs a wider view of data that has been carefully curated and is ready for analysis. The Flywheel platform enables LSOs to quickly and securely discover data across a global research network and identify potential collaborators for projects – and faster discovery leads to drug development that gets the right therapies to the right patients faster, which improves outcomes and lowers care costs.

Set up a meeting with Flywheel to see why top researchers around the world use this research data platform every day.

Flywheel

Flywheel offers comprehensive solutions for the life sciences, clinical research, and academic research industries to accelerate collaboration, enable machine learning, and streamline the massive task of data aggregation, curation and management. By leveraging cloud scalability and automating research workflows, we help organizations scale research data and analysis, improve scientific collaboration and accelerate discoveries.