Flywheel

Medical Imaging in Life Sciences R&D: *How to Remove Barriers, Accelerate Timelines and Improve Outcomes*

Instant access to high-quality, well-curated, standardized imaging data is critical

> Life sciences organizations are dramatically redefining research and development processes in the hopes of accelerating innovation and reducing the costs of drug development.

Computational techniques such as artificial intelligence and machine learning are accelerating disease identification, personalized treatments and drug discovery. With major investments in digital transformation, these organizations aim to bring innovative therapies to market faster.

Despite these promising trends, many organizations are struggling to realize the potential in the data they hold. Although an abundance of structured and unstructured data exists to support research objectives, teams are facing challenges in their attempts to fully and efficiently leverage these assets. In data science, the “80/20 rule” is well-known, referring to the fact that researchers typically spend 80% of their time finding, curating and organizing data and only 20% of their time on actual research.

Some common challenges for researchers relate to standardizing data access, curating large volumes of data and incorporating data sets into analytical workflows. Compounding these problems is the fact that many organizations are still relying on inflexible infrastructures that struggle to handle rich, complex data at scale.

Pharmaceutical companies have implemented different solutions, many of them homegrown, to try to improve data governance and facilitate research with complex objects such as medical images (MRI, PET, CT, etc.). But the complexity of large data types such as imaging data requires a sophisticated system that indexes metadata, automates curation and streamlines computation to enable data scientists to fully practice their craft.

To meet their goals, research teams need data infrastructures that help ensure compliance, support complex workflows and scale for their needs.

An efficient solution should:

Aggregate data into one centralized location where research teams can securely access it with permission-based controls in a matter of seconds, versus hours or days.

Allow data scientists to spend the bulk of their time on their actual research instead of data curation.

Facilitate the creation of high-quality data sets for AI training, with the ultimate goals of accelerating drug development, achieving financial projections and improving R&D outcomes.

The Challenges of *Leveraging Images for Research*

A variety of data modalities can be leveraged to develop AI in health care, with medical imaging becoming a highly active and promising area in recent years. Technological growth and computational power offer organizations an opportunity to store and access tremendous volumes and varieties of images. But leveraging this potential is at best an arduous process. Data at pharmaceutical companies is often siloed, harbored in legacy systems or offsite at partnering vendors such as clinical research organizations, and it can exist in an array of formats with different storage conventions.

Curating and annotating unstructured data, as well as accessing the computational power to work efficiently, can lead to barriers.[1] The 80/20 rule strikes at organizations both large and small, meaning data scientists spend a lot of their time manually finding, cleansing and organizing data, leaving little room for them to perform the task for which they were hired.

Why Getting it Right *Matters*

These inefficiencies can have ripple effects with significant consequences. If data scientists spend the bulk of their time on data housekeeping tasks before active research commences, then the development cycle slows down. What should be a go-to asset — a vast library of images rich with data to help inform decision-making — can end up hampering progress instead of driving it. Clinical trials can be delayed, and time-to-market and cost both increase as the project drags on.

What to Look for in a *Data Management Solution*

Keep these considerations in mind when evaluating potential data management platforms:

1 Avoid rigid solutions.

Tools built for specific purposes or workflows may not be flexible enough to accommodate the range of researchers' needs. Consider carefully how your data is currently handled, what immediate problems you wish to solve and what long-term goals you want to achieve. Based on that information, a partner should be able to recommend workflows and automations that can meet the full breadth of your enterprise's needs.

2 Confirm that the solution can integrate with other key tools.

Imaging data is a rich jumping-off point for many types of research, so a platform for handling it should facilitate further integrations and collaboration. For example, central nervous system researchers in pharmaceutical companies often use FreeSurfer, an open-source neuroimaging toolkit for processing, analyzing and visualizing brain images. Pharma researchers may use FreeSurfer to perform automated longitudinal analysis over the course of a trial to show how treatment interventions change brain structure. Ask any potential data management partners how "open" their platform is to any tools you rely on now. Then think ahead: What other tools might help your team if you knew they could integrate smoothly with your data platform?

3 Look to the FAIR Guiding Principles.[2]

In recent years, the FAIR principles have become adopted in numerous life sciences organizations to facilitate sustainable, innovative R&D. Under these principles, data should be:

- **Findable,** ideally in a central hub where all data resides.
- **Accessible,** via controls that enable data scientists and others to access appropriate data on demand when it is most needed.
- **Interoperable,** through a standardized process of naming, creating and understanding the context of images so they can be shared across multiple systems and platforms.
- **Reusable,** for clinical trials and real-world data analyses, among other efforts.

Best Practices in *Data Management for Pharma*

1. **Make data "FAIRification" part of the company culture.** Without this mindset, divisions may go back to siloed and disorganized processes. A commitment to FAIR, combined with a solution that adheres to it, increases the odds of success.
2. **Enforce data standardization.** Data standardization and enforcement is hard and tedious work. Look for solutions that automate standardization tasks to help the process work at scale.
3. **Encourage collaboration.** Lack of collaboration is often due to technological barriers, but modern data platforms enable data to be shared widely, with versioning tools that let researchers utilize an image and even add custom metadata, without altering the original.

Success *in action*

Pharmaceutical organizations are under intense pressure to accelerate drug discovery. But to achieve this, they need to leverage all the data at their disposal, including complex imaging data, for the machine learning processes that can improve their research timelines. To enable efficient machine learning, data management and curation processes must be automated and scalable.

Flywheel is a biomedical research data platform that was created to help solve challenges in modern pharma R&D. In one recent use case, the platform helped speed up much-needed COVID-19 research at the University of Wisconsin-Madison. Researchers there were seeking to train an AI algorithm to diagnose COVID-19 via X-ray. By tapping into the Flywheel platform, researchers were able to:

- Quickly access, curate, normalize and maintain data from multiple sources on more than 3,000 patients and 6,000 X-rays.
- Leverage sophisticated computing tools and infrastructure to facilitate rapid iteration.
- Add custom metadata to facilitate the study.
- Access unique manifests for unique training batches.

> The Flywheel solution turned the 80/20 rule on its head, allowing researchers to efficiently develop an AI algorithm that significantly outperformed human readers.[3]

The University of Wisconsin-Madison experience is just one example of how modern data management is helping organizations make vast amounts of data AI-ready and leverage it for important advances. To achieve success in this new landscape, life sciences organizations must give their teams the tools to harness the value of every piece of data available to them.

REFERENCES

1. Willemink M, Koszek W, Hardell C, et al. Preparing medical imaging data for machine learning. Radiology 2020; 295(1): 4–15. doi: 10.1148/radiol.2020192224

2. Wilkinson M, Dumontier M, Aalbersberg U, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

3. Zhang R, Tie X, Qi Z, et al. Diagnosis of coronavirus disease 2019 pneumonia by using chest radiography: Value of artificial intelligence. Radiology 2021; 298: E88-E97. doi: https://doi.org/10.1148/radiol.2020202944

Flywheel is the revolutionary research data management platform powering healthcare innovation by accelerating collaboration, enabling machine learning, and streamlining the massive task of data aggregation, curation and management. By leveraging cloud scalability and automating research workflows, Flywheel helps organizations scale research data and analysis, improve scientific collaboration and accelerate discoveries. Flywheel offers comprehensive solutions for life sciences, pharmaceutical, biotech, academic, and clinical research. Flywheel is headquartered in Minneapolis and has offices in the Bay Area, St. Louis, and Budapest. For more information on our mission and products, visit www.flywheel.io or follow us on Twitter and LinkedIn.

Learn More

Flywheel