Revolutionizing Medical Research: *Flywheel as the Imaging Solution with CAVATICA* in a Multimodal Data Ecosystem

In the evolving medical research landscape, integrating diverse data types—notably imaging and genomic data—is vital for advancing predictive analytics and precision medicine. However, the potential of a multimodal approach faces a significant challenge: the disparate and domain-specific nature of current data management and analytics platforms.

Today, medical data is typically harnessed within three independent silos:

Imaging data: Provides crucial visual insights but is often separated from other patient data, stored in systems designed solely for image archiving and retrieval.

Genomic data: Holds the key to understanding disease at a molecular level, yet it's managed and analyzed through highly specialized, domain-specific platforms.

Clinical data: Rich in patient history and treatment outcomes, this data is typically locked away in electronic health records, formatted in ways not conducive to integration with other data types.

Each data type resides in various platforms with incompatible formats and standards. This fragmentation hampers the ability to derive meaningful insights. It's like trying to solve a puzzle when the pieces are scattered across different rooms, each requiring a unique key to enter.

The challenge is made more complex with the need for de-identification and data privacy. Researchers must navigate the tricky terrain of maintaining patient confidentiality while ensuring that the corresponding data across different modalities can still be linked to the same individual for comprehensive analysis.

Scalability presents another problem. Artificial intelligence (AI) and machine learning (ML) research require large datasets for accurate model training and validation. However, the labor-intensive process of collating, harmonizing, and analyzing this multimodal data is often beyond the scope and capability of individual research teams or small institutions.

Put simply, the lack of a unified, interoperable framework that can seamlessly bring together varied data types into a cohesive, anonymized,

yet comprehensive dataset doesn't exist. Such a framework would facilitate the integration of these diverse data streams and enable scalable analysis and model building in a privacy-compliant manner. Ultimately, its goal would be to equip researchers with the ability to harness this integrated data for breakthrough discoveries and advancements in patient care.

A Customized Ecosystem for Integrating Genomics and Imaging Data

Flywheel was recently invited to be part of a multi-institutional partnership sponsored by the National Institutes of Health for the INCLUDE (INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndromE) project, an initiative to study co-occurring conditions in individuals with Down Syndrome. The project, led by Adam Resnick, PhD and supported through his colleagues Ariana Familiar, PhD and Allison Heath, PhD, was to integrate Flywheel as the platform to manage medical imaging data as an augmentation to their end-to-end genomics data management workflow including Velsera's CAVATICA platform.

CAVATICA is a data analysis and sharing platform designed to accelerate discovery in a scalable, cloud-based compute environment where data, results, and workflows are shared among the world's research community. Developed by Seven Bridges and funded in-part by a grant from the National Institutes of Health (NIH) Common Fund, CAVATICA is continuously updated with new tools and datasets.

CAVATICA's goal is to help pediatric researchers collaborate, share, interoperate, and connect with any and all other data ecosystems in order to empower data analysis across diseases, ages and geography. By connecting previously disconnected datasets, CAVATICA supports researchers and patients across the United States and throughout the world to generate new insights into pediatric diseases.

Flywheel's cloud-native medical imaging platform was deployed in an AWS environment alongside the Velsera CAVATICA platform and other tools in the research ecosystem. Flywheel and Velsera collaborated to design and build a cross-platform integration to facilitate pre-processing, cohort selection, and platform specific analytics using both genomics and imaging data.

Another important aspect of this project involved expanding the institution's cohort discovery tool, which required interoperable data models between genomics, clinical information and imaging data. The result allows researchers to select cohorts using data modality relevant criteria. For instance, searching a genomics dataset might involve selecting genotypes, while medical imaging searches might focus on modalities like MRI or CT scans, equipment vendors, or time of visit information like date of scan. This refined approach allows researchers to access, combine, and analyze data to derive insights that otherwise might have been missed.

Multi-domain data integration & analysis

Once a cohort was identified, the relevant data was added to a shared project accessible by domain-specific experts. The joint platform, using Flywheel for medical imaging and CAVATICA for genomics, enabled the NIH researchers to organize and prepare these large data sets for analysis, allowing them to extract critical derivative data, including tumor size from images and genetic markers from genomic sequences. These derivatives then formed the basis for integrated analytics to identify anticipated cross-domain relationships or patterns —which is crucial work before one can establish AI-based predictive models.

This AI-driven approach exemplifies the application of machine learning to decipher intricate patterns in data that would be challenging or impossible for humans to discern. The result is a more nuanced, efficient, and accurate analysis. The joint Flywheel/ CAVATICA framework is a testament to the power of integrating and harmonizing diverse datasets, providing a scalable, versatile, and robust platform for the future of clinical research and personalized medicine.

The teams expect to extend this effort to enable multi-site data contributions and expanded imaging modalities through standardized de-identification, review at scale, multimodal cohort selection, and combined dataset cross-discipline collaborations.

The Value and Impact of Collaboration

The broader value of this joint project lies in its potential to impact patient care, both at an individual and a population level. Pharmaceutical companies running clinical trials often focus on specific population subgroups, analyze their symptoms, look for biomarkers and other clinically relevant patterns and develop a treatment plan. However, populations not initially considered in the subgroup might be left without clear insights into their treatment options. The interoperable Flywheel and CAVATICA platforms aimed to bridge this gap, ensuring more inclusive research and treatment approaches that consider the needs and characteristics of varied patient groups.

This multimodal research ecosystem is a pattern that is emergent across visionary customers that Flywheel engages with in the research space. The model established in this effort demonstrates a framework that is extensible to enable access to multimodal datasets and analytics at scale. One can readily see how to extend the medical imaging and genomics solution to include other data types such as digital pathology or clinical data from electronic health records.

*Visit **flywheel.io** for more on how Flywheel is enabling research in high gear.*