Published by *FierceMarkets*

Real-World Evidence in the Cloud:

How Technology is Revealing the Big Picture in Pharma

aws

The collection and analysis of data generated outside of restricted clinical trials are now critical to pharma companies. Data gathered in the routine delivery of healthcare can be mined to uncover insights into treatment efficacies and outcomes in the real world. Such insights, known as real-world evidence (RWE), benefits teams working across the drug lifecycle by showing how larger, heterogeneous patient populations interact with the healthcare system and respond to medicines.

Until recently, pharma companies and other healthcare stakeholders had access to a limited range of real-world data. Outside of formalized clinical trials, companies were only able to collect basic observational research data by monitoring billing codes submitted to healthcare payers, or from registry studies, giving them a limited picture of how their products were used in practice and their effect on patients.

These data sources are now just a few parts of a much broader tapestry. Today, companies can pull in longitudinal patient information from electronic medical records (EMRs), genomic studies, social media, wearables and other sources of real-world data to build a more detailed, robust understanding of the use and value of their products.

RWE is now a key resource for every step in the life cycle from drug discovery to commercialization. This new normal grew out of the simultaneous expansion of the breadth and depth of datasets, growing focus on value—in some cases by tying payments to outcomes—and need to make processes across the drug lifecycle faster and more efficient.

The AWS Cloud is facilitating RWE by making it easier for pharma companies to ingest vast datasets from a range of sources, and by providing analytical tools, including artificial intelligence capabilities, to

efficiently analyze this data and provide insights. This centralization of data assets enables organizational intelligence by equipping companies to understand the needs of patients and other stakeholders, and how their products can meet them. The result is better-informed companies that more efficiently improve patient outcomes and address the pain points of payers and providers.

WHY THE WAY COMPANIES APPROACH RWE IS CHANGING

The increasing demand to move to a more 'value-based' health care system, where therapeutics are only reimbursed in full if they achieve certain predefined health outcomes, is accelerating the push for companies to deploy effective RWE. The success of these activities, however, rests on a company's ability to ingest, store and analyze large amounts of data from multiple external sources. Companies must efficiently and cost-effectively combine vast, disparate data sources into a single, centralized repository that is easily accessible by all their teams to prosper.

The first potential pitfall in this process is the sourcing of real-world data. Providers and payers can generate extensive data on the real-world use of medicines, whereas pharma companies largely lack such abilities. As pharma companies have less internal access to knowledge about how their drugs work in the real world than other stakeholders, they need the means to source and ingest third-party data from multiple providers.

In this model, individual departments in a pharma company address gaps in RWE assets by licensing data from external vendors and storing it in on-premises infrastructure that is siloed from the rest of the organization. This approach serves the near-term needs of the single department, but it causes inefficient sourcing and makes it hard to maximize the value of purchased data assets by sharing them with the entire company. As companies lack a clear picture of the assets acquired by different teams, multiple groups license similar data, leading to the proliferation of vast, redundant datasets stored on relatively-expensive on-premises infrastructure.

Furthermore, even if multiple groups are assembling datasets into a single in-house location, it can be a struggle to provision the right type and size of resources in physical infrastructure to optimally store and manage disparate data formats at the vast scales.

Typically companies struggle to properly gauge the level of storage and compute power they will need, either over or under buying equipment, which is exacerbated by the potentially long procurement cycles required to acquire this infrastructure. Particularly as new and unpredictable types of data may be collected in the future to support RWE programs, companies may find the questions they can ask of the data limited by the constraints of their legacy data centers.

The impact of the shortcomings of this approach has grown in step with the availability of data and need to effectively analyze and use it across the product lifecycle. Some companies are analyzing the genomes of patients and their tumors to inform steps as far back in the life cycle as target discovery. Others are mining data on drug prescriptions and EMRs to show the potential availability of patient populations that may be qualified to participate in clinical trials in different regions. However, without the capacity to effectively store, analyze, and share large-scale and disparate datasets, the value of this data cannot be realized.

With everything from target discovery to reimbursement now reliant on vast amounts of real-world data, leading companies are concluding that inefficient and ineffective approaches to RWE are no longer viable.

"We required a scalable platform that would provide us with cutting-edge analytics and enable us to work with large volumes of real world evidence that are distributed globally across disparate sources," Patrick Loerch, Celgene Senior Director Data Science, said.

Celgene, like many other companies, looked to the AWS Cloud and AWS Technology Competency Partners, to alleviate the challenges associated with large-scale data storage, analytics, and secure organizational data sharing.

HOW DATA LAKES ON THE AWS CLOUD ENABLE EFFECTIVE RWE

Companies are responding to the emerging RWE environment by replacing siloed, on-premises data repositories with cloud-based data lakes that pool and catalog all of their information assets as centralized repositories.

Instead of every department storing real-world data on their own isolated systems, companies such as Celgene are having newly in-licensed data flow into an AWS Cloud data lake. By leveraging the market tested analytics capabilities provided by Deloitte ConvergeHEALTH Miner, teams working across the product lifecycle are equipped to access, analyze, and visualize the assets to support their specific needs.

Feeding data from many sources into a single, centralized repository creates continuously-growing, multi-petabyte data lakes. The need to store large, growing amounts of real-world data makes the dynamic scalability and low cost of cloud storage critical to the viability of the concept of centralized data lakes. Working with AWS cloud also enables companies to leverage the provider's previous work on the structure, metadata system, governance and security of data lakes. This work coupled with Deloitte ConvergeHEALTH software,

that is deployable in minutes from AWS service catalog, makes it possible to create secure, easily searchable data lakes in minutes, as opposed to the potentially long lead times to procure and set up on-premises IT infrastructure.

The company-wide data access enabled by cloud-based data lakes is the first step to empowering teams to seize opportunities created by RWE. The second step is to enable workers of all levels of technical know-how to glean insights from data. Collating and mining real-world data for evidence was traditionally a technically demanding task. This is changing, though.

Celgene has created self-service, point-and-click applications to enable non-coding staff to perform analyses.[1] With the associated powerful data analytics in the cloud, Celgene reduced patient cohort analysis time from months to minutes. Using these applications, anyone can question data and produce interactive visualizations, cutting the time and resources it takes to generate insights.

The flexibility afforded by building a data lake on AWS means you can ingest data in any format and can perform data analysis on-demand, regardless of what new types of data may be available from pharma patients in the future. And hosted in the highly secure cloud infrastructure with programs to help organizations meet local and regional regulatory rules for data protection and provisioning, including HIPAA, HITRUST, and GDPR requirements.

A 2017 Deloitte survey[2] suggests companies that realize these benefits by adopting cloud systems are happier with their RWE setups than peers that use on-premises infrastructure. The survey found more than half of leading pharma companies are investing to significantly improve RWE capabilities, typically due to dissatisfaction with their existing on-premises systems.

In contrast, companies with cloud-based RWE systems are far less likely to feel the need to significantly improve their capabilities, with 80% stating they are content with their RWE capabilities. This satisfaction gap reflects the fact that the scalability, security, and speed of the cloud make it a better fit for the RWE needs of modern pharma companies.

HARNESSING ARTIFICIAL INTELLIGENCE TO ACCELERATE ANALYSIS

Easy-to-use, real-world data analysis tools are being supercharged by artificial intelligence and machine learning. Today, vast quantities of data are available but the size of the repositories and complexity of the information makes manual, human-powered analysis impractical, if not impossible. Fortunately, artificial intelligence and machine learning can sift through data far quicker than manual inspection, making it possible to generate insights within a fraction of the time it would take a human.

At the same time, AWS is lowering the barriers to entry to machine learning to expand what companies can learn from their data.[3] GE Healthcare is using services such as Amazon SageMaker to streamline the development and deployment of machine learning models to help radiologists improve health outcomes while reducing errors.

Similarly, a pharma company could mine reimbursement claims data, historic study site

1 *Invent 2017: Real World Evidence Platforms to Enable Therapeutic Innovation (LFS302).* (2017). http://paperpile.com/b/xoZLH3/4WNp

2 Getting real with real-world evidence. Available at: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-ls-2017-real-world-evidence-survey-031617.pdf. (Accessed: 26th June 2018)

3 Partnering for AI Innovation. *WSJ* (2018). Available at: https://partners.wsj.com/aws/partnering-for-ai-innovation. (Accessed: 24th July 2018)

performance records, and the details of doctors working in a particular therapeutic field for insights that accelerate clinical trial enrollment. While traditionally these datasets have been too vast to analyze manually, machine learning capabilities in the cloud can potentially reveal which sites should have access to patients in the active disease state and how they performed in the past.

This is one of a fast-growing number of ways artificial intelligence and machine learning are aiding RWE. Another is in the improvement of natural language processing services that can be used to find insights and relationships within text. Companies can utilize services, such as Amazon Comprehend, to analyze notes and other unstructured assets, enabling the mining of resources that were largely closed off from researchers in the past, but potentially hold critical insights.

Whatever the application of machine learning, the approaches are built on access to data. Cloud-based data lakes facilitate faster and more efficient insights by bringing all the data into one place and storing it in a way that supports the pooled analysis of data assets from vastly different sources.

PREPARING FOR TOMORROW

Many of the trends identified in this paper are only going in one direction. Datasets will continue to get bigger and more complex. Payer demands for evidence of real-world value will become more intense. The abilities of artificial intelligence, machine learning, and natural language processing will continue to grow and evolve, enabling teams to complete tasks in hours, not weeks. This will become the new normal for pharma companies.

Now is the time to prepare for this new normal. RWE is and will remain, a dynamic space defined by the need to quickly ingest and analyze new datasets as they become available. Executing these tasks will yield insights that drive the development of safer, more efficacious medicines and the smarter use of existing therapies.

Companies can seize these opportunities by building platforms that use AWS' cloud and services to simplify and automate the collection and analysis of real-world data. These forward-thinking companies will do more than just gain an advantage today. They will equip themselves to come out on top as RWE becomes a critical pharma battleground in the years to come.