

GLOBAL PHARMACEUTICAL COMPANY

Testing Thousands of Hypotheses with an FDA-Qualified Research Platform

Industry

Life Sciences

Location

Headquarters: North America

Use Case

- Translational bioinformatics research
- Laboratory automation
- Sales forecasting
- Market segmentation

Impact

- Successfully test tens of thousands of hypotheses with complete reproducibility to accelerate their fight against cancer and other chronic and life-threatening conditions
- Improved efficiency with an FDA-qualified platform that streamlines documentation processes

Data Science Scale

- 10,000+ hypotheses tested
- 2,000+ total Domino users (includes consumers of research projects)

Users

- 500+ data science practitioners, researchers, biostatisticians, data engineers, and IT

Solution Components

- Data Science Tools: MATLAB, Python, PyTorch, R, SAS
- Cloud Infrastructure: Amazon Web Services
- Data Science Platform: Domino

Data Science at a Global Pharmaceutical Company

While advances in research over the last decade have accelerated new insights about diseases that can lead to groundbreaking medicines and better patient outcomes, even the most efficient drug development process is time consuming. In the U.S., it typically takes more than 12 years to make it through [all five FDA-regulated steps](#), according to the U.S. National Library of Medicine¹. One way biopharmaceutical companies can improve time-to-market of new medicines is to ensure their research environments operate in lockstep with FDA requirements. A global biopharmaceutical leader has implemented an FDA-qualified research system of record on the Domino data science platform to streamline its entire research pipeline—from drug discovery to laboratory automation to customer analytics—ultimately paving the way toward faster medical advances.

By consolidating multiple technology stacks in Domino, which supports closed and open source analytical tools in a single platform, the organization has been able to:

- **Rapidly test tens of thousands of hypotheses, and explore outcomes and learnings with complete reproducibility.**
- **Accelerate the submission of test results to the FDA.**
- **Optimize pharmaceutical development to ensure robust, predictable, and scalable manufacturing processes.**
- **Catalog and understand the different tools and data sources available—including new data types, such as molecular-based data, real-world data from fitness trackers, and patient data from hospitals.**

¹ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5725284/>

Challenge

This biopharmaceutical company embarked on a digital transformation journey to modernize its statistical computing environment (SCE), standardize research systems and processes, and drive faster innovation. Key milestones on this journey included migrating legacy systems into the cloud and coalescing the research efforts taking place across all aspects of its business—discovery, development, production, and sales. They needed to make it easier for researchers, biostatisticians, data engineers, data scientists, and IT teams to:

- **Independently access diverse tools and elastic compute resources in the cloud.**
- **Share expertise and resources.** (At the time, the company's 500-plus data scientists, researchers, and statisticians typically worked in silos.)
- **Test different scenarios and reproduce results with ease.** This is critical for patient safety and regulatory requirements, but also quite challenging. To reproduce a testing result, for example, the company has to capture and track numerous dependencies including:
 - the statistical analyses selected,
 - the scripts that implement the analyses,
 - the libraries that implement statistical functions and perform the mathematical computations,
 - the operating system that runs the environment,
 - the data reduction processes,
 - and the raw data chosen.
- **Streamline workflows to productionize models more quickly and publish findings to more than 400 different "consumers."** These consumers include peers in different parts of the development process and a wide variety of internal and external partners who leverage this work in everything from documentation and validation to determining where to run their next marketing campaign for a specific drug.

Solution

The biopharmaceutical company deployed an FDA-qualified research system of record on Domino that enables researchers across the company to more quickly test hypotheses, reproduce results, build off past work, and submit their findings to the FDA. The environment is available to anyone in the company; active users represent translational medicine, observational research and data science, and business intelligence and analytics.

How does it work?

As researchers and data scientists begin their work, they use the Domino Workbench to quickly access: data sources; tools such as SAS, R, Python, PyTorch, MATLAB, and a wide range of domain-specific tools; and compute resources in Amazon Web Services. Standardized configured environments are available via a shared catalog, providing a one-stop approach for users to get up and running for typical data science workflows. Data scientists can also quickly view a snapshot of curated data pipelines and previous analyses that exist as they begin their work, in case there's something that can be repurposed for their needs.

All code, data, tools, and packages used during research and development (R&D) are automatically tracked, along with any comments and annotations, so users can quickly reproduce results, build off existing work by others, and share ideas. This capability not only accelerates research but also helps ensure that valuable intellectual property isn't lost when employees leave the company. Additionally, this reproducibility is foundational to improving chemical development and manufacturing processes, which require highly precise and sophisticated procedures to document every stage of the process, optimize ideal reaction conditions, and provide full accountability of any changes.

Domino drives documentation as a code process, making it easy for researchers and data scientists to package up all relevant information in a Domino API when they're ready to submit results to the FDA.

The company took a phased approach toward its Domino implementation, onboarding 15 teams one at a time based on their prioritized compute needs and readiness to adopt. As part of deployment, Domino provided subject matter expertise and configuration, application migration and user support, along with internal training and resources to help users get up and running on the new platform quickly.

Use Case: Cancer Research

According to the [American Cancer Society](#), in 2020 there will be an estimated 1.8 million new cancer cases diagnosed and 606,520 cancer deaths in the United States alone. Data science is vital for finding connections in data that will help improve patient outcomes and prevent cancer.

One key goal of this biopharmaceutical company's research is to improve the understanding of the relationship between genetic changes, such as DNA mutations that drive tumor formation, and how cancer cells exploit immune checkpoints to sabotage a patient's immune system.

Today, using its Domino research platform, the company can analyze data from more than 10,000 tumors to uncover possible connections. This work has led to discoveries critical in expanding survival rates for patients, including:

- **Identification of new classes of drugs** that can infiltrate tumors and the ability to correlate their impact on each individual, which help doctors identify which specific therapies offer the best outcomes for each patient.
- **New insight into how the immune system responds** to cancer cells, including the discovery of a specific chromosome that can impact immunity to cancer cells.

The Domino Effect

- **Driving discoveries:** With the help of Domino's platform, researchers can track and trace reams of genomic data aggregated from thousands of tumor samples. And they can successfully test tens of thousands of hypotheses with complete reproducibility, all to accelerate their fight against cancer and other chronic and life-threatening conditions.

- **Increasing efficiency:** Researchers and data scientists can submit research to the FDA from the same platform they use for R&D. This capability has eliminated the need for a separate team to re-write the code on a different platform.
- **Enabling innovation:** This work is foundational to the company's digital transformation initiative which will transform biopharmaceutical research with real-world health data, driving more discoveries and treatment options.

About Domino Data Lab

Domino Data Lab provides an open data science platform to help companies run their business on models. Model-driven companies like Allstate, Bristol-Myers Squibb, Dell, and Bayer use Domino to accelerate breakthrough research, increase collaboration, and rapidly deliver high-impact models. Founded in 2013 and based in San Francisco, Domino is backed by Sequoia Capital, Coatue, Bloomberg Beta, and Zetta Venture Partners. To learn more, visit dominodatalab.com.