CODE OCEAN | sema4

CASE STUDY

Sema4 Customer Case Study

Automated IT infrastructure helped growing biotech firm with secure, scalable, collaborative environment to enhance and accelerate research

Sema4 is a precision health company focused on artificial intelligence (AI)-driven analytics. Sema4 uses its Centrellis® health intelligence platform to aggregate, curate, annotate, structure, and analyze large clinical and biomarker datasets in order to build and share actionable and predictive models of human health.

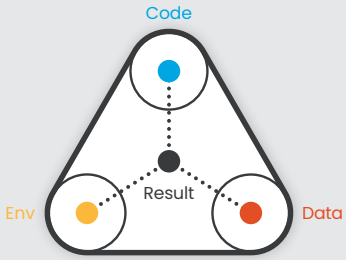
Centrellis yields insights that can indicate conditions of individual patients, drive informed decision-making across the drug development pipeline, and provide health systems with access to critical data. These insights are essential to developing information-rich genomic tests, such as Sema4 Signal™, that help detect disease, identify effective therapies, and support healthcare providers with data to help guide treatment.

Challenge

Sema4 needed a robust infrastructure for its R&D precision oncology team to analyze:

- **Clinical data** from approximately 12 million de-identified clinical records, including more than 500,000 with genomic profiles
- **Biomarker data** coming in from multiple partner sites
- **Large computational models ("AI")** required by massive new data sets

This required a cloud-based infrastructure solution where geographically dispersed, multi-disciplinary teams could develop, run, publish, view, and share computational analysis — all through a simple browser interface. The infrastructure needed to be interoperable with the Centrellis platform to provide the access, scalability, and collaboration that's so critical for teams to be effective.



Code Ocean Capsules capture the essential triplet of reproducible research — environment, code, and data.

There's no guarantee that one person's code will run perfectly on another person's machine, if at all. The individual preferences of researchers, from languages to how and when to update software, IDEs, and hardware, complicate the situation even more, not to mention the DevOps skills required to manage an IT environment. All too often, these multiple, complex challenges lead to the same outcome: results that can't be validated.

Code Ocean Capsules eliminate these discrepancies by automating every step of the process to ensure results are preserved and the lifecycle is managed as a complete package for trusted sharing, collaboration, and reuse. Automated version control using Git means that every change or edit to the code or environment is traceable and transparent. And built-in access management means teams can access and manage common datasets securely shared between Capsules.

Solution

Code Ocean's containerized Compute Capsule technology was the ideal solution for Sema4. With Code Ocean, bioinformaticians can work with their language of choice in a familiar, user-friendly Integrated Development Environment (IDE). Core reproducibility and sharing practices are automated with built-in Docker environment containerization and Git version control.

The Code Ocean Lab platform was deployed on Sema4's Amazon Web Services (AWS) virtual private cloud (VPC), enabling Sema4's complete control over data governance. As with all clients, Code Ocean worked with Sema4 to ensure compute resources were optimized for performance and cost-efficiency. This led to a solution where Sema4 has multiple cloud resources which are allocated to projects on an on-demand basis, and automatically scaled down when not in use.

> **"It's like night and day. The Code Ocean platform has made us significantly more efficient in onboarding and closing gaps in technology skills. Compute Capsules have cut down on the time it takes new team members to be up and running and allows them to focus from day one on research, rather than on getting set up for their work."**
>
> Scott Newman, PhD
> Senior Director of Precision Oncology, Sema4

Results

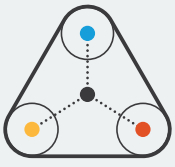
Sema4 first noticed the impact of Code Ocean when onboarding employees.

Before Code Ocean, each computational researcher had to set up a computing environment using the same language, software packages, and versions as their colleagues. This process proved to be time-consuming, inefficient, and costly, as new users struggled to learn how to set up their machines, along with workflow processes that aligned with what their predecessors had coded.

This meant initial set-up for new users could take days, if not longer, all of which delayed actual work from even getting started.

Code Ocean made it possible for Sema4 to ensure their research teams could immediately focus on data analysis by reducing the amount of time associated with onboarding. Different teams now use consistent processes to share, compare, analyze and cross-validate data results quickly and easily.

And the researchers at Sema4 know that Code Ocean's Application Science team of experienced developers and bioinformaticians are available to provide consultation and answer any questions.



See how Code Ocean can help you reach your business goals

codeocean.com/demo