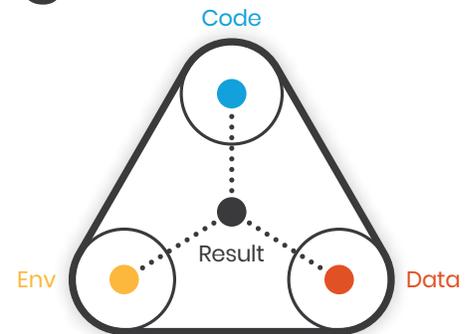# How Sema4 Is Shortening Time to Innovation

by Increasing Productivity Through the Use of Code Ocean Compute Capsules™

Sema4 is a patient-centered health intelligence company that advances healthcare through data-driven insights. The idea underlying Sema4's work is that more information and deeper analysis of health data will improve the diagnosis, treatment, and prevention of disease. To achieve this goal, the company aggregates, curates, annotates, structures, and analyzes large data sets using their analytics platform Centrellis™. This platform is then used to build predictive models of human health using artificial intelligence (AI)-based algorithms.

Centrellis yields insights that can inform on individual patients' conditions, power informed decision-making across the drug development pipeline, and provide health systems with access to critical data. The company also uses these insights to develop information-rich genomic tests, such as Sema4 Signal™, that help detect disease, support healthcare providers in identifying the best therapies for each individual patient, and guide treatment decisions.

Sema4 addresses the challenge of effectively applying high-end data, such as genomic information, to support health systems and healthcare providers in better understanding each patient and their journey and matching each patient with the best personalized care.

Goal

Set up a consistent, scalable, shared environment that allows for fast, automated and secure access by computational researchers with different expertise and experience and guarantees consistent results.

Key Benefits

- Reduces onboarding time from hours or days to minutes

- Helps close technology gaps for researchers

- Guarantees consistency and reproducibility of analyses

- Enables frictionless sharing of data and results which improves productivity

The Challenge

Sema4 collects, aggregates, and analyzes large amounts of complex clinical and genomic data. This task is complicated, especially for a fast-paced, rapidly growing company handling an extensive portfolio of projects and an ever-growing computational research team composed of researchers with diverse experiences and expertise. Sema4 faced challenges related to onboarding new employees and data-sharing between different teams, specifically:

- **Setting up the computing environments** – for each project, each computational researcher had to set up a computing environment that uses the same language, software packages, and versions as their colleagues. This process proved to be time consuming and inefficient: new users had to spend significant time finding out how to set up their machines and standardize them. For team members without a strong software engineering background, initial set-up could take days and therefore delay them starting on their actual work.
- **Setting-up security** - similarly, each user had to set up a workflow with credentials that allowed them to securely connect to databases and cloud storage with all of their computing devices. This process often took several hours to complete in each local and cloud environment. Users with little prior experience in managing security settings, in particular, lost valuable time working through this process.
- **Data sharing** – inefficiencies were encountered when sharing data and results and performing collaborative work . To review or add to the work shared by a colleague, researchers had to first ensure that they not only installed all the necessary software packages but also that their computing environment matched that of the original analysis down to the version of every package used.
- **Obtaining consistent results** – with a multitude of analysis packages to choose from, individual preferences for computing languages, and different levels of experience, consistency could vary. Even answering the same question using the same data set,

different computational biologists could get different results based on how they approached the analysis and wrote the code. Consolidating these results required time- and resource-intensive recreation of every step.

Sema4's goals were to ensure that their research teams could focus on data analysis by reducing overheads associated with set-up and security, and that different teams could use consistent processes for data analysis and share, compare, and cross-validate their results with ease. By achieving these goals, Sema4 could expedite delivery of value to their partners and patients.

> ❝ 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

The Solution

Code Ocean is a cloud-based platform for computational research that streamlines workflows for creating, organizing, and sharing complete project assets. By allowing researchers to package and share standardized Compute Capsules™ consisting of data, code, computing environment and results, Code Ocean ensures transparency and scalability across large and diverse teams guaranteeing consistency and reproducibility of research results.

Sema4 selected an initial project to determine whether Code Ocean's Compute Capsules could address their needs and learn how to configure the platform. They selected a project that involved pulling deidentified data stored in a database, analyzing these large data sets, and finally sharing and comparing their results.
In collaboration with the Code Ocean team Dr. Scott

Newman, Senior Director of Precision Oncology at Sema4, developed a Compute Capsule for data analysis and shared it with his team. The Capsule contained all the software packages needed to access the required database and perform the analysis in a standardized environment. The Capsule also allowed for instantaneous sharing and collaboration.

The Results

Using the Code Ocean Compute Capsule addressed all of the challenges the team at Sema4 faced quickly and comprehensively:

- **Setting-up the computing environment** - set-up with Code Ocean was almost immediate. Using the shared Capsule, researchers could focus on their actual work of analyzing data rather than losing time on set-up. The Code Ocean platform took care of operational tasks, e.g., infrastructure and environment set-up and resource management.
- **Setting-up security** – with Code Ocean Compute Capsules, working in a secure environment is just a few clicks away. Instead of each individual researcher setting up and managing security for their computing environments, this set-up is done once on the Capsule level by experts in the field. New users can – once they receive their credentials - access the secured and compliant Capsule and start working immediately. This process not only allows the researchers to get started quicker but also facilitates collaboration. If a researcher wants to share their existing research, a collaborator with the appropriate credentials can access and start

working on the code and data immediately rather than spending hours on setting up the needed security.
- **Data sharing** - the Capsule makes it easy for all research team members to share with each other not just results but also data, analysis pipelines and computing environments. Collaborators can now edit directly in a transparent, easy to recreate way. To address any concerns around privacy and security of both raw data and analysis results, access to the information in the Capsule can be managed on different levels allowing the Capsule owner to balance the need to comply with security and privacy concerns.
- **Obtaining consistent and reproducible results** – using the Code Ocean Capsule guarantees that analyses are carried out in a standardized way leading to highly consistent, reproducible results. Since Compute Capsule activities are version controlled, every change or edit to the code or environment is traceable and transparent, making trouble-shooting easy and fast.

Code Ocean Compute Capsules enable the computational scientists at Sema4 to focus on doing their work rather than setting it up and deliver insights faster to Sema4's partners. Capsules provide a standardized environment and uniform processes that ensure consistency and reproducibility of results.

Sema4 has now adopted the Code Ocean solution and is rolling it out across the company.

Why Code Ocean

- Allows customers to focus on their work rather than on managing tools
- Highly flexible, configurable, and secure capabilities
- Integrative computational research experience, emphasizing researchers' productivity and reproducibility
- Company focus on solving the needs of researchers today and in the future with a flexible platform
- Superb service, exceeding expectations of operational support, with transparent upgrade processes