code ocean

How Sema4 improved global onboarding time for new computational team members

sema4

The challenge

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

However, the Sema4 team needed a solution for computational research that could work with Centrellis while also handling a demanding set of requirements:

- **Multiple geographical locations:** the Sema4 team was spread across multiple different locations, and needed a way to make sure they could collaborate effectively
- **High data volume:** the team were working 12 million de-identified clinical records, including more than 500,000 with genomic profiles
- **Multiple data sources:** they were also working with biomarker data coming from multiple different partner locations

Previous approaches had been a significant time and resource sink for the Sema4 team:

- **New user IT burden:** all onboarding users required several days' worth of environment setup, aligning programming languages, software packages and versions
- **Difficulty collaborating:** because of the wide geographical spread of the business, working between different teams was often difficult

" 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

Key features used

Data dashboard

Capsules

Collections

Admin panel

code ocean

The solution

Sema4 implemented Code Ocean as their primary cloud-based computational research platform:

1.

Installed Code Ocean on their Amazon Web Services (AWS) virtual private cloud (VPC), giving them full governance and control over their data without having to upload/transfer it anywhere else

2.

Onboarded all existing users into Code Ocean, bringing them onto a common platform for computational work

3.

Established common standards for computational research that could be co-developed, shared, analyzed, and cross-validated

About Sema4

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.

The results

Onboarding
This now takes days, not weeks, and requires almost no IT support

Standardization
Consistent processes established to share, compare, analyze and cross-validate computational work

Key integrations

Amazon S3   AWS EC2   R   Python   Docker   Git