code ocean

How The Allen Institute builds and shares computational pipelines for open neuroscience

The challenge

The Allen Institute for Neural Dynamics (AIND) has a strong commitment to Open Science and wanted to share its pipelines with the wider neuroscience community. They had been running custom pipeline management code with on-prem HPCs, but this system wasn't designed for export or sharing outside of AIND.

This was a critical blocker for some of their most important projects (including OpenScope). There was also an opportunity for AIND to make Allen-validated pipelines easier to adopt in other laboratories around the world.

About the Allen Institute

The Allen Institute is a Seattle-based independent nonprofit bioscience and medical research institute founded by Microsoft co-founder and philanthropist Paul G. Allen. It conducts large-scale research through foundational science to fuel the discovery and acceleration of new treatments and cures for diseases such as Alzheimer's disease, heart disease, cancer, addiction, and more. The Allen Institute is committed to open science; one of its core values is to make all data and resources publicly available for external researchers and institutions to access and use.

"Code Ocean makes it easy for our scientists to do their work reproducibly ... New users to the platform can get far with just a little support; this gives our engineers time to focus on domain-specific challenges."

Dr. David Feng
Director of Scientific Computing, the Allen Institute for Neural Dynamics
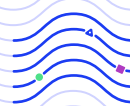
Key features used

Pipelines

Capsules

Data

code ocean

The solution

The team at the AIND used Code Ocean to:

1.

Design and build scalable pipelines using the visual pipeline builder. Pipelines created this way auto-generate Nextflow, and run on AWS Batch.

2.

Automate pipeline execution with the Code Ocean API, triggering when certain conditions are met or wherever otherwise specified by the AIND team.

3.

Use the Pipeline export function to share their pipelines with the wider neuroscience community for use outside of Code Ocean (notebook format instructions)

This was possible because of how Code Ocean is built:

- The visual pipeline builder automatically generates a Nextflow script in the background, increasing access to pipeline creation
- Code Ocean is built on open source tools like Docker, Git, and Nextflow, making export, interoperability and use outside of Code Ocean possible
- There's no lock-in to Code Ocean: all Pipelines and Capsules can be exported from Code Ocean for use elsewhere

The results

Productivity
Pipeline build time went from 12 weeks to 3, a 4X speed increase

Open science
Complete, interoperable pipelines are now shareable with a single click

Efficiency
10 AIND engineers successfully reallocated to higher impact projects

Self-service
Has allowed individual scientists to build pipelines with little support

Key integrations

| Python | AWS Batch | AWS EC2 | Nextflow | Docker | Git |