Scientific research increasingly requires sophisticated computing power and code to manage massive datasets and complex experiments. Many researchers know how to create computational algorithms, but lack the software engineering skills to effectively manage the computing environments needed to produce results. Code Ocean aims to remove the friction from computational research operations and collaboration with a platform that guarantees one of the most challenging goals of scientific work — reproducibility.

The Code Ocean integrative computational research experience is one place where great computational research is created, organized, and shared. Code Ocean delivers a unique experience with a standard, secure, and executable research package called a Compute Capsule. Using popular computational tools, compute and storage resources and environments together with a configurable workbench and an integrated repository the platform closes the usability and technology gaps needed for researchers and research teams to quickly / reliably increase flow, reproduce results, and share knowledge and discoveries. It is the best way to standardize research workflow, and track and reproduce all computations and discoveries.

A Platform for Research, Informatics, and IT

The platform continuously captures the exact version of code, data, and the execution environment that generated every result. The platform also automates labor-intensive compute infrastructure tasks to let research teams focus on science and optimize computing costs. With an integrated cloud-based workbench and secure repository, the Code Ocean platform powers a transformative boost in productivity and quality for computational research teams. Leveraging other researchers project work is easy, selecting powerful computing is one click away, and any size dataset is attached instantly to any project.
With computational research work, results are often hard to reproduce, reuse, and share, which lowers productivity and can negatively affect quality. The Code Ocean platform provides assurance on how every result is generated, with reproducibility at the heart of the platform design goals. Computational research projects are organized clearly so that data and code are distinguished and environment dependencies are provided and documented. The Code Ocean platform generates a standard, secure, and executable research package called a Compute Capsule. The Code Ocean Compute Capsule format is open, exportable, reproducible and interoperable. Each capsule is versioned and contains code, data, environment and the associated results.

Capsules are securely stored in a cloud-based repository for long-term preservation and reuse. Capsules continuously capture daily iterations, ensuring data provenance so you can reproduce projects back to any point in time.

Platform Overview
Compute Capsules and Workbench

Compute Capsules capture complete projects using a Computational Workbench

A compute capsule is the basic unit of a research project managed on the Code Ocean platform. It is a modular container for the software environment along with code and data, that runs in a browser. A compute capsule includes four components: the computational environment in which the code was run, the code, the data that was used, and the results that were obtained, including the tools used to visualize them. Capsules can be created in the Computational Workbench with simple clicks.

The computational environment is a Docker image built by user-specified Dockerfiles, and it defines what software packages the research project uses.

Compute Capsules enable reusable and reproducible computational projects

Docker offers an advanced technique to build software images so that they can package and share the same software with others. When different researchers use the same image, they use the same software even across operating systems, therefore guaranteeing consistency in research results. The Workbench’s GUI-based Dockerfile and Docker image generation allows any researcher to enjoy guaranteed reproducibility without needing to be skilled at Docker. Researchers first need to select a starter environment, then add packages for the programming language used. Once this is done, a Dockerfile is created without typing any Docker command. Then researchers just need to add code and data to make the capsule ready to run.

Compute Capsules are the basis for end-to-end reproducible analysis and processing

Compute Capsules use the power of flexible cloud computing and storage to run on an Amazon Web Services (AWS) virtual machine in the client’s Virtual Private Cloud (VPC). The Reproducible Run button executes a master script in a Compute Capsule from beginning to end. This way researchers can reproduce results from an end-to-end analysis with just one click. Researchers use Reproducible Run when they have already figured out the whole analysis procedures and want to get the end results.
A collaborative workbench streamlines all your computational tasks, while using popular computational tools

The Code Ocean Computational Workbench integrates popular computational tools such as Docker, Git, Jupyter, RStudio, Shiny, Amazon Cloud services to help research teams quickly execute computational analysis, save results, and share reproducible research in executable Compute Capsules. The Code Ocean platform speeds usability, tool integration, devops, and lifecycle tasks by closing technology gaps with a highly intuitive, ready-to-use research user experience.

Interact with code and data using popular tools in a browser
Researchers run capsules via the Cloud Workstation feature of the workbench when they want to explore the dataset or just run part of the code. This allows the user to analyze data interactively when running the entire analysis is either unnecessary or not yet ready. It is mostly used when researchers are still trying to figure out the entire processing pipeline by interacting with data.

View and edit code without consuming compute time
Reproducible Run and Cloud Workstation features both consume compute hours. However, just editing and viewing the code within the integrated workbench Capsule Viewer does not consume any compute time, offering a cost-effective way to edit code.

Save changes to Git without a command line interface
Git-tracked changes made to the code are usually done via a command line interface. To track any edited change, a scientist needs to type in commands to stage and commit the files in the terminal. Capsules within the Code Ocean Computational Workbench are integrated with Git; no command line is needed for the most frequently used Git features. The Capsule UI automatically saves and stages changes, then users can commit the changes with just one click.

Sync Code Ocean workbench projects with existing GitHub repositories
Capsules are Git repositories themselves, hosted on the server where the Code Ocean platform is deployed. Researchers can bring their private GitHub repositories into the Code Ocean platform as capsules and keep them in sync.

Enable seamless collaboration
Researchers can allow immediate access for internal or external collaborators to work on their projects. With the Code Ocean platform, collaboration is as easy as sharing a link and granting permission. Collaborators can be added to any capsule from the Capsule Viewer with access levels of view only or edit.
Share research assets within your team or organization
Once a project is done, the owner of the capsule can publish the capsule so everyone from the organization can see and reuse it. This makes it easier for other members of the organization to find projects of interest, leverage and reuse written code, and build on others’ templates.

Manage pipelines for reproducible multi-step processing
Computational biologists use workflow management tools like Nextflow, Snakemake, Cromwell, etc. to manage complex pipelines that connect different processing steps for the entire analysis. These workflow management tools can be installed and used in Compute Capsules, thus allowing completely transparent and reproducible end-to-end pipelines, which are easily shared.

No platform lock-in
Everything created in the Code Ocean is exportable, downloadable, and cross-compatible with outside tools. Users can export capsules in whole as a zip file, and rebuild them using open source tools like Docker. Each component of the capsule — code, data, environment, results, and metadata — can also be downloaded separately either as a zip file or individual files.

Create, organize, and share interactive analysis for everyone

Create and publish analysis Apps for all scientists and researchers
- Self-service, ready-to-use Apps
- Used in Standard Web Browsers
- Easy to Share and Collaborate
- Reuseable, Easy to Manage

Self-service web app allows easy access for bench scientists to execute complex algorithms and visualize results, without touching the code.

Using the Code Ocean Workbench App Panel, computational researchers are able to create interactive analysis apps for any Compute Capsule with zero additional code for the web browser front-end. The App Panel generates and delivers easy to use, point-n-click, self-service Apps to teams of researchers and scientists in a Web Browser and without any IT, interaction with the code or using the command line. This self-service web app allows bench scientists to explore code and data, and leverage the complex algorithms developed by their computational colleagues. In this way, any researcher can independently rerun the analysis using new parameters, without touching the code.
The platform automates all the required compute infrastructure resources to optimize use of tools, cloud compute, and storage

Computational researchers might need to spend precious time performing (or waiting for someone with the cloud or IT engineering skills to perform) operational tasks like infrastructure and environment setup, configuration, and data management. The Code Ocean platform takes care of tedious resource management, such as selecting, launching, switching, and terminating compute machines. It enables fast setup and onboarding of researchers, and also keeps all tools current with the latest security and other updates.

The Code Ocean platform offers scalable computing with easy access to exactly the capacity researchers need. Users can execute projects using cloud computing resources such as AWS Elastic Compute Cloud (EC2), Simple Storage Service (S3), and Elastic File System (EFS). The platform allocates each project with the needed CPU cores, GPUs, memory, and storage on any cloud instance type.

Easily switch between compute resources
Ordinarily, compute resource selection takes many steps to do, even for an experienced engineer, and can be a nightmare for the average researcher. The Code Ocean platform makes it simple. Compute Capsules run on a selected cloud instance, and researchers can easily adjust the compute resources needed for the computation with a few clicks in the Computational Workbench.

Scale compute resources automatically
The Code Ocean platform enables smart cost management using automated resource allocation and cost transparency. The platform works behind the scenes to provision and deprovision compute and storage resources and optimize their use. Compute resources used by compute capsules are automatically scaled out when the demand of compute power exceeds the current machine capacity. Similarly, when capsules are idling but not shut down, the Code Ocean platform’s autoscaling mechanism is triggered to scale in the cloud instances, which reduces costs.
Centrally organize, share, and preserve assets in a private Code Ocean Hub

Provides an easy-to-use, secure place for teamwork

- Easily invite collaborators, teamwork, and project sharing
- Share and reuse Compute Capsules, data, results
- Role-based
- Advanced search & tagging

The Code Ocean Hub is a foundation for “FAIRifying” computational research data with an easy-to-use application and repository for organizing, publishing, sharing, and reusing Code Ocean Capsules, project datasets, and project results. The Code Ocean Hub works seamlessly with the Computational Workbench creating a collaborative experience for the individual researcher and the entire research and development team. The Hub is available in three different forms, for organizations in a Virtual Private Cloud, for the community of academic researchers with the Code Ocean SaaS offering, and available for publishers to incorporate into submissions process via a Hub publishing API.

Secure assets in the Hub repository

- Integrated Git and any service
- Managed input data, results and environments
- Version control
- Architected for FAIR Principles

Findable Accessible Interoperable Reusable
Openly publish and find research in the public Code Ocean Capsule Repository

Code Ocean hosts an open, public Capsule Repository for the global community of computational research. Individual researchers have access to thousands of research results preserved and organized for fast access, sharing, and reuse. The repository is integrated with the Code Ocean Computational Workbench to make viewing and navigating capsule metadata, code, data, environment, and the results simple and intuitive providing a valuable open environment for “FAIRifying” computational research data. Cloning previous research projects allows researchers to tweak and build on top of existing work by directly modifying code or data without wasting time on setup. Users can find and select a capsule from the repository and launch it at any time within the platform.

Cloud hosted, Open Repository, Exportable with No Lock-in

- Easily find and reuse 1000’s of research capsules
- Seamlessly navigate for associated metadata, code, data, environment
- Verification for guaranteed reproducibility
- Architected for FAIR Principles
- Preservation with the CLOCKSS network and DOI minting

FINDABLE
ACCESSIBLE
INTEROPERABLE
REUSABLE

Secure Deployment keeps projects and data behind your organization’s firewall

Code Ocean is deployed on the AWS cloud with a dedicated VPC. The Code Ocean platform provides self-service provisioning of all required AWS resources including scaling EC2 compute, S3 storage, and scheduling backups. With automated backups and available restore scripts users can address disaster recovery requirements. The Code Ocean VPC deployment allows companies to have the entire Code Ocean platform deployed to a company private cloud for security, privacy, and reproducibility in one place. This virtual private cloud gives you all the benefits of the platform while working with sensitive data not to be shared with third parties.

- AWS Virtual Private Cloud (VPC) deployment
- Provisions all required AWS resources in client’s AWS account
- Monitoring and usage reporting for customer tracking and analytics
- Data governance control
- Single Sign On (SSO)
An open, robust research platform ecosystem

All popular and open source tools, environments, languages and packages are seamlessly supported by the Code Ocean research experience.

Any Open Source Languages

Interactive Development Environments

Data Visualization Tools

Any Cloud Data Sources

Any Cloud Computing Resource

Any Computational Packages
Cromwell, Kallisto, Custools, Salmon, Bowtie, Cufflinks, Tophat, STAR and much more

For more information:
www.ebsco.com | information@ebsco.com | (800) 653-2726