

Empowering Reproducible Open Science through the Cloud Computing Platform (CCP) by D4Science

Thursday, 3 October 2024 09:20 (20 minutes)

Abstract

The Cloud Computing Platform (CCP), developed under the aegis of D4Science [1], an operational digital infrastructure initiated 18 years ago with funding from the European Commission, represents a significant advancement in supporting the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, open science, and reproducible data-intensive science. D4Science has evolved to harness the “as a Service” paradigm, offering web-accessible Virtual Laboratories [2] that have also been instrumental in facilitating science collaborations [3]. These laboratories simplify access to datasets whilst concealing underlying complexities, and include functionalities such as a cloud-based workspace for file organisation, a platform for large-scale data analysis, a catalogue for publishing research results, and a communication system rooted in social networking practices.

At the core of the platform for large-scale data analysis, CCP promotes widespread adoption of microservice development patterns, significantly enhancing software interoperability and composability across varied scientific disciplines. CCP introduces several innovative features that streamline the scientific method lifecycle, including a method importer tool, lifecycle tracking, and an executions monitor with real-time output streaming. These features ensure that every step—from creation, through execution, to sharing and updating—is meticulously recorded and readily accessible, thus adhering to open science mandates. CCP supports a broad range of programming languages through automatic code generation, making it effortlessly adaptable to diverse scientific requirements. The robust support for containerisation, utilising Docker, simplifies the deployment of methods on scalable cloud infrastructures. This approach not only reduces the overhead of traditional virtualisation but also enhances the execution efficiency of complex scientific workflows. The platform’s RESTful API design further facilitates seamless interactions between disparate software components, promoting a cohesive ecosystem for method execution and data analysis.

Significantly, CCP embodies the principles of Open Science by ensuring that all scientific outputs are transparent, repeatable, and reusable. Methods and their executions are documented and shared within the scientific community, enhancing collaborative research and enabling peers to verify and build upon each other’s work. The platform’s design also includes comprehensive provenance management, which meticulously tracks the origin and history of data, thus providing a record for scientific discoveries.

CCP serves as a platform for large-scale data analysis of the (i) EOSC Blue-Cloud2026 project VRE, which by leveraging digital technologies for ocean science, utilises CCP to perform large-scale collaborative data analytics, significantly benefiting from CCP’s robust, scalable cloud infrastructure and tools designed for extensive data processing and collaboration, and of the (ii) SoBigData Research Infrastructure that, with its focus on social data mining and Big Data analytics, integrates CCP to facilitate an ecosystem for ethical, scientific discoveries across multiple dimensions of social life.

Keywords: Open Science, Cloud Computing, FAIR Principles, Reproducibility, Data-intensive Science, Containerisation, Microservices

Topic

EOSC Developments and Open Science: Reproducible Open Science

Primary authors: Dr ASSANTE, Massimiliano (CNR); Mr DELL’AMICO, Andrea (CNR); Mr LETTERE, Marco (Nubisware S.r.l.); PAGANO, Pasquale (CNR); Mr PANICHI, Giancarlo (CNR)

Presenters: Dr ASSANTE, Massimiliano (CNR); Mr LETTERE, Marco (Nubisware S.r.l.)

Session Classification: Reproducible Open Science: making research reliable, transparent and credible