



Contribution ID: 11

Type: Oral presentation

DARE: Integrating solutions for Data-Intensive and Reproducible Science

Wednesday, 16 October 2019 09:50 (20 minutes)

The DARE (Delivering Agile Research Excellence on European e-Infrastructures) project is implementing solutions to enable user-driven reproducible computations that involve complex and data-intensive methods. Technology developed in DARE enables Domain Experts, Computational Scientists and Research Developers to compose, use and validate methods that are expressed in abstract terms. Scientists' workflows translates to concrete applications that are deployed and executed on cloud resources offered by European and international e-infrastructures, as well as in-house institutional platforms and commercial providers. The platforms' core services enable researchers to visualise the collected provenance data from runs of their methods for detailed diagnostics and validation, in support of long-running research campaigns involving multiple runs. Use cases are presented by two scientific communities in the framework of EPOS and IS-ENES, conducting research in computational seismology and climate-impact studies respectively. DARE enables users to develop their methods within generic environments, such as Jupyter notebooks, associated with conceptual and evolving workspaces, or via the invocation of OGC WPS services interfacing with institutional data archives. We will show how DARE exploits computational facilities adopting software containerisation and infrastructure orchestration technologies (Kubernetes). These are transparently managed via the DARE API, in combination with registries describing data, data-sources and methods. Ultimately, the extensive adoption of workflows (dispel4py, CWL), methods abstraction and containerisation, allows DARE to dedicate special attention to portability and reproducibility of scientific progress in different computational contexts. We will show how choices of research developers as well as the effects of the execution of their workflows are captured and managed, which enable validation, monitoring and reproducibility. We will discuss the implementation of the provenance mechanisms that adopt workflow's provenance types, lineage services (S-ProvFlow) and PROV-Templates, to record and interactively use context-rich provenance information in W3C PROV compliant formats.

Primary authors: Dr SPINUSO, Alessandro (KNMI); Dr KLAMPANOS, Iraklis (NCSR Demokritos); Dr PAGÉ, Christian (CERFACS); Dr ATKINSON, Malcolm (University of Edinburgh)

Presenter: Dr SPINUSO, Alessandro (KNMI)

Track Classification: Workshop: Building reproducible workflows for earth sciences