

Data Repositories and Science Gateways for Open Science

Wednesday, 11 November 2015 16:00 (20 minutes)

The steep decrease of costs of large/huge-bandwidth Wide Area Networks has fostered in the recent years the spread and the uptake of the Grid Computing paradigm and the distributed computing ecosystem has become even more complex with the recent emergence of Cloud Computing.

All these developments have triggered the new concept of e-Infrastructures which are being built since several years both in Europe and the rest of the world to support diverse multi-/inter-disciplinary Virtual Research Communities (VRCs) and their Virtual Research Environments (VREs).

E-Infrastructure components can indeed be key platforms to support the Scientific Method, the “knowledge path” followed every day by scientists since Galileo Galilei. Distributed Computing and Storage Infrastructures (local High Performance/Throughput Computing resources, Grids, Clouds, long term data preservation services) are ideal both for the creation of new datasets and the analysis of existing ones while Data Infrastructures (including Open Access Document Repositories –OADRs –and Data Repositories –DRs) are essential also to evaluate existing data and annotate them with results of the analysis of new data produced by experiments and/or simulations. Last but not least, Semantic Web based enrichment of data is key to correlate document and data, allowing scientists to discover new knowledge in an easy way.

However, although big efforts are being done in the last years, both at technological and political level, Open Access and Open Education are still far from being pervasive and ubiquitous and prevent Open Science to be fully established. One of the main drawbacks of this situation is the limiting effect it has on the reproducibility and extensibility of science outputs which are, since more than four centuries, two fundamental pillars of the Scientific Method.

In this contribution we present the Open Access Repository (OAR), a pilot data preservation repository of INFN and other Italian Research Organisations’ products (publications, software, data, etc.) meant to serve both researchers and citizen scientists and to be interoperable with other related initiatives both in Italy and abroad. OAR is powered by the INVENIO software and is both an Open Access Initiative conforming and an official OpenDOAR data provider, able to automatically harvest resources from different sources, including the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3), using RESTful API’s. It is also one of the official OpenAIRE archives, compliant with version 3.0 of its guidelines.

OAR allows SAML-based federated authentication and it is one of the Service Providers of the eduGAIN inter-federation; it is also connected to DataCite for the issuance and registration of Digital Object Identifiers (DOIs). But what makes OAR really different from other repositories is its capability to connect to Science Gateways and exploit Distributed Computing and Storage Infrastructures worldwide, including EGI and EUDAT ones, to easily reproduce and extend scientific analyses. In this presentation some concrete examples related to the data of the ALEPH and ALICE Experiments will be shown.

Links, references, publications, etc.

<http://www.openaccessrepository.it>

<http://science-gateway.chain-project.eu>

Additional information

The authors wish to thank Sunje Dallmeier-Tiessen (CERN), Giuseppe Fiameni (Cineca), Paola Gargiulo (Cineca), Maurizio Lancia (CNR), Giacomo Mariani (Cineca), Salvatore Mele (CERN), Laura Rueda (CERN), Massimiliano Saccone (CNR), Alberto Salvati (CNR), Tim Smith and Tibor Simko on behalf of the whole INVENIO Team at CERN, Heinrich Widmann (DKRZ) and Wojciech Ziólek (CERN) for their help with the deployment of the Open Access Repository, its interfacing to CNR, EUDAT and SCOAP3 services and the implementation of the ALEPH use case.

Primary authors: CARRUBBA, Carla (University of Catania); MENASCE, Dario (INFN Milano Bicocca); INSERRA, Giuseppina (University of Catania); MAGGI, Marcello (INFN Bari); RICCERI, Rita (University of Catania); BARBERA, Roberto (University of Catania and INFN); BIANCO, Stefano (INFN LNF); BOCCALI, Tommaso (INFN Pisa)

Presenter: BARBERA, Roberto (University of Catania and INFN)

Session Classification: Data without boundaries: metadata interoperability