



Contribution ID: 54

Type: **Presentation short (15 min)**

On demand data analysis tools for the EOSC Resource Catalogue

Tuesday, 19 October 2021 14:25 (15 minutes)

OpenAIRE (www.openaire.eu) aims to establish an open and sustainable scholarly communication infrastructure responsible for the overall management, analysis, manipulation, provision, monitoring and cross-linking of all research outcomes. One of the contributions of OpenAIRE to the European Open Science Cloud (EOSC) is its research graph (<https://graph.openaire.eu>), one of the largest open scholarly record collections worldwide, which constitutes the EOSC Resource Catalogue. Conceived as a public and transparent good, populated out of data sources trusted by scientists, the EOSC Resource Catalogue aims at bringing discovery, monitoring, and assessment of science back in the hands of the scientific community.

It includes metadata records and links (i) collected from 70k+ scholarly communication sources from all over the world, including Open Access institutional repositories, data archives, journals; (ii) inferred by data mining algorithms, and (iii) provided by users of the OpenAIRE portals thanks to the Link functionality.

OpenAIRE is already using the catalogue to power its portals and to support researchers, funders, organisations, research communities and infrastructures at discovering and tracking research products. The catalogue is also openly accessible via data dumps and APIs (<https://develop.openaire.eu>), so that its content can be used by any researchers for their own research activities.

In order to ease the usage and analysis of the EOSC Resource Catalogue, this presentation proposes an integration with the EGI notebooks on demand (<https://marketplace.egi.eu/44-notebooks>).

EGI Notebooks is a browser-based tool for interactive analysis of data using EGI storage and compute services based on the JupyterHub technology. Notebooks are offered on-demand to single researchers or research communities.

The idea is to offer EGI notebooks capable of requesting “slices” of the EOSC Resource Catalogue that are relevant for the end user and support the definition of functions for the analysis of such slices.

Thanks to the EGI Notebook, the end users will be able to analyse the EOSC Resource Catalogue on top of the stable and scalable infrastructure of EGI.

For example, if the end users want to analyse the outputs of H2020 projects, the notebook will download the dump of H2020 research outputs from Zenodo (<https://doi.org/10.5281/zenodo.4559725>) and execute a number of predefined functions to give an overview of the data to the end-users.

The EGI notebook should also be able to address requests for slices of the catalogue that are not already published on Zenodo by OpenAIRE. In such cases, the users should be able to define the selection criteria that will be applied to the full dump (<https://doi.org/10.5281/zenodo.3516917>) or used to build a query to the OpenAIRE or EOSC discovery API.

Speaker:

Michele De Bonis is born in Italy. Computer Engineer and PhD student at ISTI-CNR aiming to study Deep Learning techniques for graph processing. Currently responsible for the implementation and the maintenance of the OpenAIRE services.

Most suitable track

Innovating services together

By submitting my abstract, I agree that my personal data is being stored in accordance to conference Privacy Policy

Primary authors: DE BONIS, Michele (CNR-ISTI); BARDI, Alessia (CNR-ISTI); MANGHI, Paolo (OpenAIRE)

Presenter: DE BONIS, Michele (CNR-ISTI)

Session Classification: Innovating Services Together - Presentations