OpenBio.eu: An extrovert bioinformatics research object repository and workflow management system

Alexandros Kanterakis[1], Lefteris Koumakis[2], Manolis Koutoulakis[2], Konstantina Pitianou[2], Nikos Kanakaris[3], Nikos Karacapilidis3, George Potamias2

[1] Foundation for Research and Technology, Heraklion, Greece, Email: kantale@ics.forth.gr

[2] Foundation for Research and Technology, Heraklion, Greece

[3] IMIS Lab, MEAD, University of Patras, Greece

Despite the plethora of Workflow Management Systems (WMS), guidelines (i.e FAIR), source code repositories and Open Access Journals, reproducibility is still a major issue in bioinformatics and computational biology. This is mainly due to the unwillingness of authors to share ROs (Research Objects, i.e. Tools, Data, Workflows) and the introvert nature of modern WMSs. By introvert, we refer to the difficulty to interconnect ROs that belong to different WMSs (i.e. Galaxy, Nextflow and Snakemake), a task that requires above average IT skills. Moreover, ROs are still treated as components that are tightly coupled with their corresponding WMSs, whereas important social features like discussion, rating and crowdsourced editing are poorly supported.

In this work, we present the enhanced version of OpenBio.eu, an environment that treats ROs as independent "first class citizens". Namely, Data, Tools and Workflows are actual executable software components that contain the BASH commands that download, install, and run them without requiring any other dependency. The environment facilitates the indexing and combination of these ROs into new ROs without introducing any new DSL (Domain Specific Language). Exporting/importing to/from existing WMSs is a basic feature and a primary focus of development. Currently, we support Workflow export/import to Common Workflow Language (CWL) and we also support export in Airflow format. In OpenBio.eu, users are incentivised to import their ROs by taking credit when others are using them. Importing ROs is as easy as installing them in a PC. Users can directly download, execute, rate and comment on any RO. They can also "Fork" it and create a personal version that can edit as they wish. Similarly, users can compile Workflows by Drag and Dropping ROs in a graph. Workflows can be nested indefinitely, whereas conditional execution and iteration is naturally supported. We also present an execution engine that tights a resource manager (NetData) and a Workflow Execution Engine (Airflow) into a Docker container. Users can "plug" multiple execution engines and manage them through their personal profile page. OpenBio.eu also includes a knowledge graph environment that visualizes the discourse conducted on ROs and can help researchers choose the appropriate resources for their task.

Overall, OpenBio.eu is a free, extrovert and social environment aiming to maximize reproducibility and enhance the visibility of research in life sciences.