



## Towards Reproducible Bioinformatics: The OpenBio-C Workflow Management Platform

Alexandros Kanterakis<sup>1#</sup>, Galateia Iatraki<sup>1</sup>,  
Konstantina Pityanou<sup>1</sup>, Lefteris Koumakis<sup>1</sup>,  
Nikos Kanakaris<sup>2</sup>, Nikos Karacapilidis<sup>2</sup>, and  
George Potamias<sup>1\*</sup>

<sup>1</sup> Computational BioMedicine Laboratory (CBML), Institute of Computer Science, FORTH

<sup>2</sup> Industrial Management and Information Systems Lab, MEAD, University of Patras

<sup>#</sup>Presenting author: ICS-FORTH, email: [kantale@ics.forth.gr](mailto:kantale@ics.forth.gr)

<sup>\*</sup>Corresponding author: ICS-FORTH email: [potamias@ics.forth.gr](mailto:potamias@ics.forth.gr)

### ABSTRACT

Published and highly-ranked biomedical research results are very often not *reproducible*. We argue that this phenomenon is mainly caused by the separation between scientific reporting and the analytical pipelines used to produce the respective findings, and it is one of the origins of the *reproducibility crisis* [1]. One may even state that “*reproduce or perish*” should become the new motto for contemporary science [2]. This makes it harder to translate scientific discoveries into credible clinical protocols and recommendations, a fact that resembles a “*lost in translation*” situation! [3]. Trapped in such a problematic research landscape, and guided by an important statement that *one-size-fits-all is unlikely to work*, as it is made evident by the decline in popularity of well-known and widely utilized bioinformatics *workflow management systems* / WfMS (e.g., Taverna, Galaxy [4]), the challenge for a new conceptualization of WfMS is raised. In such a conceptualization scientific workflows and pipelines: (a) should be designed as *community building infrastructures* that work best when interlinked with other objects and activities around them; (b) their constituents should comply to the so-called *FAIR* (Findable, Accessible, Interoperable and Reusable) principles; and (c) reproducibility should be *intrinsic* attribute of them.

The aforementioned needs and requirements underlie the motivations for the development of the **OpenBio-C** platforms<sub>s</sub>, and determine its fundamental objectives. OpenBio-C ([www.openbio.eu/platform](http://www.openbio.eu/platform)) is a web-based and open WfMS that: (i) offers graph-based workflow synthesis, execution and monitoring operations by simply dragging and dropping tools and other workflows in an online graph editor; (ii) is equipped with semantically annotated research objects ([www.researchobject.org](http://www.researchobject.org)) repository (i.e., tools, data, workflows) and respective management services; (iii) supports Q&A services for the engaged bioinformatics researchers coupled with a bibliographic reference manager and a user profile builder; and (iv) offers argumentation-based collaborative services enabled by a graph-based discourse visualization component. The platform requires no IT knowledge and supports import / export of tools from a variety of existing WfMS. For the description of tools, “gluing” them together and composing workflows, OpenBio-C utilizes BASH ([www.gnu.org/software/bash](http://www.gnu.org/software/bash)), satisfying two basic conditions regarding interoperability and reproducibility of scientific workflows: (a) no need to “wrap” the installation in any “domain specific language”, and (b) easy import and export to other workflows.

### REFERENCES

- [1] Munafò MR, et al. 2017. *Nat. Hum. Behav.*, 1:21.
- [2] Grimes DR, et al. 2018. *R. Soc. Open Sci.*, 5(1).
- [3] Levin LA and H. V. Danesh-Meyer, HV. 2010. *JAMA*, 303(15):1533–1534.
- [4] Kanterakis A, et al. 2019. *BIBE 2019* (to be published).

§The work presented in this paper is supported by the OpenBio-C project ([www.openbio.eu](http://www.openbio.eu)), which is co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project id: T1EDK- 05275). GitHub: <https://github.com/kantale/OpenBioC>