Portable workflow and tool descriptions with the CWL

Document Version
Final published version

Citation for published version (APA):
Amstutz, P., Tijani, N., Soiland-Reyes, S., Kern, J., Stojanovic, L., Pierce, T., Chilton, J., Mikheev, M., Lampa, S., Ménager, H., Frazer, S., Malladi, V. S., & Crusoe, M. R. (2015). *Portable workflow and tool descriptions with the CWL*. Abstract from Bioinformatics Open Source Conference, Dublin, Ireland. https://github.com/common-workflow-language/2015-bosc/blob/master/bosc_abstract/cwl_abstract_bosc.md

OPEN ACCESS

Title	Portable worflow and tool descriptions with the CWL
Authors	Peter Amstutz[0], Nebojša Tijanić[1], Stian Soiland-Reyes[2], John Kern[3], Luka Stojanovic[1], Tim Pierce[0], John Chilton[4], Maxim Mikheev[5], Samuel Lampa[6,7], Hervé Ménager[8], Scott Frazer[9], Venkat Sai Malladi[10], <u>Michael R. Crusoe[11]</u>
Affiliations	0. Curoverse Inc. 1. Seven Bridges Genomics, Inc.
	2. University of Manchester, School of Computer Science
	3. AccuraGen Inc. 4. Penn State University, The Galaxy Project 5. BioDatomics Inc.
	6. Uppsala University, Department of Pharmaceutical Biosciences
	7. BILS (Bioinformatics Infrastructure for Life Sciences) 8. Institut Pasteur
	9. The Broad Institute 10. Stanford University
	11. Michigan State University, Microbiology and Molecular Genetics
Contact	mcrusoe@msu.edu
URLs	http://common-workflow-language.github.io/
	https://github.com/common-workflow-language/common-workflow-language/
License	Apache License, Version 2.0

Bioinformatics workflow platforms provide provenance tracking, execution and data management, repeatability, and an environment for data exploration and visualization. Example F/OSS bioinformatics workflow platforms include Arvados, Galaxy, Mobyle, iPlant DiscoveryEnvironment, Apache Taverna and Yabi. Each one presently represent workflows using different vocabularies and formats, and adding new tools requires different procedures for each system.

Neither the description of the *workflows* nor the descriptions of the *tools* that power them are usable outside of the platforms they were written for. This results in duplicated effort, reduced reusability, and impedes collaboration.

Three engineers (Peter Amstutz, John Chilton, and Nebojsa Tijanic) from leading bioinformatics platform teams (Curoverse, Galaxy Team, and Seven Bridges Genomics) and a tool author (Michael R. Crusoe / khmer project / Michigan State University) started working together at the BOSC 2014 Codefest with an initial focus on developing a portable means of representing, sharing and invoking command line tools which was then the basis for portable workflow descriptions. The group placed high value on re-using existing formats and ontologies; they governed themselves with a lazy consensus / do-ocracy approach.

On March 31st, 2015 the group released their second draft of the Common Workflow Language specification. The serialized form is a YAML document that is validated by an Apache Avro schema and can be interpreted as an RDF graph using JSON-LD. The documents are also valid Wf4Ever 'wfdesc' descriptions after a simple transformation. Future drafts will include the use of the EDAM ontology to describe the tools enabling discovery via the ELIXIR tool registry.

Seven Bridges Genomics, the Galaxy Project, and the organization behind Arvados (Curoverse) have started to implement support for the Common Workflow Language, with interest from other projects and organizations like Apache Taverna, BioDatomics and the Broad Institute. Developers on the Galaxy Team are exploring adding CWL tool description support with plans to add support for the CWL workflow descriptions. Tool authors and other community members will benefit as they will only have to describe their tool and workflow interfaces once. This will enable scientists, researchers and other analysts to share their workflows and pipelines in an interoperable and yet human readable manner.