

---

<b>Title</b>	Unifying workflow systems with the CWL
<b>Authors</b>	Peter Amstutz <sup>0</sup> , Nebojša Tijanić <sup>1</sup> , Stian Soiland-Reyes <sup>2</sup> , Abram Connelly <sup>0</sup> , John Kern <sup>3</sup> , Luka Stojanovic <sup>1</sup> , Ward Vandewege <sup>[0]</sup> , Tim Pierce <sup>[0]</sup> , John Chilton <sup>[4]</sup> , Maxim Mikheev <sup>[5]</sup> , Samuel Lampa <sup>[6][7]</sup> , Hervé Ménager <sup>[8]</sup> , Scott Frazer <sup>[9]</sup> , Venkat Sai Malladi <sup>[10]</sup> , _Michael R. Crusoe_ <sup>[11]</sup>
<b>Contact</b>	mcrusoe@msu.edu
<b>URLs</b>	<a href="http://common-workflow-language.github.io/">http://common-workflow-language.github.io/</a> <a href="https://github.com/common-workflow-language/common-workflow-language/">https://github.com/common-workflow-language/common-workflow-language/</a>
<b>License</b>	Apache License, Version 2.0

---

[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

Bioinformatic workflow platforms provide provenance tracking, execution and data management, repeatability, and a platform for data exploration and visualization. Example F/OSS bioinformatic workflow platforms are Arvados, DiscoveryEnvironment, DNANexus, Galaxy, Mobyle, and Yabi. Each one represents workflows using a different vocabulary and format and adding a new tool requires a different procedure for each system.

Neither the description of the workflows nor the descriptions of the tools that power them are usable outside of the workflow platforms they were written for. This results in duplicated efforts and reduced reusability of these components.

Three engineers (Peter Amstutz, John Chilton, and Nebojsa Tijanic) from leading open source bioinformatic workflow groups (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 command line tool description format. As the group grew the scope expanded to include 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 Avro schema and can be interpreted using JSON-LD. The documents are also valid Wf4Ever descriptions after a simple transformation. Future drafts will include the use of the EDAM ontology to describe the tools enabling tool discovery via the ELIXIR tool registry.

The popular bioinformatics platform Galaxy is going to be accepting CWL tool and workflow descriptions with planned support for exporting workflows into the CWL format. This will scientists, researchers and other analysts to share their workflows and pipelines in an interoperable and yet human readable manner. Tool authors and other community members will also benefit as they will only have to describe their tool's interface once. Uptake is expected by Curoverse and Seven Bridges Genomics as they had dedicated considerable engineering resources to the project.