Dockstore 2.0: Enhancing a community platform for sharing cloud-agnostic research tools

Denis Yuen1, Brian O’Connor2, Cricket Sloan2, David Steinberg3, Natalie Pfaff2, Walt Shands2, Gary Lu1, Andrew Duncan1, Charles Overbeck2

1 Ontario Institute for Cancer Research, MaRS Centre, Toronto, Ontario 2 UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, USA

Background
Dockstore is a platform that was created in response to the many challenges faced during the Pan-Cancer Analysis of Whole Genomes (Pan-Cancer Analysis of Whole Genomes). It is a repository of scientific workflows described using popular workflow languages with their dependencies distributed in Docker images. This powerful combination of technologies allows for improved sharing and portability of scientific workflows.

Docker is used to describe the environment that the workflow will run in along with all of the dependencies required.

Workflow languages are used to describe the steps involved in running a workflow and how they are dependent on each other, including all inputs and outputs. We currently support CWL, WDL and Nextflow. These descriptor documents can be stored on GitHub, Bitbucket, GitLab, or directly on dockstore.org.

Benefits
- Searchable catalogue of tools and workflows accessible via a proposed GA4GH standard API
- Launch workflows locally and in a variety of cloud platforms
- Create organization landing pages for your team, lab, grant, or institution
- Social features such as starring, labels, and discussion threads

Community
Dockstore is thankful to its many contributors, users, and partners. Here we present a few to give a sense of what is going on in this space.

Highlighted New Features
Dockstore 1.7.0 (projected)
- Services Prototype: experimental support for long-lived services such as genome browsers and reference data servers for running workflows
- Easy Registration Prototype: registration using GitHub Apps, automatically keep repos up to date via GitHub hooks
- Immutable and DOI issuing: freeze workflow releases and issue DOIs for them using Zenodo
- Improved support for CWL, WDL
  - testing with a toolset supporting CWL 1.1
  - testing with WDL 1.0 parsing libraries
- Display testing logs for verified workflows

Dockstore 1.6.0 (2019-04)
- Dockstore Organizations: Create landing pages to describe and group workflows based on institution, grant, theme, etc.
- Beta WES Client: launch a Dockstore workflow on any GA4GH WES-compatible platform
- Better CWL, WDL, and Nextflow Language Support:
  - Nextflow parsing improvements
  - owtIoI and Cromwell compatibility update

Dockstore 1.5.0 (2018-09)
- Hosted Tools/Workflows: also store tool/workflow descriptors and test parameter files directly on Dockstore.org
- Nextflow Support: register, search, and visualize Nextflow workflows
- DRS File Plugin: provision files via the GA4GH-DRS standard for our CLI

Future Work
Looking for collaborators / testers
- Alternative Containerization Support: Singularity and/or uDocker support
- Additional Workflow Languages: Support for more languages and platforms
- Signing of entries on Dockstore: Verify ownership and integrity of Docker images
- GI Environment: Automated testing for workflows across cloud platforms

References
- Arnoldiz, Peter; Croese, Michael R; Tjandra, Neljouja; Chapman, Ben; Gillson, John; Hauer, Michael; Karamuzov, Andrey; Lefer, Dan; Ménager, Henri; Nedeljkovic, Maya; Seides, Matt; Soland-Reyes, Kevin; Stojanovic, Luka (2016): Common Workflow Language, v1.0. F1000Research doi:10.7490/f1000research.1114631.1

Run workflows on a variety of cloud platforms using a web browser

Community
- CWL
- WDL
- Nextflow

Dockstore Organizations
- Create landing pages to describe and group workflows based on institution, grant, theme, etc.

CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

The CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

The r-f-core series of high quality Nextflow workflows has been registered on Dockstore allowing users to see them alongside WDL and CWL workflows

Other highlighted contributors to Dockstore include workflows from the Cancer IT (Sanger Institute), the Broad Institute, the TOPMed (Trans-Omics for Precision Medicine) program and DataSTAGE (Storage, Toolspace, Access and analytics for big data empowerment).

Additional Workflow Languages

CWL and WDL
- Improved support for CWL, WDL
  - testing with a toolset supporting CWL 1.1
  - testing with WDL 1.0 parsing libraries
- Display testing logs for verified workflows

CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

The CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

The CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

The CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

Dockstore 1.7.0 (projected)
- Services Prototype: experimental support for long-lived services such as genome browsers and reference data servers for running workflows
- Easy Registration Prototype: registration using GitHub Apps, automatically keep repos up to date via GitHub hooks
- Immutable and DOI issuing: freeze workflow releases and issue DOIs for them using Zenodo
- Improved support for CWL, WDL
  - testing with a toolset supporting CWL 1.1
  - testing with WDL 1.0 parsing libraries
- Display testing logs for verified workflows

Dockstore 1.6.0 (2019-04)
- Dockstore Organizations: Create landing pages to describe and group workflows based on institution, grant, theme, etc.
- Beta WES Client: launch a Dockstore workflow on any GA4GH WES-compatible platform
- Better CWL, WDL, and Nextflow Language Support:
  - Nextflow parsing improvements
  - owtIoI and Cromwell compatibility update

Dockstore 1.5.0 (2018-09)
- Hosted Tools/Workflows: also store tool/workflow descriptors and test parameter files directly on Dockstore.org
- Nextflow Support: register, search, and visualize Nextflow workflows
- DRS File Plugin: provision files via the GA4GH-DRS standard for our CLI

Future Work
Looking for collaborators / testers
- Alternative Containerization Support: Singularity and/or uDocker support
- Additional Workflow Languages: Support for more languages and platforms
- Signing of entries on Dockstore: Verify ownership and integrity of Docker images
- GI Environment: Automated testing for workflows across cloud platforms

References
- Arnoldiz, Peter; Croese, Michael R; Tjandra, Neljouja; Chapman, Ben; Gillson, John; Hauer, Michael; Karamuzov, Andrey; Lefer, Dan; Ménager, Henri; Nedeljkovic, Maya; Seides, Matt; Soland-Reyes, Kevin; Stojanovic, Luka (2016): Common Workflow Language, v1.0. F1000Research doi:10.7490/f1000research.1114631.1

This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-NGS). Funding for the Ontario Institute for Cancer Research is provided by the Government of Ontario. Additional in-kind assistance is provided by the UC Santa Cruz Genomics Institute.