



HAL
open science

A curated Domain centric shared Docker registry linked to the Galaxy toolshed

François Moreews, Olivier Sallou, Yvan Le Bras, Grosjean Marie, Cyril Monjeaud, Thomas A Darde, Olivier Collin, Christophe Blanchet

► **To cite this version:**

François Moreews, Olivier Sallou, Yvan Le Bras, Grosjean Marie, Cyril Monjeaud, et al.. A curated Domain centric shared Docker registry linked to the Galaxy toolshed. Galaxy Community Conference 2015, Jul 2015, Norwich, United Kingdom. hal-01160514

HAL Id: hal-01160514

<https://inria.hal.science/hal-01160514>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A curated Domain centric shared Docker registry linked to the Galaxy toolshed

François Moreews¹, Olivier Sallou², Yvan le Bras², Marie Grosjean³,
Cyril Monjeaud², Thomas Darde⁴, Olivier Collin², Christophe Blanchet³

1 Genscale team -IRISA -Rennes, France

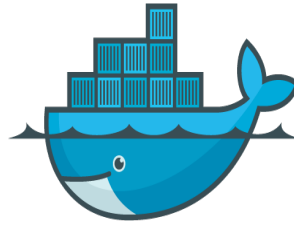
2 Genouest Bioinformatics facility – INRIA/IRISA – Rennes, France

3 French Institute of Bioinformatics – CNRS IFB-Core UMS3601 – Gif-sur-Yvette, France

4 INSERM U625 – Rennes France



Docker : presentation



“**Docker** is an open-source engine to easily create **lightweight, portable, self-sufficient containers** from any application. The same container that a developer **builds** and **test** on a laptop can run at **scale**, in **production**, on Vms,[...], public **clouds** and more.”

Docker : presentation

Why using Docker containers to build, deploy and execute applications ?

- **Efficient** (no virtualization)
- **Isolation**
- **Build one time, execute “anywhere”**, independently of the execution platform (laptops, clusters and clouds with linux kernels)

...

Build : Dependencies & Dockerfile

```
FROM ubuntu:12.0.4
```

```
ADD . /script
```

```
ENTRYPOINT ["perl /script/analysis.pl"]
```

Run Docker

```
docker -run  
  containerUniqueID  
  -k=31 -i input.fastq -o output.bam
```

-The docker run command acts as a wrapper of the tool command line.

-Host directories (input, output,work...) can be mounted inside the container.

Google Container Engine

- A container based cloud architecture
- “With container-based computing, application developers can focus on their application code, instead of on deployments and integration into hosting environments”.

Docker on academic HPC clusters

- **Google Kubernetes** : an open source technology for containers life cycle management.
- **Docker Swarm** : allows to create and access to a pool of Docker hosts.
- **Genouest GO-DOCKER** : a batch scheduler like SGE, submitting jobs in Docker containers on top of Swarm..

Bioinformatics tools benchmarks with Docker

- cami-challenge.org : Critical Assessment of **Metagenomic** Interpretation
- <http://nucleotid.es> : continuous, objective and reproducible evaluation of genome **assemblers** using docker containers
- bioboxes.org : **interchangable** bioinformatics software containers

Galaxy Docker integration

- Docker can be used in Galaxy to :
 - manage tools dependencies : one tool , on Docker
 - Distribute populated Galaxy Distribution related to one topic

Shared registries : Docker Hub

- Not structured
- Not curated
- Not domain centric
- Not community driven

Shared registries : BIOSHADOCK

BIOSHADOCK

An initiative of the French Bioinformatics Institut & the Genouest Bioinformatics Facility

Goals :

- Federate bioinformatics tools deployment procedures for the IFB cloud infrastructure
- Generate customized Galaxy cloud instances on the fly.
- Docker image indexation (service registry & searches)

Shared registries : BIOSHADOCK

BioShaDock registry

A **Bioinformatics Shared Docker** registry

 Browse BioShaDock

BioShaDock is a Bioinformatics Docker registry.

Here are hosted Docker images dedicated to a broad spectrum of Biological communities as represented by the [Biogenouest](#) Western France network.

In particular, you will find here :

-Command line tools ;

-Complexe web server frameworks.

-Galaxy Docker images that you can use with specific Docker Galaxy tools thanks to recent developments through the e-Biogenouest project (<https://www.e-biogenouest.org/>), (see [GUGGO](#) and our [Toolshed](#));

The number of images available will grow following the community needs.

The Docker GenOuest core facility team :

François Moreews, Cyril Monjeaud, Yvan Le Bras, Olivier Sallou



Shared registries : BIOSHADOCK



Home / Images

Images

Search for an image

Images (34/34)

- abyss
- bio-linux
- biojava
- bioperl
- biopython
- blast
- bowtie
- bwa
- centos
- centos7
- chado_database
- clustal-omega
- comet
- discosnp
- emboss
- idba
- ipython-notebook
- megahit
- minia

docker-ui.genouest.org/#/tag/library/stacks_galaxy/latest/5f778585c808b968fe865159eda886a48f5e1bf82



Home / Images / stacks_galaxy / latest

Details for tag library / stacks_galaxy : latest

want to use this image ?

#pull command :

```
docker pull docker-registry.genouest.org/stacks_galaxy:latest
```

#run test command:

```
docker run -i -t docker-registry.genouest.org/stacks_galaxy:latest /bin/bash
```

remark: if the docker client generates an ssl certificate error you can solve this issue adding on option "--insecure-registry docker-registry.genouest.org" in the docker daemon command line. modify the docker daemon command line adding the following option :
DOCKER_OPTS="--insecure-registry docker-registry.genouest.org" the docker daemon command should be something like "docker -d --insecure-registry docker-registry.genouest.org"
To do this on Ubuntu, modify /etc/init/docker.conf, on Debian , modify /etc/init.d/docker. Do not forget to restart the service : (service docker restart)

Image Details

General information

[Image Ancestry](#)

Author

Comment

add STACKS .py wrappers for Galaxy on /usr/bin/

Created

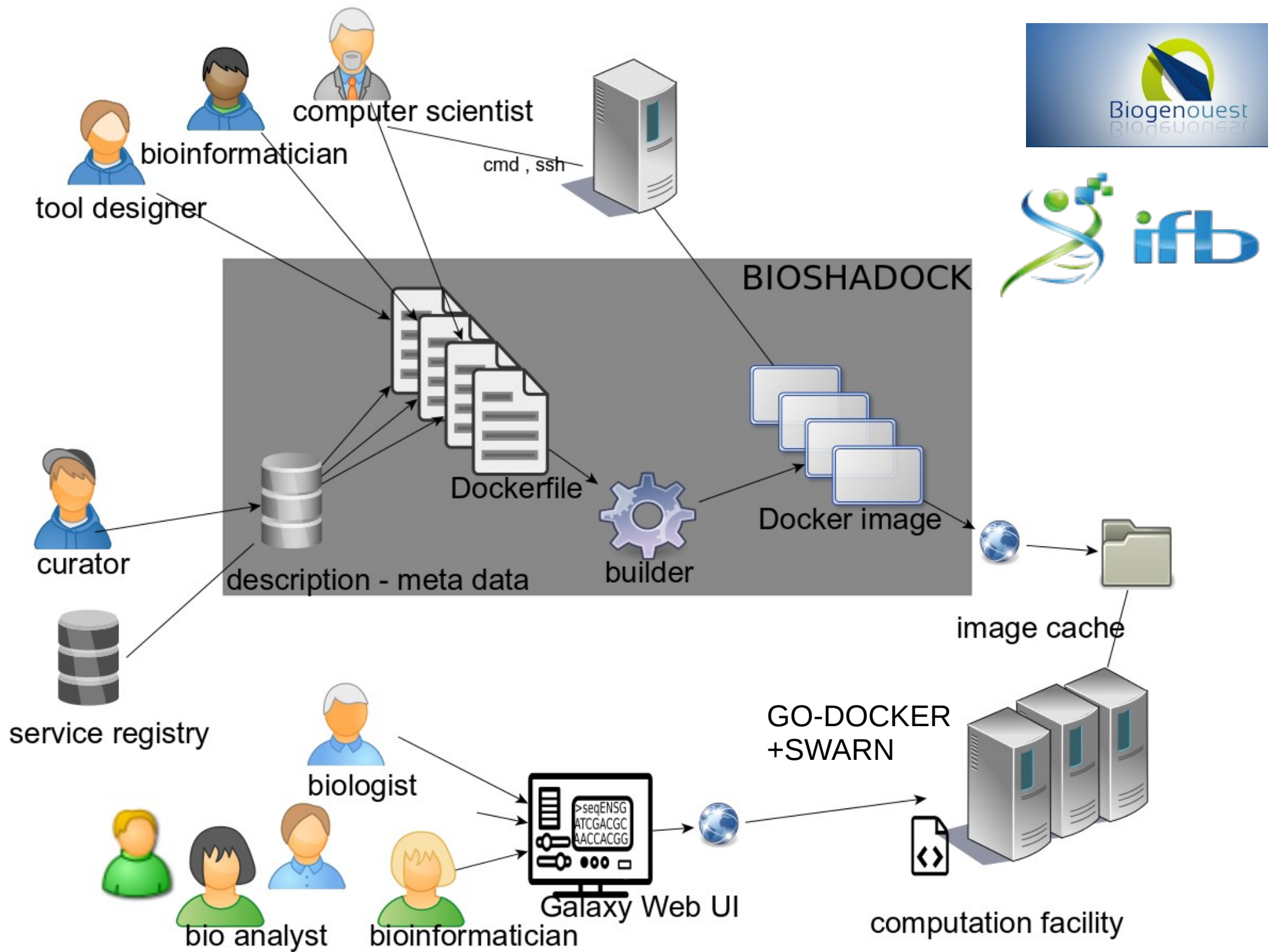
18 days ago (2014-11-07 10:45:19 +0100)

Docker version

1.3.0

Shared registries : BIOSHADOCK

- BIOSHADOCK
 - Focuses on the model “on tool, one docker image”
 - Allows Dockerfile build
 - Manages permissions (private/ public images)
 - May integrate meta data to facilitate query and service registry searches
 - One unique repository for softwares with or without tool.xml => SAAS + CMD
 - Integrated to Galaxy by redefining tools dependencies in a Toolshed



BIOSHADOCK TOOLSHED integration

The screenshot displays the Galaxy Tool Shed interface. At the top, the header includes the Galaxy Tool Shed logo, the text "164 valid tools on Jul 06, 2015", and navigation links for "Repositories", "Help", and "User". The main content area is titled "Repositories by Category" and features a search input field with the placeholder text "search repository name, description". Below this is a table with two columns: "Name" and "Description". The table lists various categories such as "Alignment", "Data management", "Data manager", "Data source", "Fasta / Fastq manipulation", "France Genomique Assembly", "Metadata management", "Metagenomics", "Microarray", "Next Generation Sequencing", and "Phylogeny". A sidebar on the left contains navigation links for "Search", "Valid Galaxy Utilities", "All Repositories", and "Available Actions". The browser address bar at the bottom shows "toolshed.genouest.org".

Galaxy Tool Shed Repositories Help User

164 valid tools on Jul 06, 2015

Search

- Search for valid tools
- Search for workflows

Valid Galaxy Utilities

- Tools
- Custom datatypes
- Repository dependency definitions
- Tool dependency definitions

All Repositories

- Browse by category

Available Actions

- Login to create a repository

Repositories by Category

Name	Description
Alignment	Alignment tools
Data management	Tools for data management
Data manager	Biological data management
Data source	data source
Fasta / Fastq manipulation	Fasta / Fastq manipulation
France Genomique Assembly	Tools for NGS assembly as part of France Genomique Project
Metadata management	Tools for ISA manipulation
Metagenomics	NGS metagenomics analysis tools
Microarray	Analyse gene expression data through microarray technology
Next Generation Sequencing	NGS analysis tools
Phylogeny	phylogeny tools

toolshed.genouest.org

BIOSHADOCK cluster /cloud integration using GO-DOCKER

Go-Docker

Home

Resources

Login

Go-Docker

0

Containers pending

0

Containers running

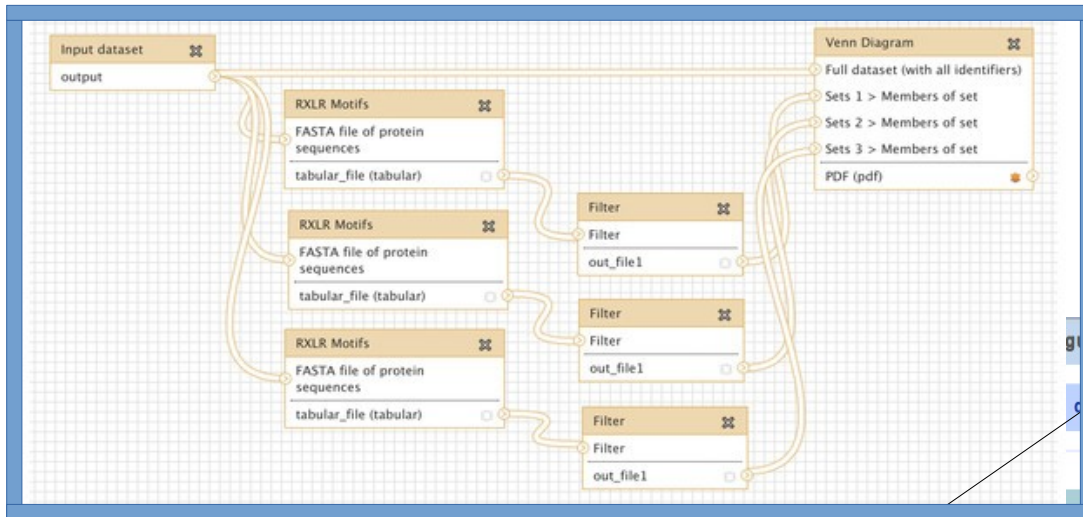
164

Containers to date



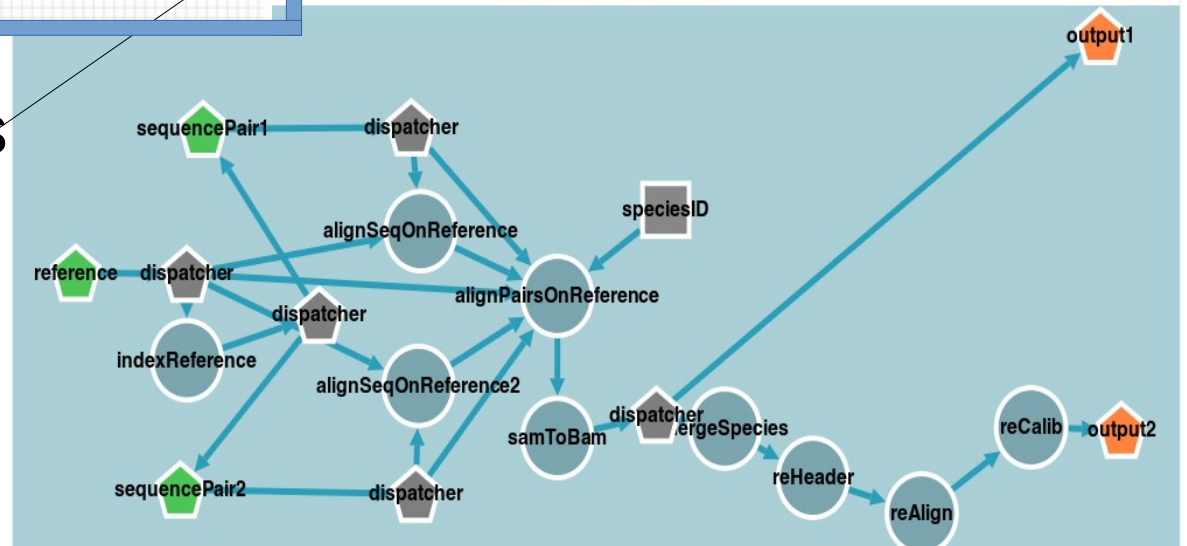
Build it one time, use it as you want

Command lines



Job-ID	Prior	Name	User	State	Submit/Start	CPU%
321	0	Slingshot	deadline	r	04/25/2005 13:16:17	1
323	0	Slingshot2	deadline	t	04/25/2005 19:17:07	2
324	0	Slingshot2	deadline	qu	04/25/2005 19:15:55	0
325	0	Slingshot4	deadline	qu	04/25/2005 19:16:09	0
326	0	P12	deadline	qu	04/25/2005 19:16:22	0
327	0	P14	deadline	qu	04/25/2005 19:16:27	0
328	0	P14	deadline	qu	04/25/2005 19:16:29	0
329	0	P14	deadline	qu	04/25/2005 19:16:31	0
331	0	Slingshot2	deadline	qu	04/25/2005 19:16:48	0
332	0	Slingshot4	deadline	qu	04/25/2005 19:16:52	0

Galaxy tools & workflows



BIOSHADOCK

Other SAAS tools

References

- Genouest GO-DOCKER : <http://www.genouest.org/?p=246>
- Google Kubernetes, Docker container cluster management : kubernetes.io
- BioShaDock, a Bioinformatics Shared Docker registry : <http://docker-ui.genouest.org>
- GUGGO Galaxy Tooshed : <http://toolshed.genouest.org>
- Nucleotid.es, continuous, objective and reproducible evaluation of genome assemblers using docker containers : <http://nucleotid.es>
- ELIXIR Tools and Data Services Registry : <https://elixir-registry.cbs.dtu.dk>
- Bioboxes, a standard for creating interchangeable bioinformatics software containers : <http://bioboxes.org>
- IFB academic Cloud : <http://www.france-bioinformatique.fr/?q=en/core/e-infrastructure-team/ifb-cloud>