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

François Moreews[1], Olivier Sallou[2], Yvan le Bras[2], Marie Grosjean[3],
Cyril Monjeaud[2], Thomas Darde[4], Olivier Collin[2], Christophe Blanchet[3]

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 virttualization)
- **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.

Docker on the commercial Cloud

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

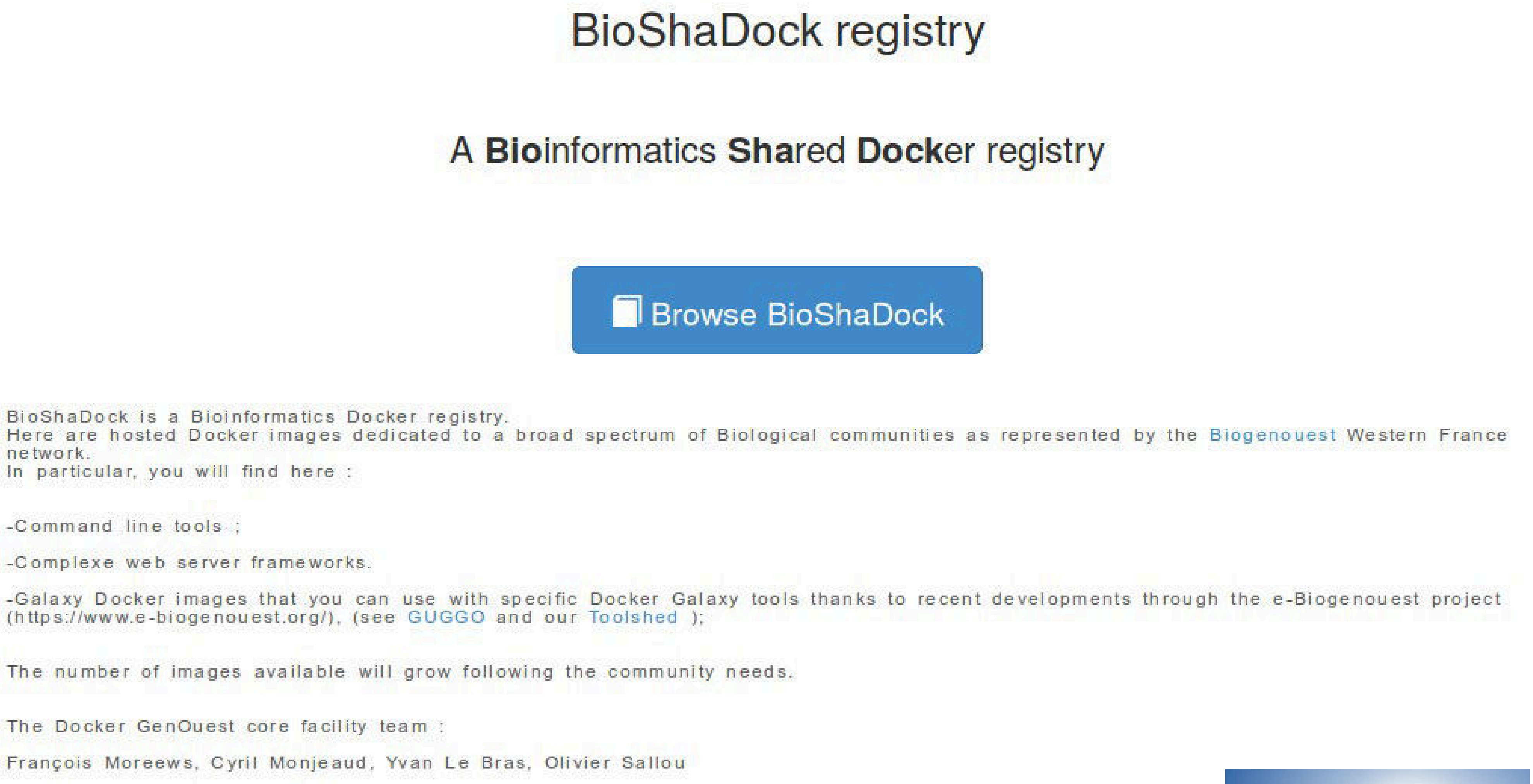
BIOSHADOCK

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 **Bio**informatics **Sha**red **Dock**er registry

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);

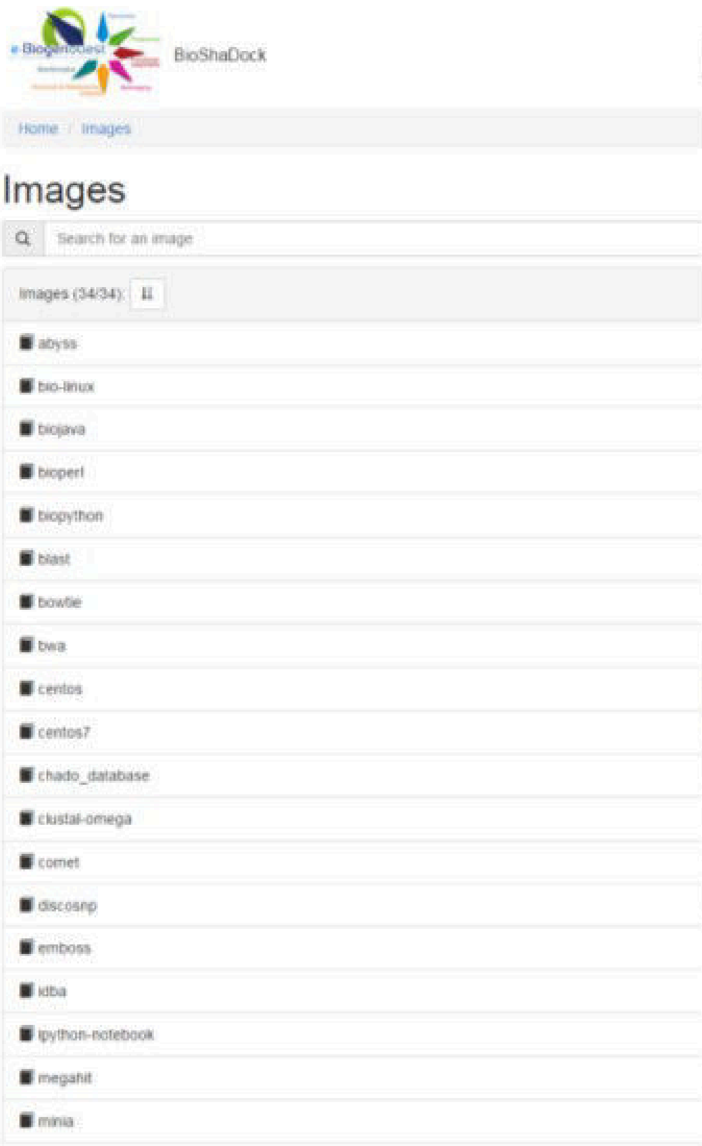
The number of images available will grow following the community needs.

The Docker GenOuest core facility team :

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

Biogenouest

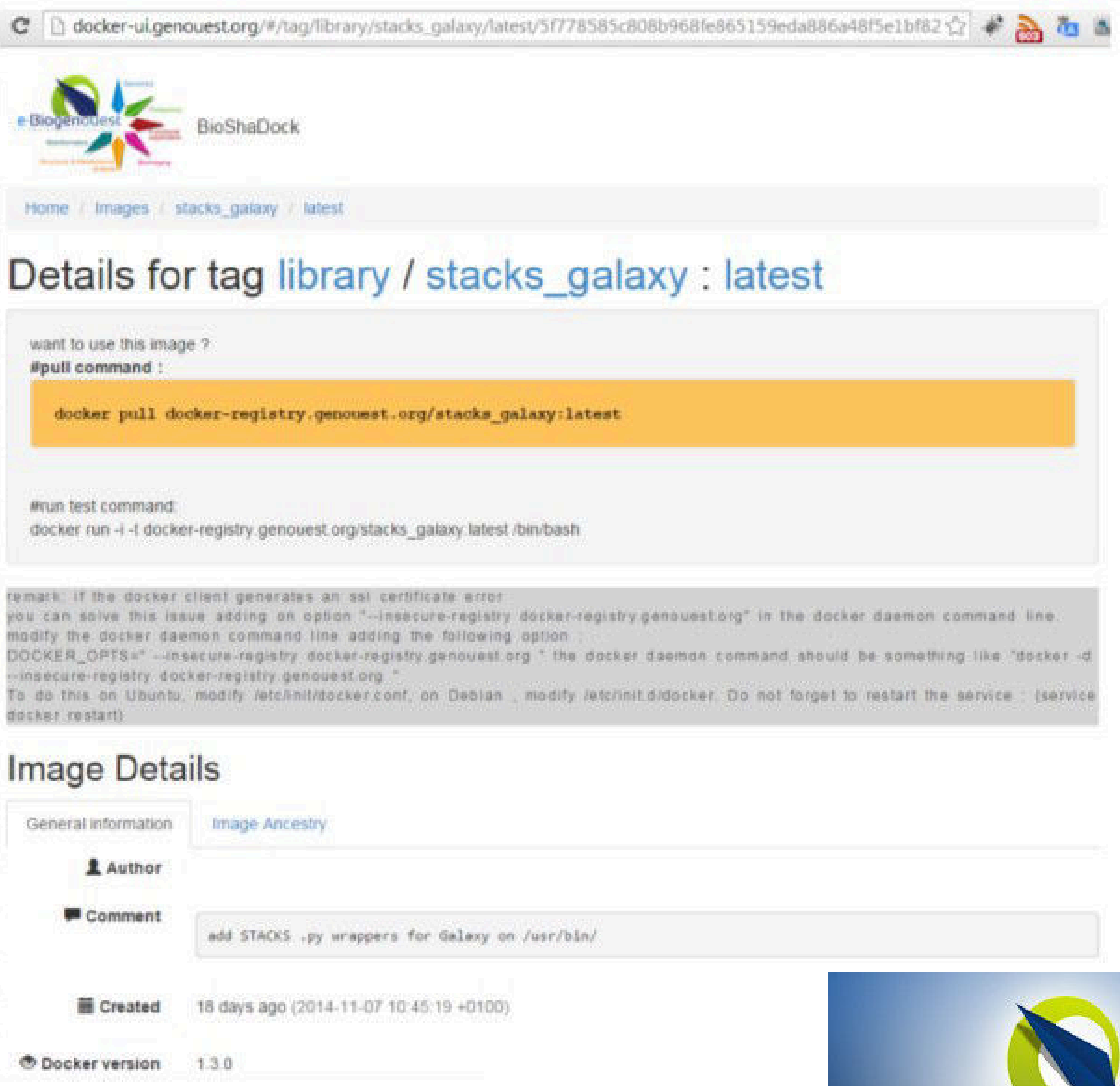
Shared registries : BIOSHADOCK



Images

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



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

Author

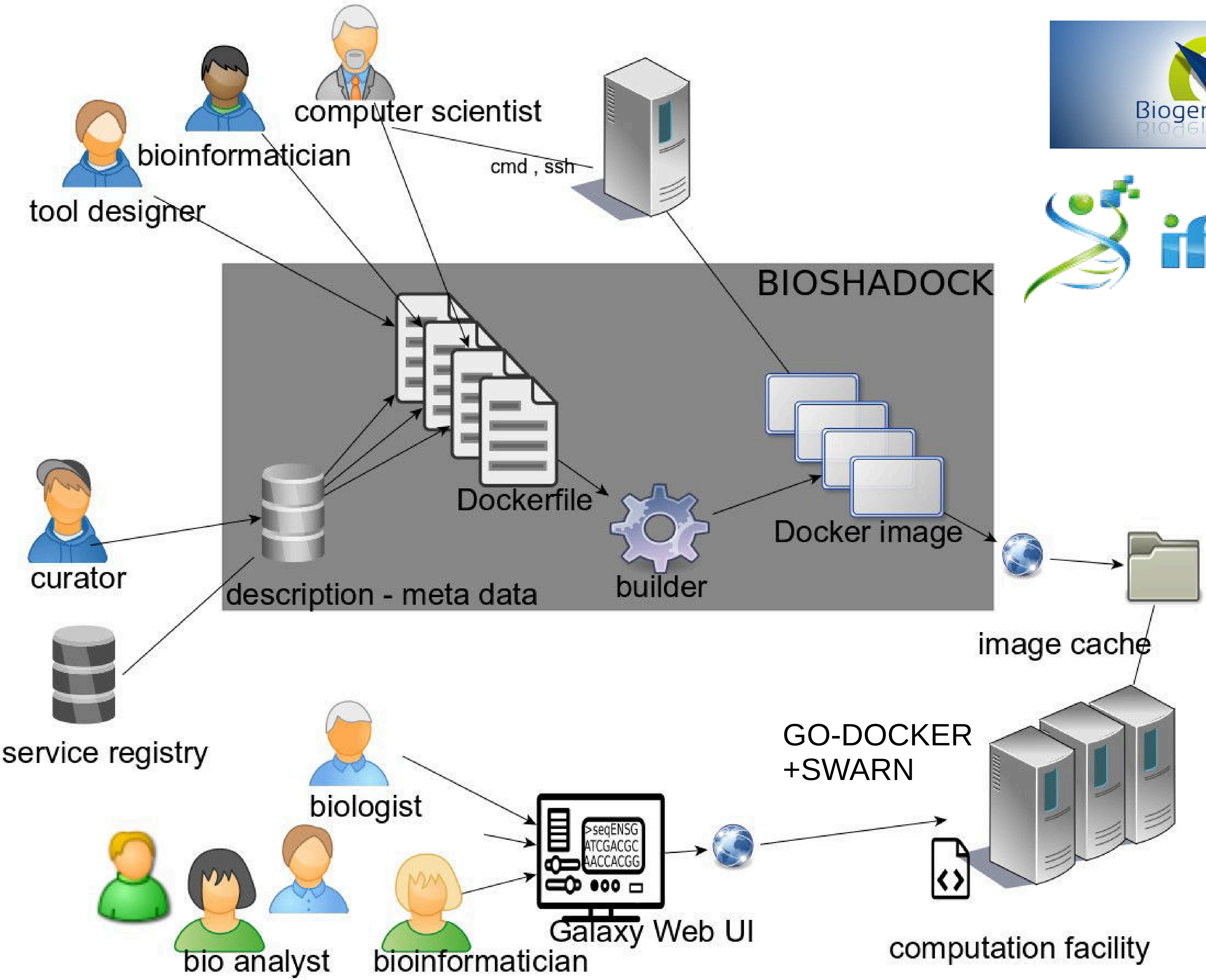
Comment add STACKS .py wrappers for Galaxy on /usr/bin/

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 repositity for softwares with or without tool.xml => SAAS + CMD
  - Integrated to Galaxy by redefining tools dependencies in a Toolshed



computer scientist

bioinformatician

tool designer

cmd , ssh

BIOSHADOCK

Dockerfile

builder

Docker image

curator

description - meta data

service registry

image cache

GO-DOCKER
+SWARN

biologist

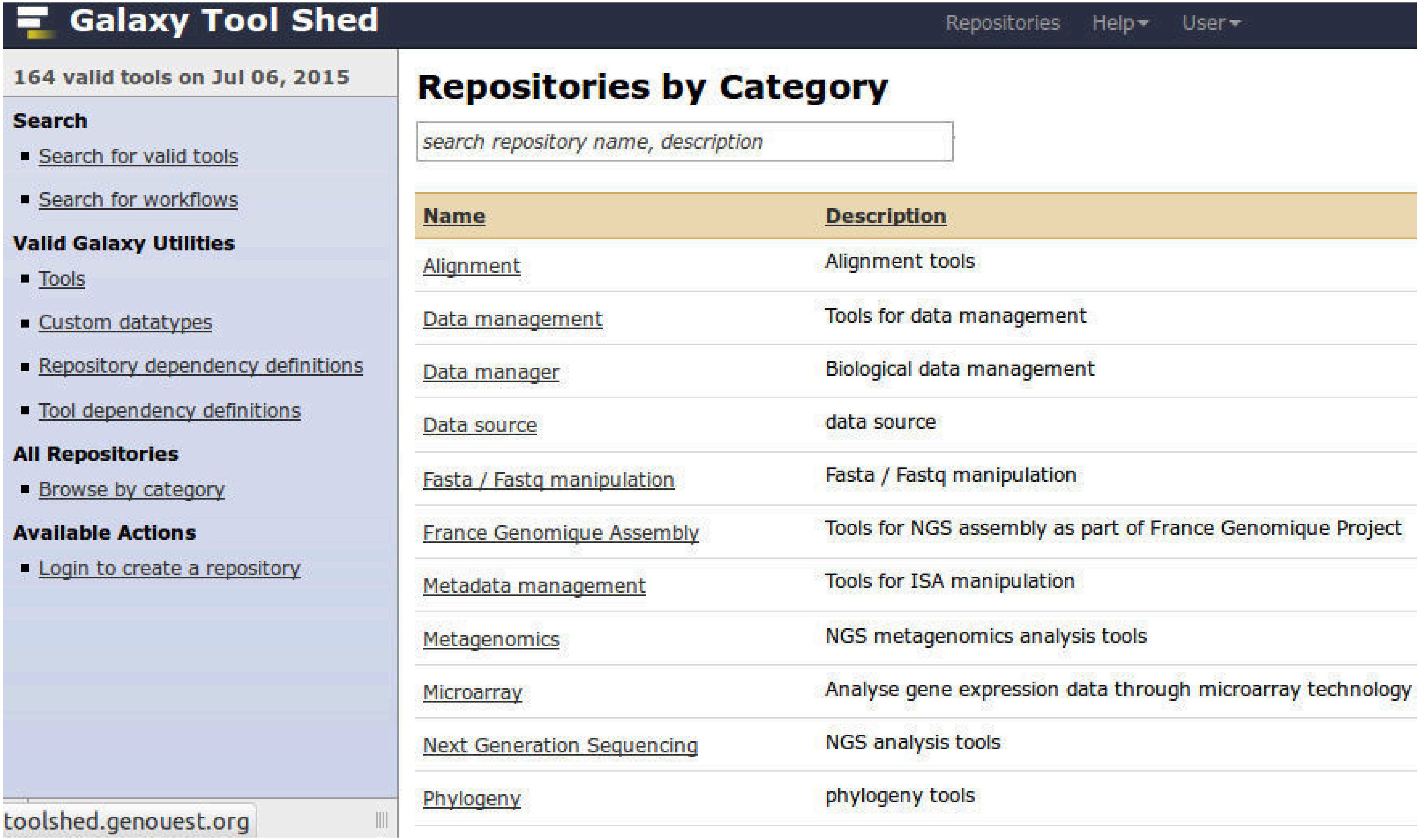
bio analyst

bioinformatician

Galaxy Web UI

computation facility

BIOSHADOCK TOOLSHED integration



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

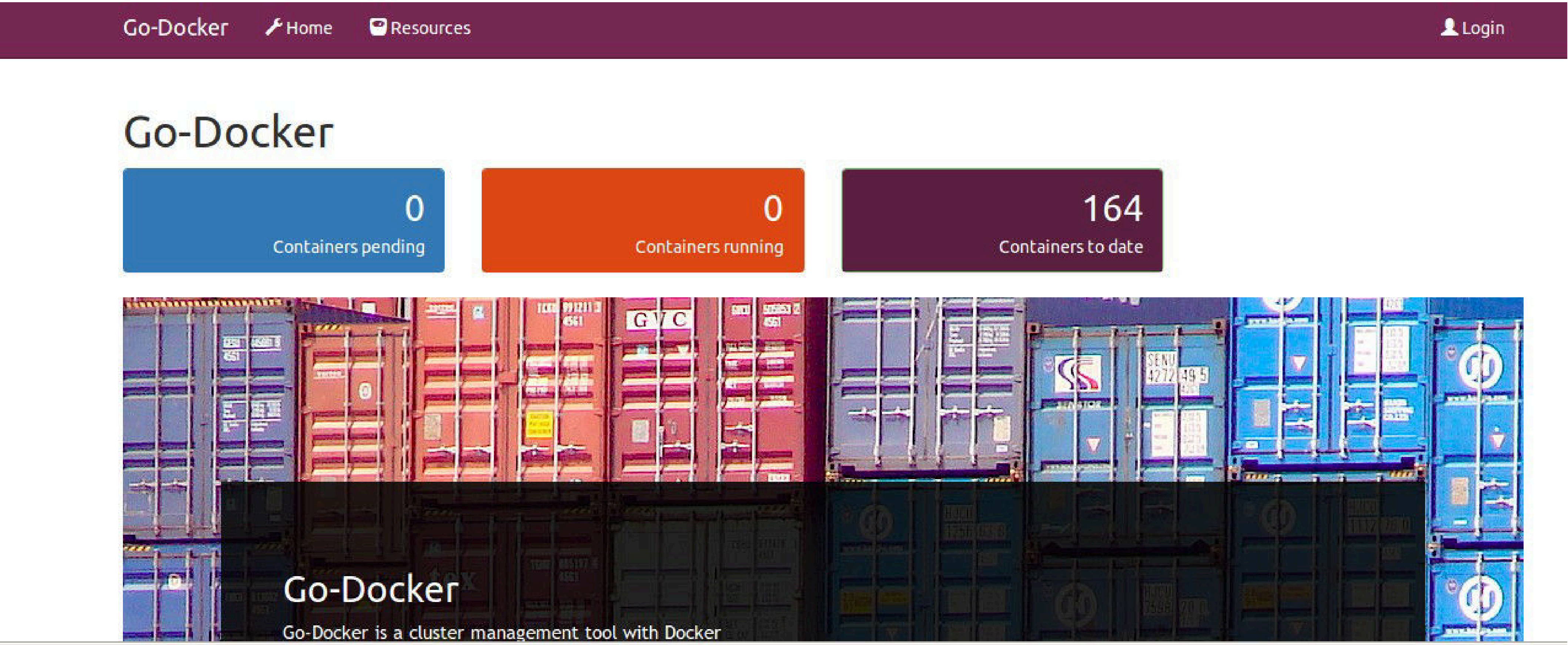
toolshed.genouest.org

Repositories by Category

search repository name, description

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

BIOSHADOCK cluster /cloud integration using GO-DOCKER



Go-Docker

0
Containers pending

0
Containers running

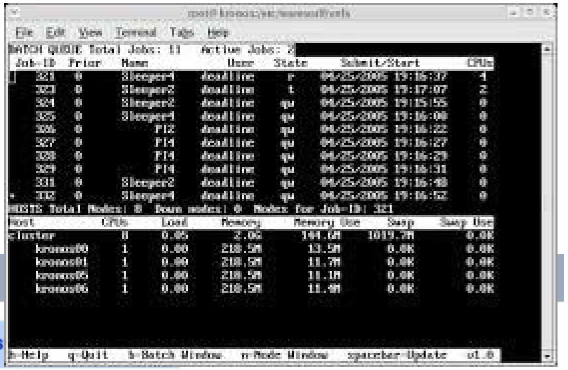
164
Containers to date

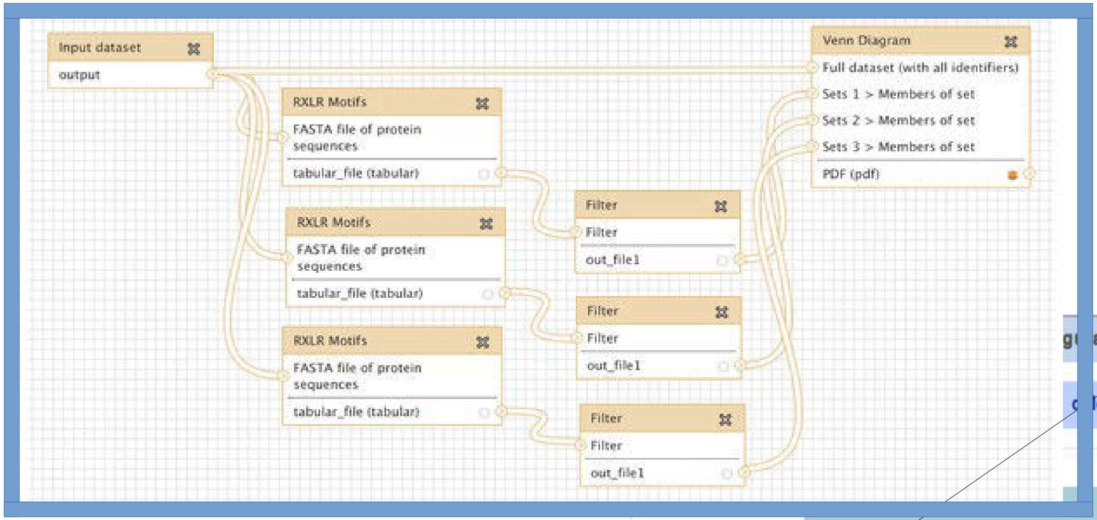
Go-Docker

Go-Docker is a cluster management tool with Docker

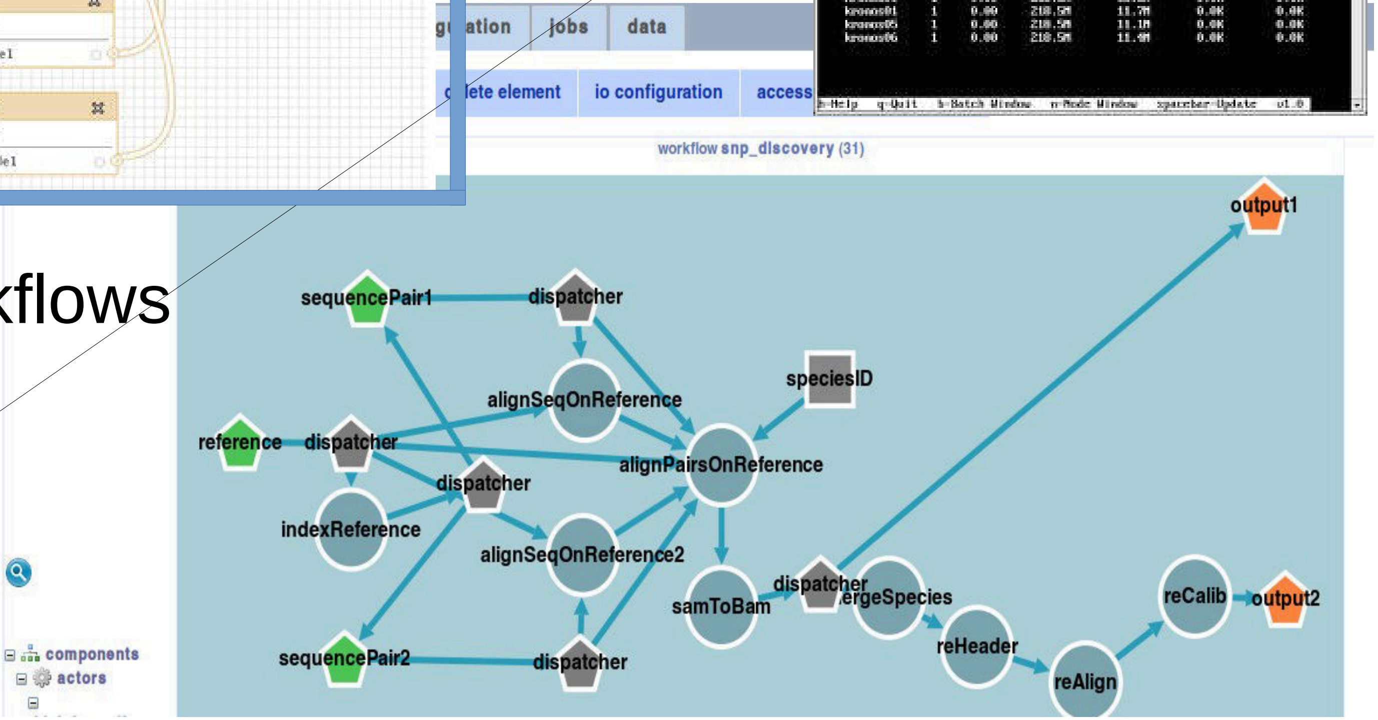
Build it one time, use it as you want

Command lines





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 interchangable bioinformatics software containers : http://bioboxes.org
- IFB academic Cloud : http://www.france-bioinformatique.fr/?q=en/core/e-infrastructure-team/ifb-cloud