Outline

• Summary of PhenoMeNal (H2020 project)
• BioMedical Metabolomics Facility Leiden
• BioMedical Metabolomics Research
  – Challenges / opportunities
• Services
  – Portal
  – Galaxy, Jupyter, Luigi
• Software stack
  – Infrastructure
• Software lifecycle
  – Development pipeline (CI/CD)
  – How to contribute software
  – Open Source contributions
• Feedback
PhenoMeNAL

Large-Scale Computing for Medical Metabolomics

Summary
Phenome and Metabolome aNalysis (PhenoMeNal)

Combination of experts:

- Bioinformatics
- High performance computing (HPC)
- Bio-Medical Metabolomics labs

Build a comprehensive and standardised e-infrastructure for data processing and metabolomics analysis pipelines
Phenomenal-h2020.eu
@PhnmlH2020
PhenoMeNal is funded by European Commission's Horizon2020 program, grant agreement number 654241.
Our role

- WP2: Sustainability

- Release management
  - WP5 + WP6 + WP9 = release

- BioMedical Metabolomics facility (Leiden)
  - Standardised platforms
  - Data processing and analysis software (in-house)
BioMedical Metabolomics Facility Leiden
• Research:
  – Clinical application of metabolomics
  – Highly sensitive analysis using 3D cell culture model
• > 50,000 samples/year
• GC- and LC-MS
  – Targeted
  – Screening (full-scan)
• Instrument & Method development
BioMedical Metabolomics Research

Challenges

• Instruments produce more data (high-throughput, more precise/accurate)
• Data integration with other omics (genomics, imaging)

Opportunities

• On-demand access to public and commercial cloud resources.
• Open Source
  • Workflow engines: Galaxy, Luigi,…
  • Interactive environments: IPython/Jupyter/Rstudio
  • Software distribution and deployment: Docker, Terraform, Ansible, Kubernetes …
Resources
Public data

- Metabolomic Repository Bordeaux
- Metabolomics Workbench (NIH)
- MetaboLights (EU/UK)

721 public data sets

28-09-2017
Services (VRE)

for HPC cloud or workstation
Build a comprehensive and standardised e-infrastructure for data processing and metabolomics analysis pipelines

Virtual Research Environment (VRE)
Portal

- Create/access your VRE(s)
- App Library (showcase)
- Help(desk)
- Demo instance (Galaxy)
Portal

• Elixir SSO (Google, ORCID, LinkedIn)
Portal

- Terms and Conditions
- ELSI
  - Ethical,
  - Legal and
  - Social Implications
Portal

PhenoMeNal Cloud (public instance)

Wizards for on premises and commercial cloud infrastructure deployment

- OpenStack (coming soon)
- Amazon (AWS)
- Google (GCP)
VRE

Galaxy Workflow Engine

Etienne Thévenot

Pierrick Roger Mele
VRE

- Fetch data MetaboLights
- Data pre-processing
- Statistics
- Annotation
- Results back to MetaboLights

Of LC-MS, GC-MS, NMR, and Fluxomics
VRE

Jupyter (Ipython) integration

By Björn Grüning and Eric Rasche

- Use Galaxy user/auth
- Before, during, or after workflow execution

https://github.com/bgruening/galaxy-ipython
VRE

Luigi (Spotify)

work in progress

Marco Capuccini
Data Scientist, Uppsala University
Software stack

start a live demo!
Your own instance @Google/AWS?
Architecture

Research enabling services

Containerised tools

Kubernetes cluster

Google, Amazon, OpenStack (workstation using Minikube)
Infrastructure
- on demand
- setup in minutes
- pay-per-use

Software
- containerized software that runs everywhere
- shared resources, without conflicting dependencies
- orchestration instead of installation

https://portal.tsi.ebi.ac.uk/#!/
Software lifecycle
Build a comprehensive and standardised e-infrastructure for data processing and metabolomics analysis pipelines

“Lead by example” in software development, distribution and deployment for metabolomics
Workflows and tools

• Portability via containers
• Workflow ready
• Packaged and ready for distribution (HPC or workstation)
How to make your software available through PhenoMeNal?

Docker

- Write a Dockerfile and add it to the phnmnl repository @github
- Add tests
- Add to Jenkins CI

All successful builds are stored in the PhenoMeNal container registry and Biocontainers.pro

https://docs.docker.com/engine/userguide/eng-image/dockerfile_best-practices/
Docker

- Add second job to Jenkins to test the container
- Runs after a successful build of the container, tests run on a real Kubernetes cluster.

Helm is a tool for managing Kubernetes charts. Charts are packages of pre-configured Kubernetes resources.

https://github.com/phnmnl/jenkins-scripts
https://github.com/phnmnl/helm-charts
Galaxy

- Create a tool wrapper
galaxyproject.org/admin/tools/add-tool-tutorial
planemo.readthedocs.io/
- Add tool to Galaxy
  - Forking container-galaxy-k8s-runtime
  - Add Galaxy tool wrapper
  - Register tool in job_conf.xml
  - Add to menu tool_conf.xml

```xml
<destination id="ramid-container" runner="k8s">
  <param id="docker_repo_override">container-registry.phenomenal-h2020.eu</param>
  <param id="docker_owner_override">phmmnl</param>
  <param id="docker_image_override">ramid</param>
  <param id="docker_tag_override">latest</param>
  <param id="max_pod_retrials">3</param>
  <param id="docker_enabled">true</param>
</destination>
```

```xml
<section name="Fluxomics" id="pheno-fluxomics">
  <tool file="phenomenal/fluxomics/ramid/ramid.xml"/>
  ...
```
Galaxy

• Test it on local Kubernetes instance (minkube)
  github.com/phnmnl/phenomenal-h2020/wiki/galaxy-with-k8s
  github.com/phnmnl/phenomenal-h2020/wiki/QuickStart-Installation-for-Local-PhenoMeNal-Workflow

• Send pull request back to PhenoMeNal repository
Open Source contributions

PhenoMeNal runtime for Galaxy running inside a container orchestrator

https://github.com/phnmnl/container-galaxy-k8s-runtime

Pablo Moreno
https://github.com/phnmnl/container-galaxy-k8s-runtime
Open Source contributions

KubeNow
provisioning Kubernetes clusters
http://kubenow.readthedocs.io/

A Kubernetes cluster up and running in less than 10 minutes (provisioned with `kubeadm`)
Weave networking
Traefik HTTP reverse proxy and load balancer
Cloudflare dynamic DNS integration
GlusterFS distributed file system
Continue Live demo!

Your own instance @Google/AWS?
Take home messages

- Same user experience in the cloud and on a local workstation
- Automate the development and deployment (DevOps) of software
- Could not have done this without Open Source!
Please help get the Netherlands to the first position!

Community building!

portaldev.phenomenal-h2020.eu

github.com/phnmnl/phenomenal-h2020
Acknowledgements

- **EBI**: Chris Steinbeck*, Claire O’Donovan, Namrata Kale, Kenneth Haug, Reza Salek, Sijin He, Pablo Moreno


- **IPB-Halle**: Steffen Neumann, Kristian Peters, Christoph Rutkkiest, Daniel Schöber

- **CEA**: Etienne Thevenot, Pierrick Roger-Melee.

- **Imperial College London**: Robert Glen, Tim Ebbels, Jianlian Gao, Noreddin Sadawi.


- **U. of Barcelona**: Marta Cascante, Carles Foguet, Vitaly Selivanov, Pedro de Atauri

- **EBI TSI and Cloud Teams**: Dario Vianello, David Ocana

- **CRS4**: Gianluigi Zanneti, Marco Enrico Piras, Luca Pireddu.

- **U. of Leiden**: Merlijn van Rijswijk, Thomas Hankemeier, Michael van Vliet

- **INRA**: Fabien Jourdan, Benjamin Merlet

- **SIB**: Sven Bergmann, Rico Ruedii, Roger Mallol

- **CIRMMP**: Antonio Rosato

- **U. of Birmingham**: Mark Viant, Ralf Webber, Michelle Thompson

- **BBMRI-ERIC**

- Communities around many open source projects (Galaxy, Kubernetes, Jupyter, Luigi, etc).
Questions or feedback?
Contact information LACDR

Michael van Vliet
Scientific Programmer @LU

Leiden Academic Centre for Drug Research (LACDR)
Leiden University, Gorlaeus Laboratories

Einsteinweg 55, 2333 CC Leiden
PO Box 9502, 2300 RA Leiden
The Netherlands

E: m.s.vanvliet@lacdr.leidenuniv.nl