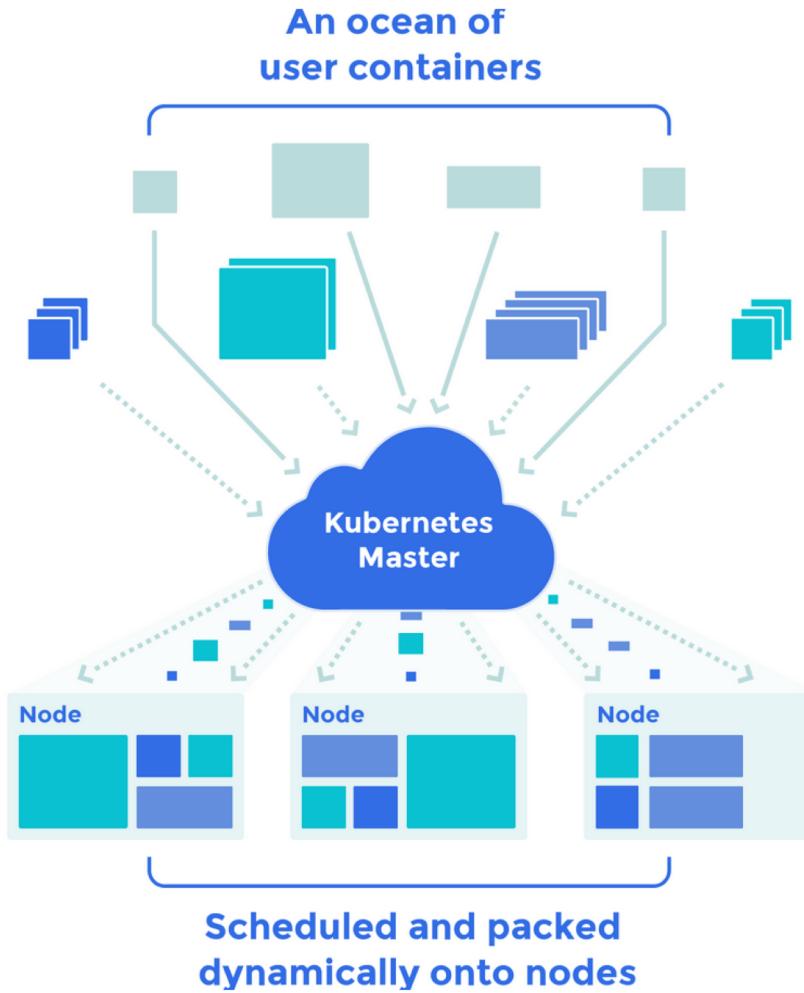


# Galaxy on Kubernetes

Pablo Moreno

Gene Expression Team

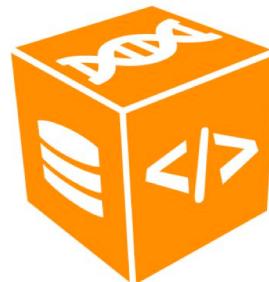
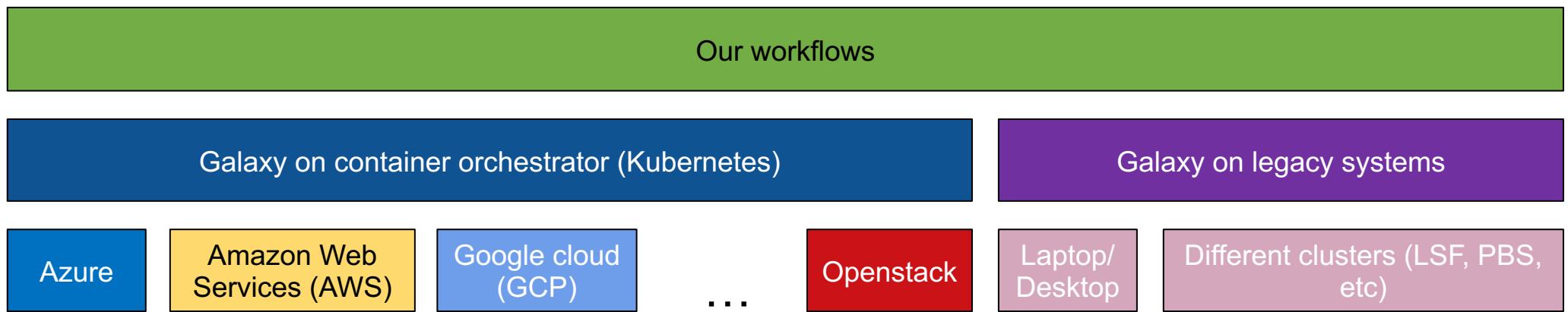
# Container orchestrator



# Same workflows – multiple clouds



2015-2018



BioContainers

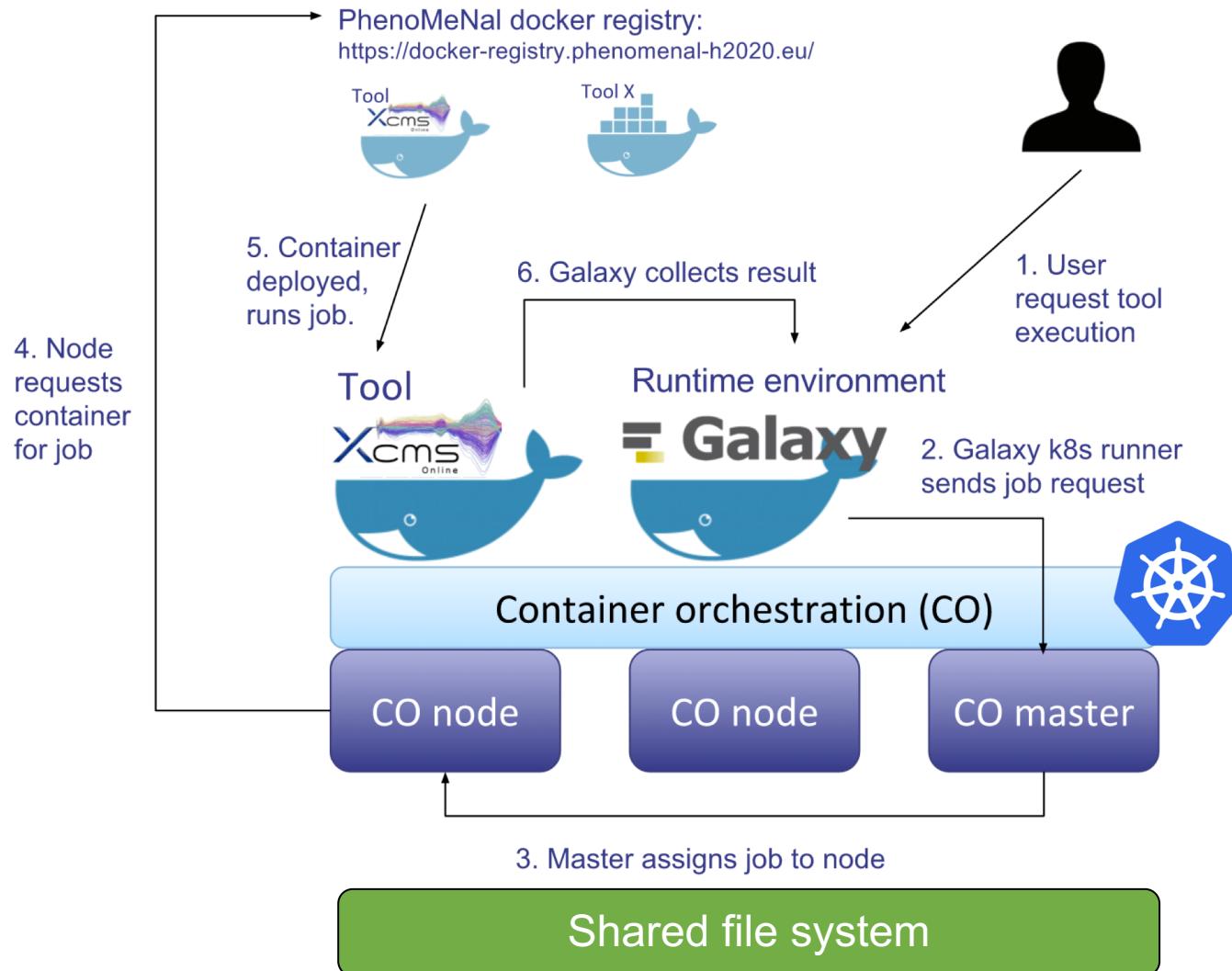
**BIOCONDA**<sup>®</sup>

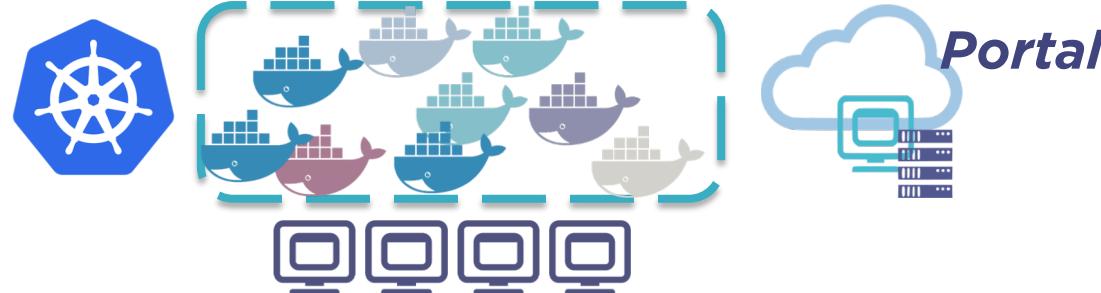
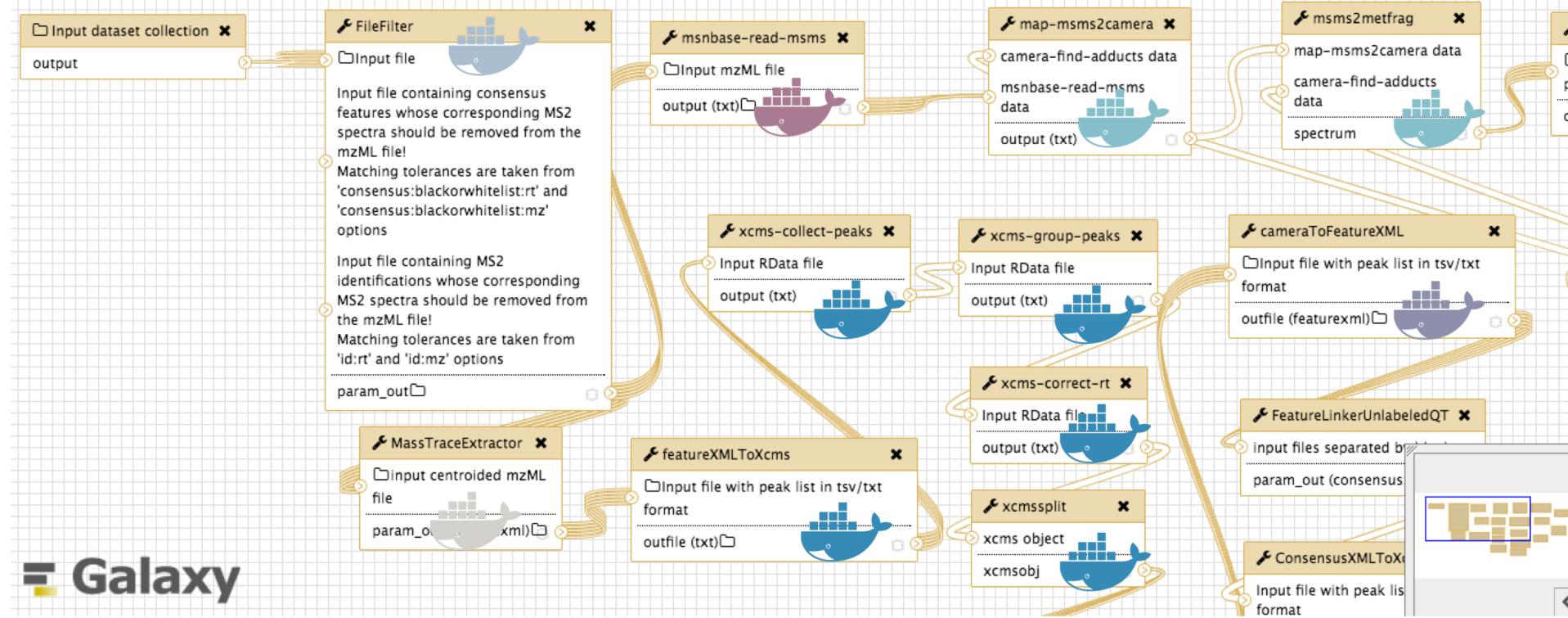
EMBL-EBI The EMBL-EBI logo, which is a green circular pattern of dots forming a grid-like structure with a red dot in the center.

# Workflows on containers

## Containers

- Galaxy-web
- Galaxy-init
- Postgres
- ProFTP
- Grafana





# Architecture layers

<https://github.com/kubernetes-sigs/kubespray>



K8s package manager



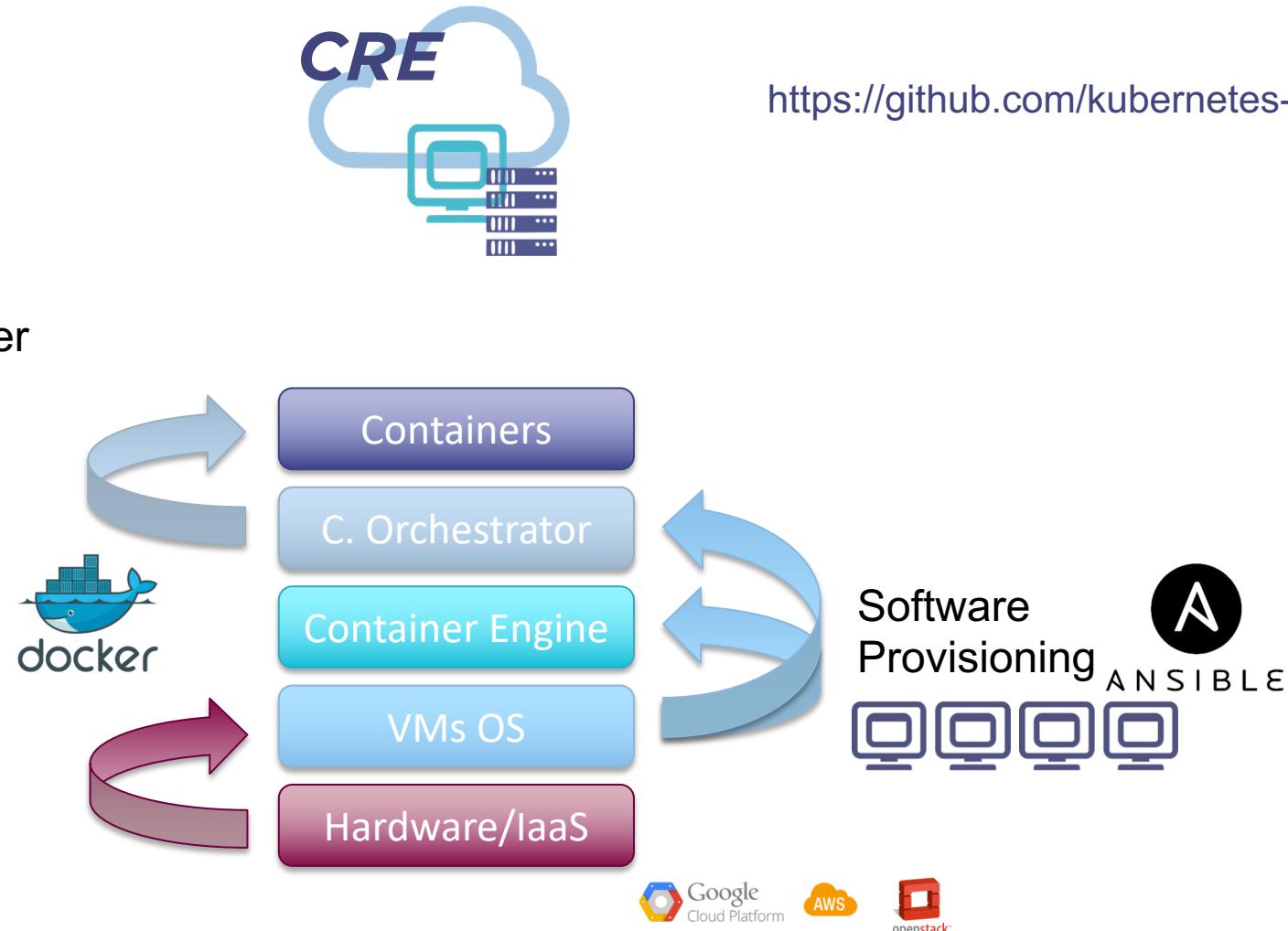
Container  
Orchestration

kubernetes



Infrastructure  
Provisioning

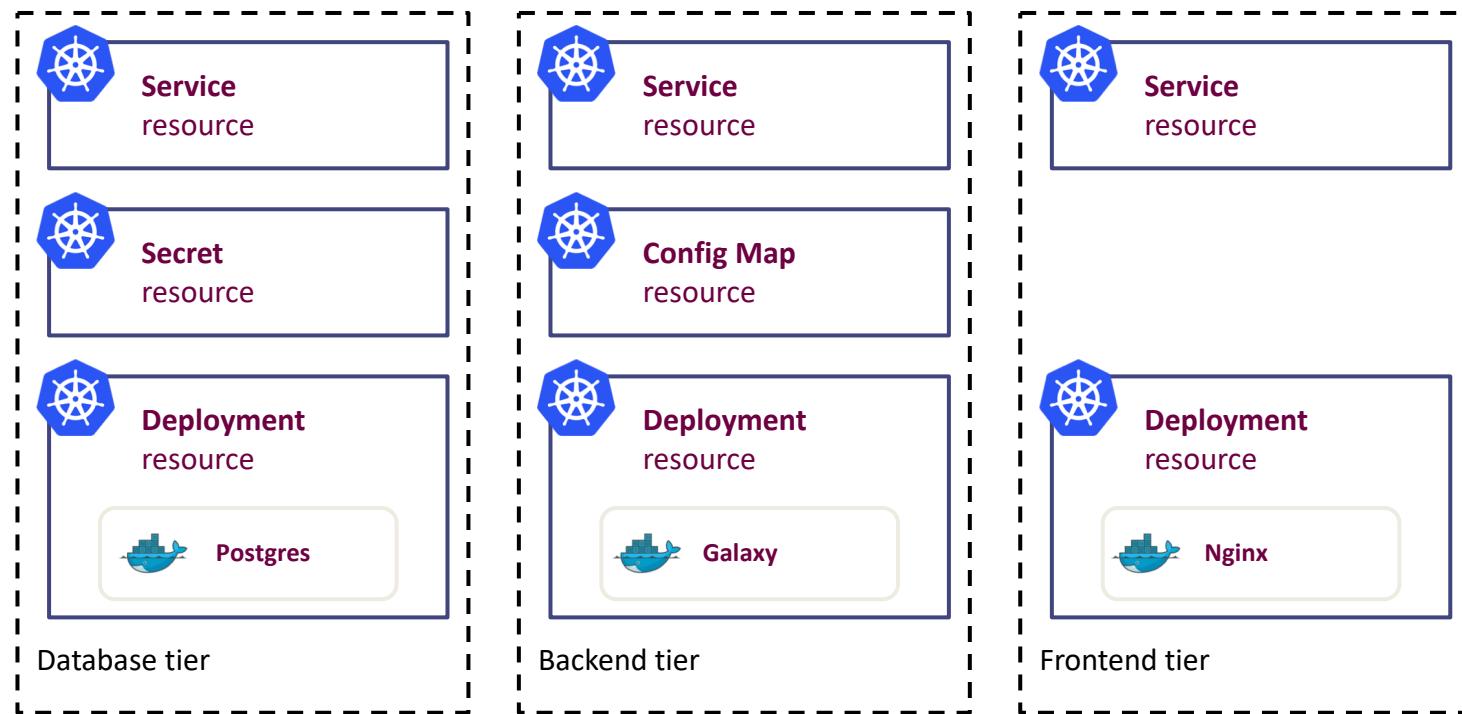
TERRAFORM



# Shared file system alternatives

- Anything supported by Kubernetes on RWX mode (see <https://kubernetes.io/docs/concepts/storage/persistent-volumes/#access-modes> ).
- Tried so far:
  - GlusterFS (direct deployment or heketi)
  - CephFS (through rook.io)
  - NFS
- Other shared fs – could work as HostPath PVs
- Let me know if you try: AWS FsX or GCP NetApp NFS

# *Spinning up Galaxy in k8s*



Adapted from: <https://www.slideshare.net/AdnanAbdulhussein/kubernetes-cicd-with-helm>

# *Example Deployment*

```
apiVersion: v1
kind: Deployment
metadata:
  name: galaxy
spec:
  replicas: 3
  template:
    metadata:
      labels:
        app: galaxy
    spec:
      containers:
        - name: my-app
          image: container-registry/phnmnl/galaxy-k8s:v1.0.0
          ports:
            - containerPort: 8080
              livenessProbe:
                httpGet:
                  path: /
                  port: 8080
                initialDelaySeconds: 120
                timeoutSeconds: 5
```

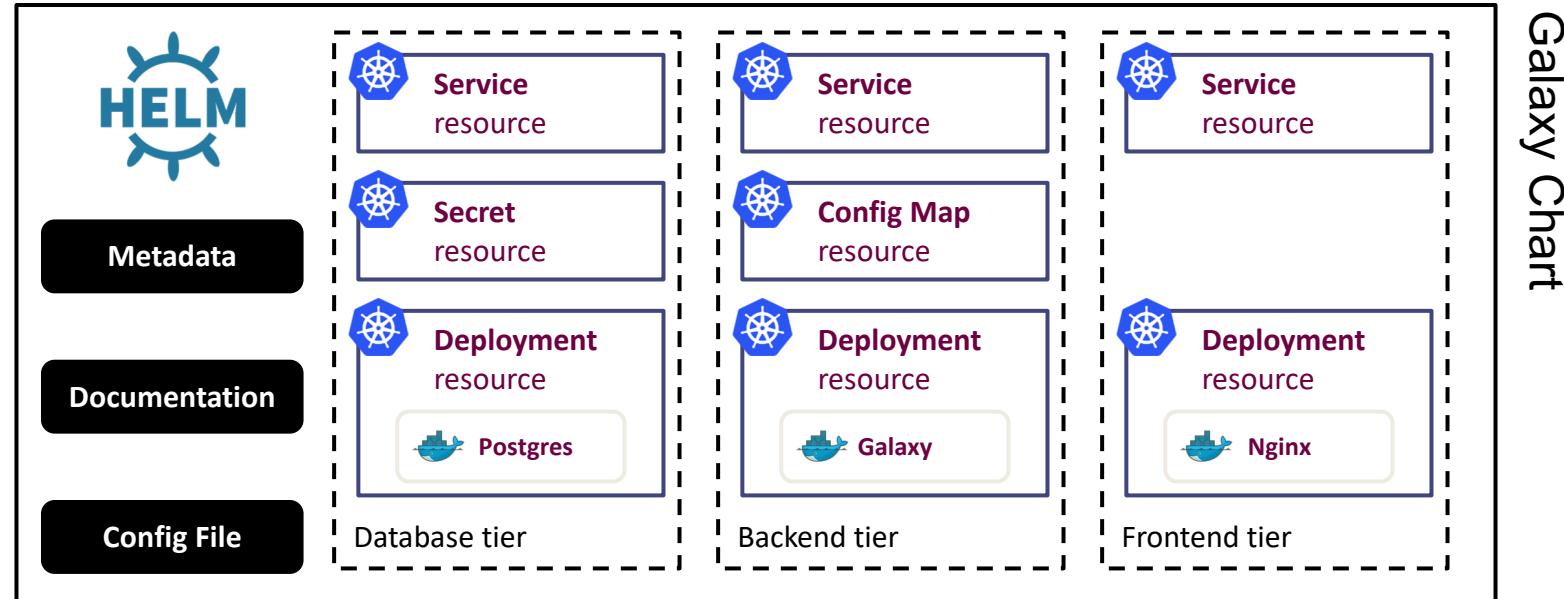
# *Example manual deployment*

```
$ kubectl apply -f ./postgres-secret.yaml  
secret "postgres" configured  
$ kubectl apply -f ./postgres-deployment.yaml  
Deployment "postgres" configured  
$ kubectl apply -f ./postgres-service.yaml  
service "postgres" configured  
  
$ kubectl apply -f ./galaxy-configmap.yaml  
configmap "galaxy" configured  
$ kubectl apply -f ./galaxy-deployment.yaml  
deployment "galaxy" configured  
$ kubectl apply -f ./galaxy-service.yaml  
service "galaxy" configured  
  
$ kubectl apply -f ./nginx-deployment.yaml  
deployment "nginx" configured  
$ kubectl apply -f ./nginx-service.yaml  
service "nginx" configured
```

Managing raw  
manifests can  
be difficult

- No template parameterization
- No application lifecycle hooks
- No history of releases

# *Deployment with Helm*



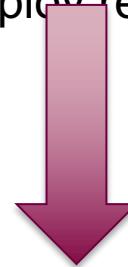
`helm install galaxy-helm-repo/galaxy`

`helm install --set admin_email=admin@domain galaxy-helm-repo/galaxy`

```
replicaCount: 1
galaxy:
  init:
    image:
      repository: pcm32/galaxy-init-pheno-flavoured
      tag: v18.01-pheno-dev
      pullPolicy: Always
    resources: {}
  backend:
    postgres: true
  image:
    repository: pcm32/galaxy-web-k8s
    tag: v18.01-pheno-dev
    pullPolicy: Always
  tools:
    destination: "/export/tools"
  destinations_default: "k8s"
admin:
  email: "admin@email.do"
  password: "admin-pass"
  api_key: "qwertyqweqwe"
  username: admin
```

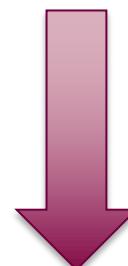
config.yaml

helm install –f config.yaml galaxy/galaxy-stable  
(produced my-deploy revision 1)



New version, change Tag of container

helm upgrade –f config.yaml my-deploy galaxy/galaxy-stable  
(produced my-deploy revision 2)



Change replication or policy.

helm upgrade –f config.yaml my-deploy galaxy/galaxy-stable  
(produced my-deploy revision 3)

## Variables

Parameter	Description		
<code>export_dir</code>	Export directory for Galaxy compose	<code>admin.allow_user_creation</code>	Configures <code>allow_user_creation</code> Galaxy config environment variable.
<code>galaxy_conf.brand</code>	Branding text displayed on Galaxy	<code>galaxy_conf.smtp_server</code>	SMTP server for Galaxy password reset functionality
<code>init.image.repository</code>	Repository for the docker image: <code>&lt;server&gt;/&lt;owner&gt;/&lt;image-name&gt;</code> for Galaxy init.	<code>galaxy_conf.smtp_username</code>	SMTP username for Galaxy password reset functionality
<code>init.image.tag</code>	Image tag for Galaxy init image.	<code>galaxy_conf.smtp_password</code>	SMTP password for Galaxy password reset functionality
<code>init.image.pullPolicy</code>	Pull policy for the Galaxy init image	<code>galaxy_conf.email_from</code>	SMTP email_from for Galaxy password reset functionality
<code>init.resources</code>	k8s resources map for the init process	<code>galaxy_conf.smtp_ssl</code>	SMTP ssl for Galaxy password reset functionality
<code>image.repository</code>	Repository for the docker image: <code>&lt;server&gt;/&lt;owner&gt;/&lt;image-name&gt;</code> for Galaxy main process.	<code>galaxy_conf.url</code>	Incoming URL label for Galaxy password reset functionality, shown on reset email to identify instance.
<code>image.tag</code>	Image tag for Galaxy image.	<code>galaxy_conf.allow_user_deletion</code>	Allows the admin to delete users
<code>image.pullPolicy</code>	Pull policy for the Galaxy image.	<code>galaxy_conf.allow_user_creation</code>	Allows the admin to delete users

<https://github.com/galaxyproject/galaxy-kubernetes/>

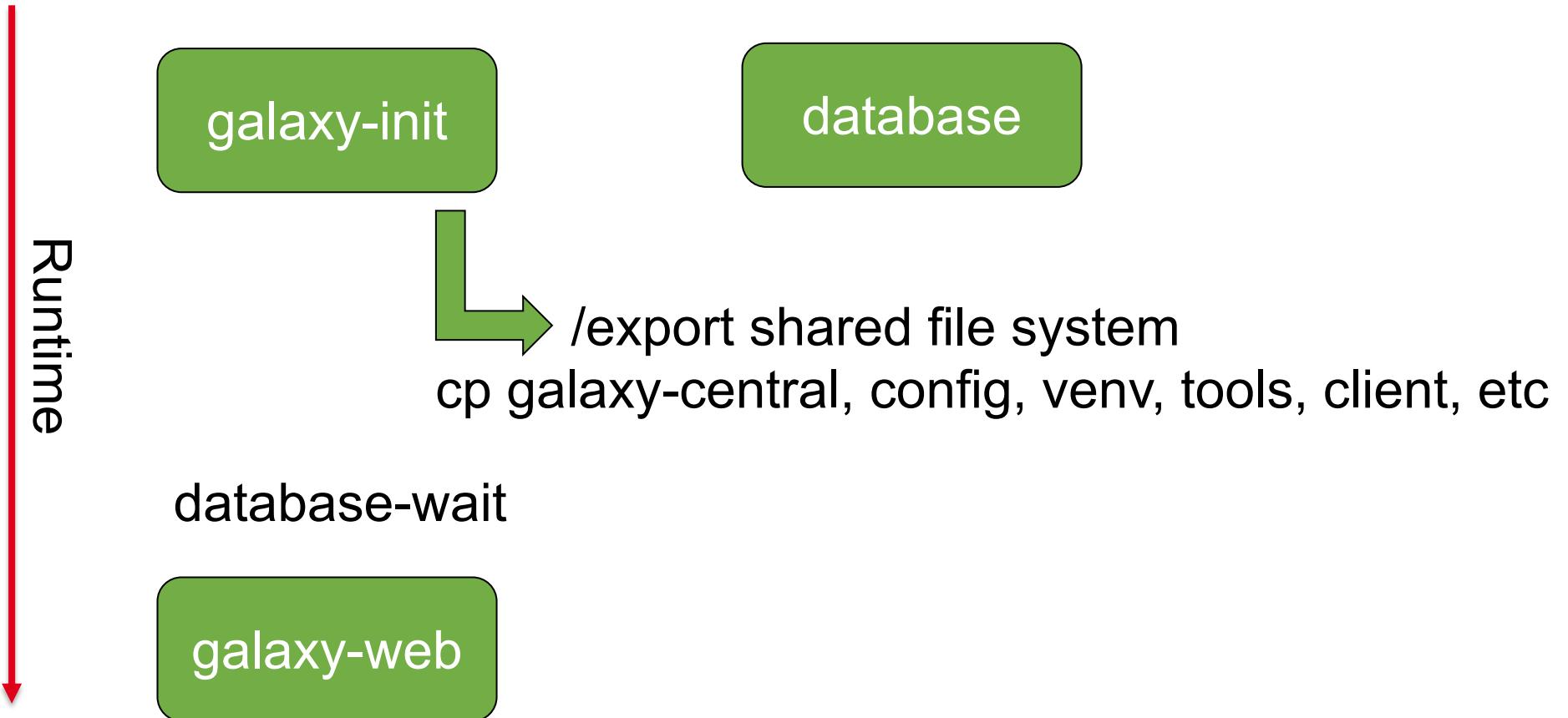
# Timeline

- Late 2015 - First versions of Kubernetes Galaxy Runner [KR]
- Mid 2016
  - First versions of Helm chart [HC]
  - [KR] Use k8s service account
  - Galaxy 16.07: first release including Kubernetes runner
  - 16.08 available to users on PhenoMeNal public instance.
- 2017
  - First PhenoMeNal release – AWS, Google Cloud
  - [KR] [HC] Non-privileged file system access & Ingresses
  - [KR] [HC] Resources requests/limits & RBACs

# Timeline

- 2018
  - [KR] [HC] Configurable namespace & Mechanism to trust jobs in cluster
  - Automatic deployment on OpenStack through PhenoMeNal
  - [KR] Improved job error handling – no more Zombie jobs.
  - [HC] Major revamp to 2.0 by @afgane & @nuwang
  - [HC] Moved to containers from docker-galaxy-stable
  - Use biocontainers/mulled natively.
  - Galaxy-Kubernetes bioarxiv paper <https://doi.org/10.1101/488643>
- 2019
  - Major revamp of the Kubernetes runner by @mvdbekk – 19.05 scheduled

# Container model under docker-galaxy-stable



# Use of docker-galaxy-stable

- Advantages
  - Reduce maintainability burden
  - Lots of good practices built-in
  - Production quality
  - Only extend init
- Disadvantages
  - Very heavy images (serve too many purposes).
  - Slow startup
  - Orchestration model is awkward (due to ^^^)

## Current limitations/issues

- Shared file system
- One PVC per deployment
- Postgres sub-chart is old and doesn't expose all configuration needed.
- Handlers and web on the same container
- Slow startup
- Resubmission not working
- Heavy images

# The future

- Deployment of handlers in their own containers as per scalation need (@afgane & @nuwang have a prototype of this at GVL) and other orchestration layout improvements.
- Smaller images & shorter startup times.
- Interactive environment support (@pcm32 halfway through).
- Manage to run without a shared file systems, relying more on object stores. (some ongoing by @jmchilton Galaxy wise).
- Auto-restarts on out of memory (should be possible after @mvdbeek revamps).
- Metrics integrations (Prometheus or similar).
- Talk to federated Kubernetes clusters

# Where to contribute

- Galaxy k8s runner:  
<https://github.com/galaxyproject/galaxy/blob/dev/lib/galaxy/jobs/runners/kubernetes.py>
- Galaxy helm chart: <https://github.com/galaxyproject/galaxy-kubernetes>
- Docker-galaxy-stable: <https://github.com/bgruening/docker-galaxy-stable/tree/dev/compose>

# Making your own setup

- Extend galaxy-init, very much like we do here: [https://github.com/ebi-gene-expression-group/container-galaxy-sc-tertiary/blob/develop/Dockerfile\\_init](https://github.com/ebi-gene-expression-group/container-galaxy-sc-tertiary/blob/develop/Dockerfile_init)
- Push images
- Modify an example helm config to use your images:
  - [https://github.com/galaxyproject/galaxy-kubernetes/tree/develop/example\\_configs](https://github.com/galaxyproject/galaxy-kubernetes/tree/develop/example_configs)
  - <https://github.com/ebi-gene-expression-group/container-galaxy-sc-tertiary/tree/develop/helm-configs>
- helm install –f <your-config-file> galaxy-helm-repo/galaxy-stable

# Demo

- Run on minikube

# Summary

- Kubernetes abstracts most of the cloud particularities.
- Galaxy-kubernetes setup allows deployment on scalable k8s clusters.
- Has been battle tested for a few years now, different cloud providers.
- 19.05 will have an improved version of the runner.
- Space to improve

# Acknowledgments - PhenoMeNal

- **EBI:** [Chris Steinbeck\\*](#), Claire O'Donovan, Namrata Kale, Kenneth Haug, Reza Salek, Sijin He, Pablo Moreno
- **Uppsala U.:** [Ola Spijuth](#), [Kim Kultima](#), Stephanie Hermann, Payam Emami, Marco Capuccini, Anders Larson.
- **IPB-Halle:** [Steffen Neumann](#), Kristian Peters, Christoph Ruttkies, Daniel Schober
- **CEA:** [Etienne Thevenot](#), Pierrick Roger-Mele.
- **Imperial College London:** [Robert Glen](#), [Tim Ebbels](#), Jianlian Gao, Evangelos Handakas, Noreddin Sadawi.
- **U. of Oxford:** [Susanna-Asunta Sansone](#), Philippe Rocca-Serra, David Johnson, Alejandra Beltran-Beltran.
- **U. of Barcelona:** [Marta Cascante](#), Carles Foguet, Vitaly Selivanov, Pedro de Atauri
- **EBI TSI and Cloud Teams:** Steven Newhouse, Dario Vianello, David Ocana, Jose Dianes, Amelie Cornelis, Santiago Insua, Charles Short
- **CRS4:** [Gianluigi Zannetti](#), Marco Enrico Piras, Luca Pireddu.
- **U. of Leiden:** [Merlijn van Rijswyk](#), [Thomas Hankemeier](#), Michael van Vliet
- **INRA:** [Fabien Jourdan](#), Benjamin Merlet
- **SIB:** [Sven Bergmann](#), Rico Ruedii, Roger Mallol
- **CIRMMT:** [Antonio Rosato](#)
- **U. of Birmingham:** [Mark Viant](#), Ralf Webber, Michelle Thompson, James Bradbury
- **BBMRI-ERIC**
- Communities around many open source projects (Galaxy, Kubernetes, Jupyter, Luigi, etc).



PhenoMeNal is funded by European Commission's Horizon 2020 program, grant agreement number 654241.

# Acknowledgements



Irene Papatheodorou

Group Leader

Gene  
Expression

Curation & Submissions



Anja Fullgrabe – Nancy George – Silvie Fexova

Software Engineering & Web



Alfonso Fuentes – Haider Iqbal – Monica Jianu – Lingyun Zhao

Irene Papatheodorou

Group Leader

Gene  
Expression

Curation & Submissions

Bioinformatics & Data Production



Pablo Moreno – Suhail Mohammed – Jon Manning

Kerstin Meyer's Team @ Sanger

Carlos Talavera – Ni Huang

Galaxy Project, Bioconda, Kubespray,  
Biocontainers, CloudLaunch,  
SC Community