

DATAVERSE ON KUBERNETES

WHAT, WHY, HOW

 github.com/IQSS/dataverse-kubernetes

 dataverse-k8s.rtf.d.io



24.01.2020 | Oliver Bertuch



\$ WHOAMI

OLIVER BERTUCH

Research Software Engineer + DevOp

Working at [RDM division of Central Library @ Jülich Research Centre](#)

Hacking on [Dataverse](#), [Dataverse on K8s](#) and [Jülich DATA](#)

Maintainer of [Dataverse on Kubernetes](#) project

\$ WHOAREYOU

QUICK FINGERS, PLEASE...

1. Who heard about Kubernetes?
2. Who knows what a Pod is?
3. Who tried or even uses Kubernetes?
4. Who has Kubernetes on their laptops for dev?

🐦 2010 IS CALLING.

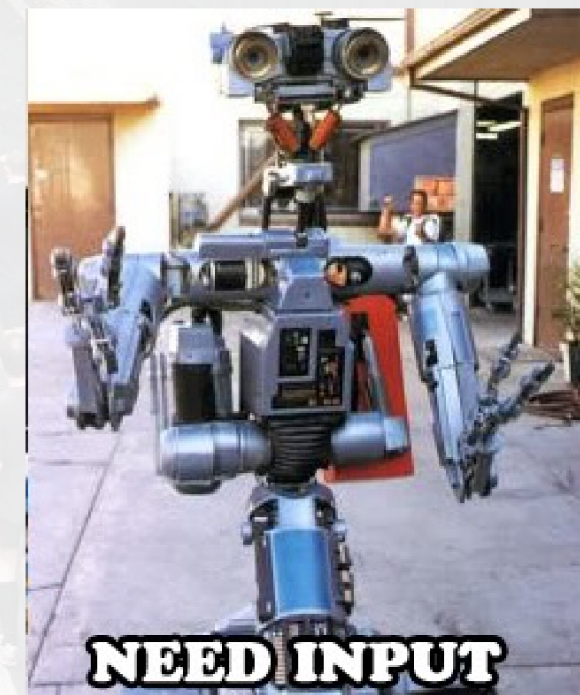


A rant about the cloud hype and how stupid private clouds are. [Heise article](#)

If you think clouds are the evil twin of all evil:
feel free to take a nap for ~30 minutes. 😊

AGENDA

- Very briefly explain K8s (K8s?)
- What's that project up to?
- Building blocks and demo
- Please wake up again...





KUBERNETES

A VERY SHORT INTRODUCTION (1/2)

- Docker != Kubernetes, but Kubernetes can use Docker as its container runtime
- Supports both Windows and Linux containers
- Huge open source project, available in commercial products (VMWare, OpenShift, public clouds, ...)
- Orchestrator for Container Management since 2014

KUBERNETES

A VERY SHORT INTRODUCTION (2/2)

Ideas:

- *Declarative* managment¹⁾ of objects first
- Some things done *imperative*
- Abstraction all over:
 - Hardware & OS (containers)
 - Storage (volumes/SDS)
 - Network (SDN)
- Replace rather than upgrade

Goals:

- Initially in 2014:
 - easy massive scaling of stateless apps
 - *Don't shepherd pets, grow to like cattles*
- Today:
 - universal tool to manage all kind of apps as a cloud (yet statefull is complex with persistance)

MORE INTRODUCTION MATERIALS:

- [K8s in 10 min](#)
- [K8s in < 3h](#)
- [Kubernetes upstream docs](#)

1) see Puppet/Ansible for similar ideas

KUBERNETES

WHAT GIVES...?

FOR RDM

- Enable moving (reproducible) processing¹⁾ nearer to data
(see also Data Commons talk earlier today)
- Distributed infrastructures simplify scaling (not only for heavy scientific computing)

FOR DEVS

- Concentrate on running apps and development of 'em
- Less "works on my machine":
consistent env from dev to prod
- Adapt modern good practices as CI & CD, GitOps, staging and testing in ephemeral envs, ...

FOR ADMINS

- Get teams working together:
 - Devs become responsible for production, too, getting fixes and features shipped faster
 - Admins still have their powers unleashed and maintain clear responsibilities
 - Does *DevOps* sound familiar?

1) see WholeTale, Bindiverse, Jupyter and many more integrations for Dataverse

KUBERNETES


ANY OBJECTIONS?

- *But it's so complex!*
 - Grow up.
 - Persistence somewhat complex due to initial design.
- *But I'm Google!*
 - Nope (unless you are).
 - Beware: your custom cluster? You pay those experts!
 - Use managed or supported, public and/or private.
 - Reduce overhead with corporate or consortium wide clusters
- *But it's so complex! And so many tools...*
 - Be brave. You'll master the learning curve.
 - Start with *kubect!* 🍷 You're done 90%.

THE "DATAVERSE ON " PROJECT

THIS AIN'T IQSS.

☆ They focus on the app and its features. (Thank you!)

We focus on cloud-enabling. 

☰ We're all connected every (work)day, sharing and learning.
Pushing both forward together. ▶▶

GOALS AND TARGETS

1. Create a common base for non-classic installations
2. Containers - provide building blocks for prod and dev
3. Documentation - home sweet home for things omitted from upstream
4. Kubernetes - provide sustainable, reusable objects, workflows & tests
5. No forks of upstream. Stay community-driven.



CONTAINER BUILDING BLOCKS

-  [iqss/dataverse-k8s](#): Glassfish, Dataverse & Scripts in a box
-  [iqss/solr-k8s](#): Solr, Configuration & Schema in a box



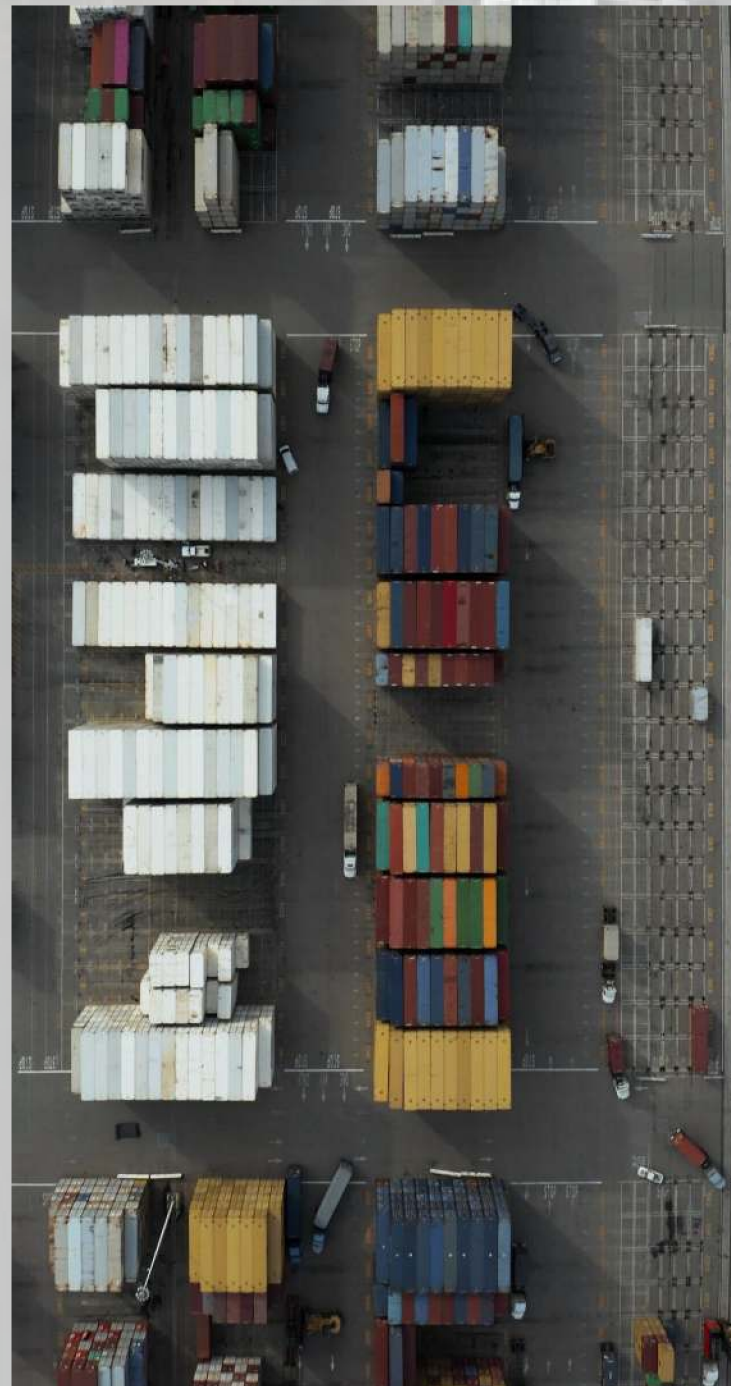
- Tags for upstream release versions. Sec updates for current.
- latest means master of project. Might be unstable.

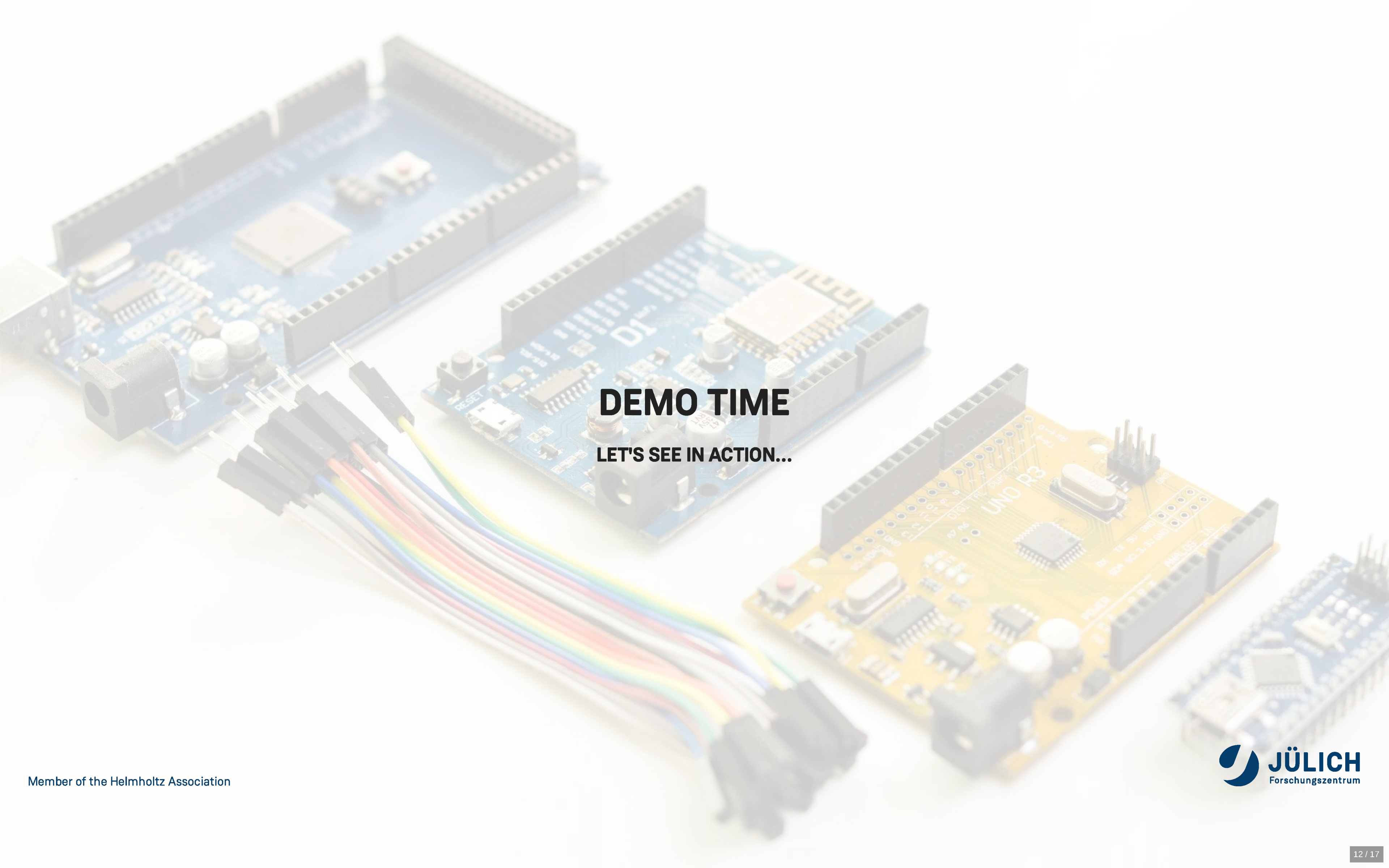


 Image building done via (reusable) public Jenkinsfiles
Thank you [Don Sizemore @ ODUM/UNC](#) for donating  community service!



Future: provide develop snaps, Payara [#63](#) and sec scans [#115](#).





DEMO TIME
LET'S SEE IN ACTION...

K8S OBJECTS

Declarative descriptions how we want Dataverse to look like in our cloud.

- Builds on our container images.
- Actual base to build your own deployment on.
- Reusable with kustomize.io¹⁾.
- See [deploy section](#) of guide and demo (later).
- Maintenance tasks as imperative jobs.

1) No Helm Chart or CRD Operator - yet?

K8S OBJECTS

Imperative jobs: bootstrapping, custom metadata support, ...

Use it for developing Dataverse or new integrations!

Which leads us to...

The Dataverse Cloud & Container Guide

[Table Of Contents](#)Deploying, Running and
Dataverse on Kubernetes

Deploying, Running and Using Dataverse on Kubernetes



A VERY QUICK DEMO

Dataverse v4.18 kubeval valid image dataverse-k8s image solr-k8s license Apache-2.0 docs failing IRC chat #dataverse

<https://dataverse-k8s.rtf.d.io>

This community-supported project aims to provide simple to re-use Kubernetes objects on how to run Dataverse on a Kubernetes cluster.

It aims at day-1 deployments and day-2 operations.
(Say hello to friendly @dataversebot at upper left 🙋)

- Documentation: <https://dataverse-k8s.rtf.d.io>
- Support: <https://github.com/IQSS/dataverse-kubernetes/issues>
- Roadmap: <https://dataverse-k8s.rtf.d.io/en/latest/roadmap.html>

If you would like to contribute, you are most welcome. Head over to the [contribution guide](#) for details.

✓ Important

This is a **community driven and supported project**, unsupported by [IQSS](#), Harvard.
Current main drivers are [Forschungszentrum Jülich](#) and [DANS KNAW](#).

If you need help, please [open an issue](#) or find us on [IRC](#) or Twitter.

ROADMAP

IT'S JUST THE BEGINNING...

- Enhance images (see before)
- Provide detailed production usage guidance. Generator?
- Integrations and demos (Monitoring!, Rserve, DCM, TRSA, S3, KeyCloak, ...)
- Support auth configuration rollout.

BTW: Do not use Shibboleth as SP in the cloud.

- PoC docker-compose usage. What about podman?
- Move to Payara: make Dataverse cloud-ready
(faster startup, MicroProfile Config & Health API, Cluster to scale/multi-instance, batch for ingress/index/zip/...)
- Run unit, integration and end-to-end tests on ephemeral deployments

⌚ THANK YOU.

QUESTIONS?

- Let's talk (crazy) ideas:
 - How does this fit into your projects, your ideas?
 - What about screencasts with demos of integrations?
 - What do you need in the future?
 - What would be your priorities?
 - What about more contributors?
 - What about creating a sustainable project at GDCC?
- Also feel free to ask about OpenID Connect support.
- Please reach out... 😊
 - 💬 IRC #dataverse
 - 🐙 [poikilotherm](#)
 - 🐦 [@poi_ki_lo_therm](#)
 - ✉️ o.bertuch@fz-juelich.de
- Licensed under  
- Credits:
 - Icons by [FontAwesome](#), CC-BY 4.0
 - Backgrounds: <https://unsplash.com>