

Cloud Migration of DataverseNO

Philipp Conzett*, Vyacheslav Tykhonov**,
Rolf Andersen*, Loïc Bergeron* Obiajulu Odu*

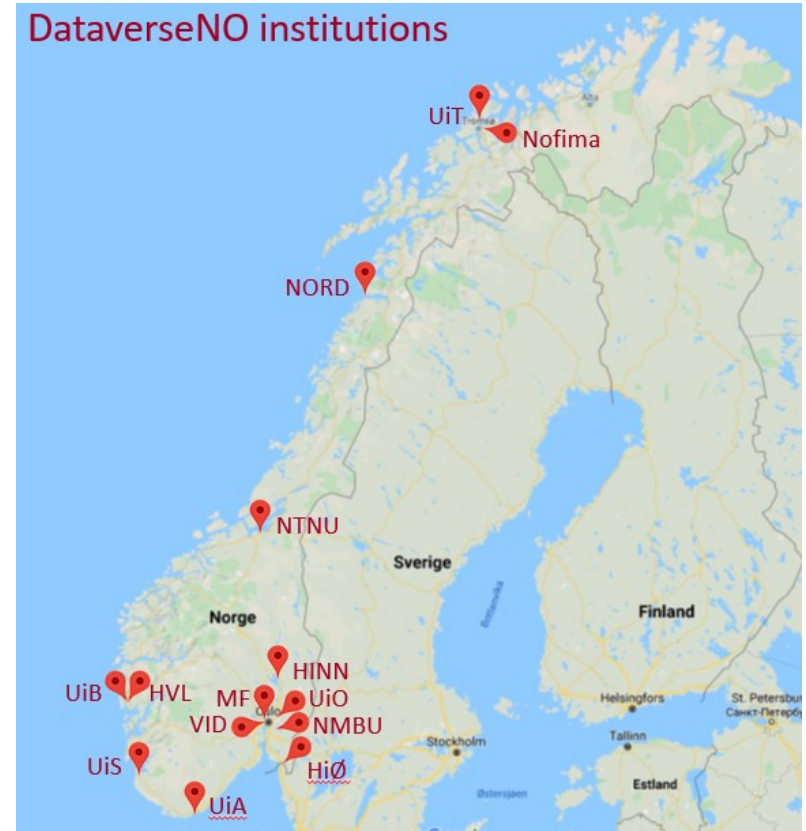
* UiT The Arctic University of Norway

** Data Archiving and Networked Services (DANS), Netherlands

Dataverse Community Meeting 2023
University of Minho, Braga, Portugal
June 5-7, 2023

DataverseNO

- ❑ a **national**,
- ❑ **generic** repository
- ❑ for **open** research data
- ❑ mainly from researchers at **Norwegian** research institutions
- ❑ currently **14 member institutions**
- ❑ operated at **UiT The Arctic University of Norway**
- ❑ As part of the strategic plan for IT infrastructure at UiT, DataverseNO was earlier this year **migrated to cloud infrastructure**.

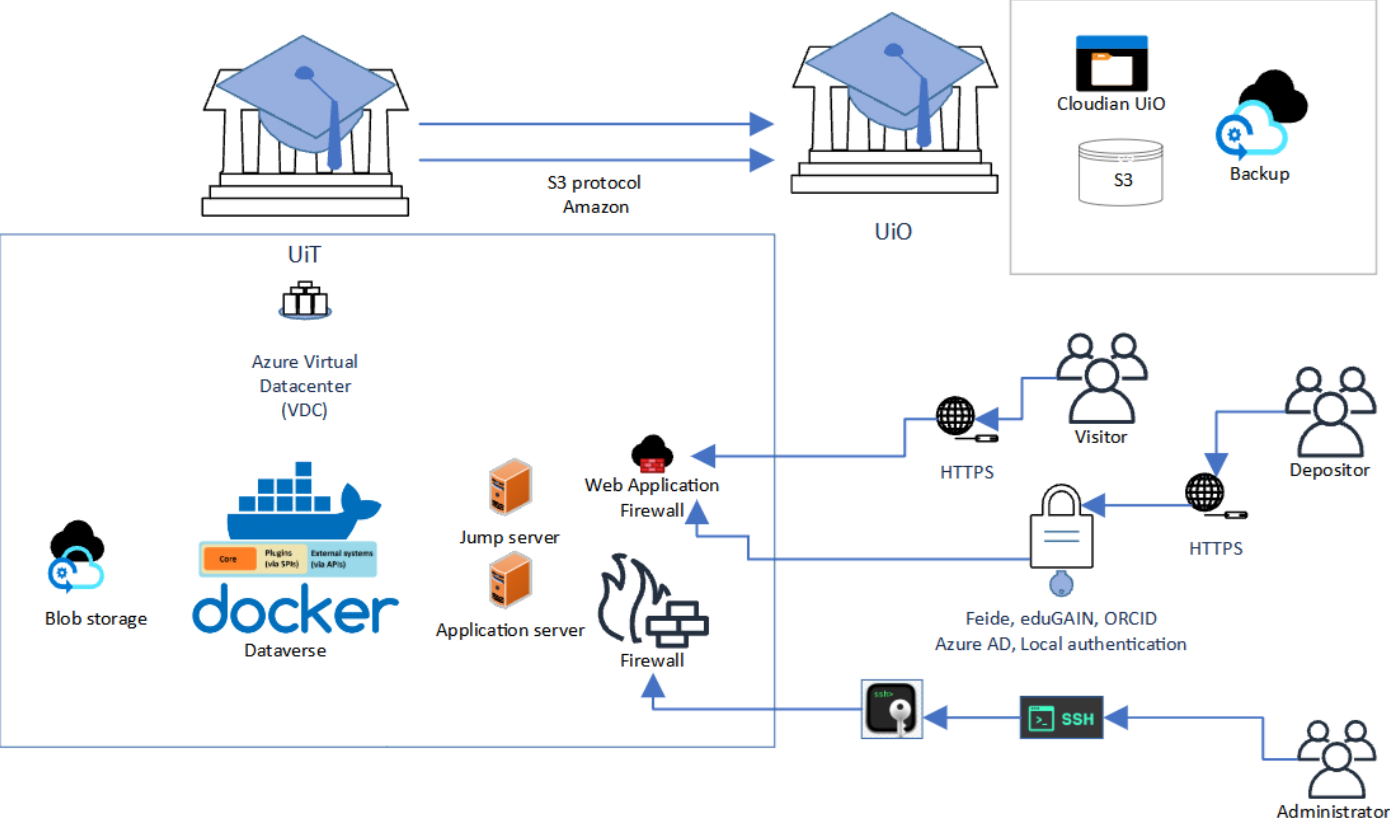


Chosen technologies, methods, and configurations

Migrated DataverseNO application now running on Dockerized infrastructure including the following features:

- Hosting ‘Archive in a box’ installation developed as a part of SSHOC project (see Tykhonov et al. 2022) and deployed in Microsoft Azure Cloud.
- Feide, a national Norwegian federated authentication service, was Dockerized and integrated with the DataverseNO Cloud instance, and keeps up-to-date affiliation of all users.
- Storage is cloud-based and S3-compatible with Cloudian storage at the University of Oslo as basic storage layer and with Azure Blob Storage as additional backup.
- Direct upload and download support with DVWebloader tool developed by GDCC and funded by DataverseNO allows to keep folder structure.
- Some customized changes were Dockerized and available out-of-the-box as part of “Archive in a box” supporting new releases of Dataverse.
- During migration the team has established setup to get overview of DataverseNO metrics.

Overview



Slightly adapted from: DataverseNO Disaster Recovery Plan

Benefits

Running the DataverseNO application based on Docker containers with external storage offers several benefits, including:

- Manual maintenance efforts are significantly reduced by running common infrastructure shared with other community members like DANS.
- DataverseNO customizations managed through github repo with some provenance info
- Upgrades to new Dataverse versions are done almost automatically.
- Docker images can be shared with other interested parties, for example, for Feide-based on Shibboleth authentication service and CloudBeaver.
- Storage layer allows to separate data for different partner institutions and keep them in the different buckets linked to their subdataverses.
- Infrastructure as a service can be shared with other partner institutions regarding to networking in Azure Cloud and other critical components setup. This allows to get extra resources from other partners running the same infrastructure in their place.
- New setup allows fast integration of other third party services within DataverseNO

Challenges

During cloud migration we met some challenges, including:

- Some complications related to keep Azure configuration in stable working state after upgrades.
- Difficulty to keep all affiliations up-to-date within Feide and local users
- In some cases, it's problematic to link user accounts between different authentication types.
- Difficult to diagnose errors/failures when uploading many/large files to Cloud storage.

Future plans

DataverseNO is planning to improve the current cloud deployment with features like the following:

- Run DataverseNO on Kubernetes platform in more sustainable setup
- Further improve CI/CD workflow and automated testing for integrations
- Implement immutable storage backup as extra security layer
- Implement metrics portal with more advanced options for curators, depositors, and other stakeholders

Thank you for your attention!

References

Wittenberg, Marion, Vyacheslav Tykhonov, Eko Indarto, Wilko Steinhoff, Laura Huis in 't Veld, Stefan Kasberger, Philipp Conzett, Cesare Concordia, Peter Kiraly, & Tomasz Parkoła. (2022). D5.5 'Archive in a Box' repository software and proof of concept of centralised installation in the cloud. Zenodo.

<https://doi.org/10.5281/zenodo.6676391>

More info on DataverseNO



dataverse.no



@DataverseNO



info.dataverse.no