Contribution ID: **85** Type: **Long Talk**

Sunet Drive - An Academic Toolbox for FAIR Data Storage, Analysis and Publication

Wednesday, 2 October 2024 12:15 (15 minutes)

Sunet Drive is a national file storage infrastructure for universities and research institutions in Sweden. It is based on a Nextcloud setup and is comprised of 54 nodes, one prepared and provisioned for each institution. The aim of Sunet Drive is to become an **Academic Toolbox** capable of collecting, storing, analyzing, and publishing research data, supporting FAIR principles. We present Sunet Drive as an integrated solution comprised of four essential building blocks:

- File sync and share based on Nextcloud and S3 as the underlying storage entities
- eduGAIN login using SeamlessAccess and added security through step-up authentication and MFA zones
- Scalable JupyterHub integration for flexible and reproducible data analysis
- Research Data Services for easy publication to public and curated repositories

Participating organizations co-manage their Sunet Drive node as part of a global scale setup, meaning that every node is governed by the operating organization, while being able to collaborate and share data with users within the federation, but also external partners through open cloud mesh protocol (OCM), such as the ScienceMesh. S3-compatible buckets are used as logical storage entities that can be assigned for different purposes: research projects, institutions, laboratories. They are technically independent from the EFSS layer and their life-cycle can be managed beyond the lifetime of the selected EFSS software, an important step towards long-term sustainability of FAIR data.

Collaboration is encouraged by allowing access through eduGAIN and subsequently accept documents, shares, and data from their collaboration partners. External collaboration is enabled via Eduid.se. Added security can be provided through step-up-authentication, adding a second authentication factor for identity providers that have not added support for 2FA yet. Further security can be added by activating MFA-zones, mandating the receiver of a file or folder to add a second authentication factor, such as TOTP or a WebAuthn device.

During the runtime of a research project, data can be processed and analyzed directly through a scalable JupyterHub integration, an open source application developed by Sunet and funded by the GÉANT Project Incubator. Compute resources are intelligently managed in a kubernetes environment and can be allocated on a per-project basis, which includes support for CPU and GPU flavours.

The integration of Research Data Services, RDS, enables the preparation and direct publication of datasets directly. This includes services like InvenioRDM (e.g., Zenodo), Harvard Dataverse, or Doris from the Swedish National Dataservice, SND. Research object crates (RO-Crate) are used as an intermediate lightweight package for the data, and respective metadata, connectors ensure compliance with each publication service. Domain-specific customizations include the integration of different publishing paradigms: While data is being actively pushed to repositories such as InvenioRDM or OSF, the SND Doris connector uses a more lightweight approach where the metadata is pushed to Doris, with the data storage remaining under the sovereignty of the publishing institution.

Providing researchers with an Academic Toolbox with streamlined support for authentication, data management, analysis, and publication helps to ensure compliance with local, national, and international guidelines for storing of research data, including FAIR principles.

Topic

Needs and solutions in scientific computing: Platforms and gateway

Primary author: FREITAG, Richard (SUNET)

Co-authors: Mr ANDERSSON, Magnus (SUNET); Mr NORDIN, Micke (SUNET)

Presenter: FREITAG, Richard (SUNET)

Session Classification: Maximizing Data Efficiency and Availablity: Innovations in Data Management Platforms