Contribution ID: **73**   Type: **Presentation**

Sunet Drive - An Academic EFSS Packaged for EOSC

Monday 6 March 2023 16:30 (15 minutes)

Sunet Drive is a federated and scalable Enterprise File Sync and Share solution, that has been developed, deployed, and packaged as part of the European Open Science Cloud and can be transparently extended to new participating organizations. The two main building blocks of Sunet Drive are Nodes and Buckets, both elements designed to promote data sovereignty and FAIR principles. 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 that support the open cloud mesh protocol (OCM), such as the ScienceMesh. New organizations have been and can be onboarded by migrating existing provisioned users to a full node associated to an organization or institution. Buckets, specifically 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 lifecycle can therefore be managed beyond the lifetime of the selected EFSS software, an important step towards long-term sustainability for FAIR handling of data.

The infrastructure stack is implemented in collaboration with the commercial actor Safespring, and data generally resides in at least two different data centers. This ensures a scalable stack built on best practice open source components, together with experience in running large scale deployments. Certification standards such as ISO 27001 guarantee a mature handling of the infrastructure and data in the solution.

By having chosen a state-of-the-art EFSS solution for Sunet Drive, researchers, scientists, and their collaborators can align the requirements of their funding body and associated data management plans with their primary data sources by using modern like synchronization clients and mobile applications. On the other hand, integrated and connected services ensure that scientists will be able to collaboratively work on their projects without having to leave the ecosystem.
Collaboration is encouraged by allowing any eduGAIN connected identity provider to provision user accounts, and subsequently accept documents, shares, and data from their collaboration partners. The lack of support for a discovery service on the EFSS side has been solved by using a global site selector through a SaToSa proxy that delegates users to their respective Sunet Drive Node. External collaboration is enabled via Eduid.se.

During the runtime of a research project, research data can be curated and prepared for publication. The integration of Research Data Services, RDS, enables the preparation and publication of datasets directly from the EFSS solution. This includes external services like InvenioRDM (e.g., Zenodo), Harvard Dataverse, or Doris from the Swedish National Dataservice, SND. This includes domain-specific customizations. While research object crates (RO-Crate) are used as an intermediate lightweight package for the data, and respective metadata, connectors will ensure compliance with each publicatoion service. While data is being actively pushed to InvenioRDM, the Doris connector uses a more lightweight approach where the metadata is pushed to Doris, while the data storage remains under the sovereignty of the publishing institution.

Having data stored in S3-compatible buckets associated to a federated EFSS node managed by specific organizations ensures data sovereignty and helps to ensure compliance with local, national, and international guidelines for storing of research data, including FAIR principles. After a project has finished, ownership of the data can be transferred during the data retention period, and for long-term archival purposes.

Authors: Mr PAUES, Gabriel (Safespring); Mr FREITAG, Richard (SUNET)

Co-author: ABEN, Guido (SUNET)

Presenters: Mr PAUES, Gabriel (Safespring); Mr FREITAG, Richard (SUNET)

Session Classification: Site Reports

Track Classification: Main session: CS3 Community Site Reports