SUNET

Sunet Drive

2021-10-14

SUNET

Agenda

General and background

- Research data – a model
- Data Integration
- Nodes and buckets
- Sunet Drive Overview
- Follow your data
- Single Sign On
- Own your data
- MSBFS2020:7
- Pricing

- drive.sunet.se
  - What do I get?
  - Separation of concerns
  - Onboarding process
  - Sunet Forum
  - Future
- Sunet Drive – Now and Then
  - Agenda 2021-10-27
  - Storage incident (for the curious)

SUNET

Anecdote

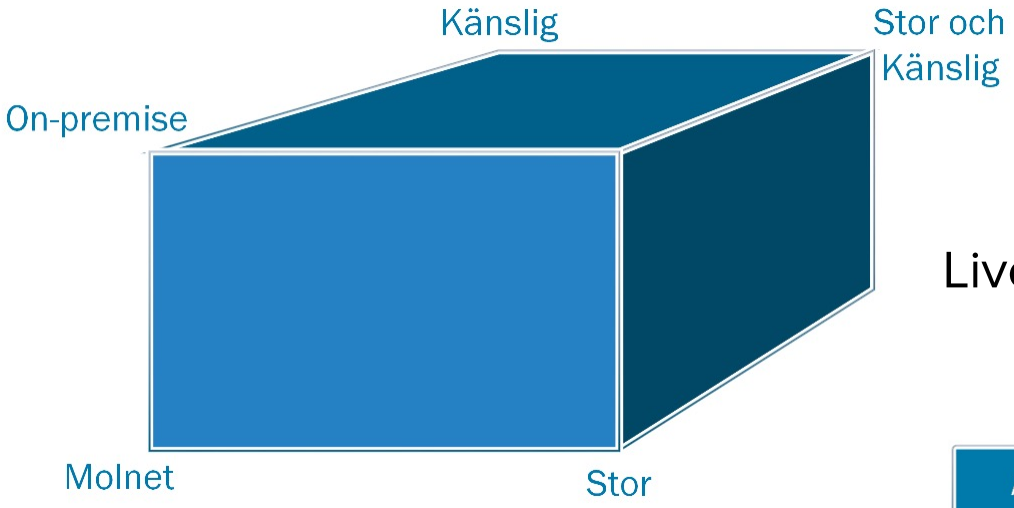
Thanks for the detaljer information! I will be ending my employment this Saturday, so am looking for some solution where the data will not be deleted due to the deactivation of my akka id and where it would be accessible by my PI in the future.

Wednesday, 10:51

SUNET

Research Data – A model

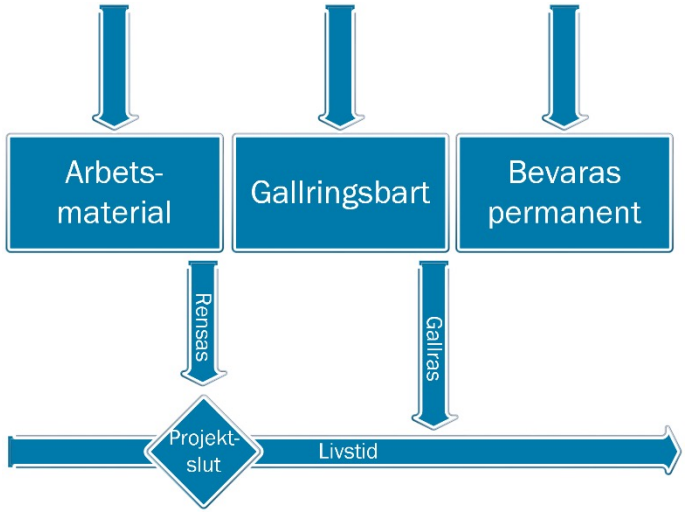
Classification



How do you access data?

Where is my data?

Livecycle



Access

"Klassisk" Mappar

"Modern" Appar

SUNET

Research Data – A model

Classification

Känslig

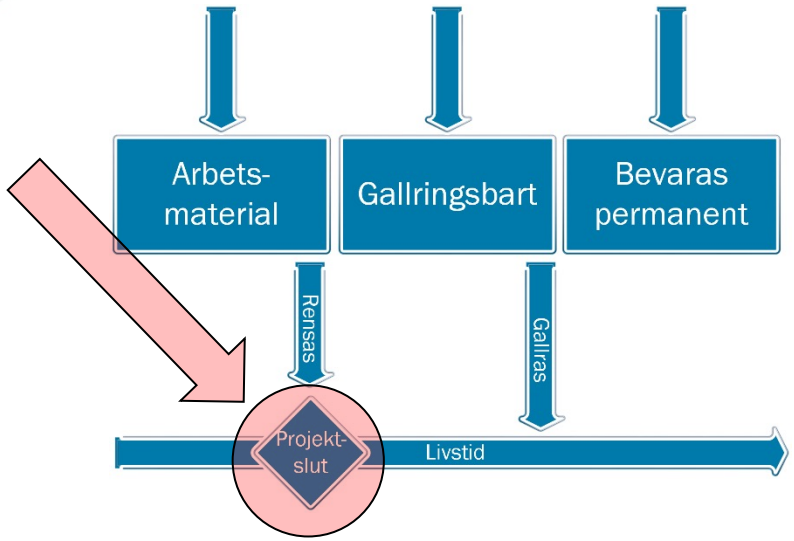
Stor och Känslig

On-premise

How do you access data?

Where is my data?

Livecycle



This is where the money ends!*

Molnet

Stor

Access

”Klassisk” Mappar

”Modern” Appar

*for many projects

SUNET

Example - Integration

Primary data sources

UV-Vis

IVIUM

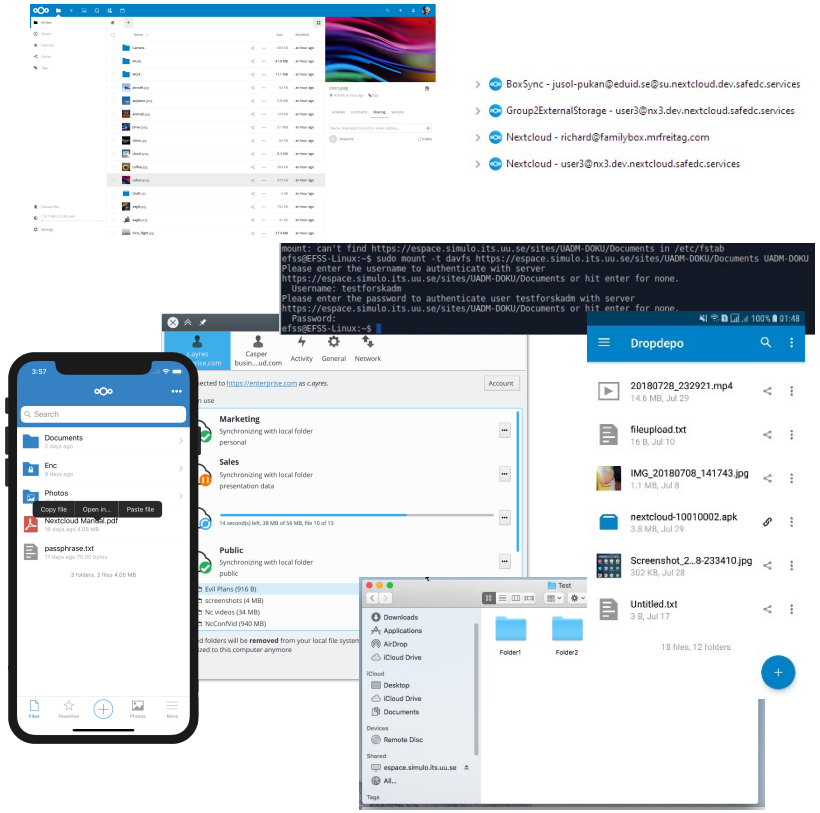
...

PIA

IPCE

Storage (S3)

Access



SUNET

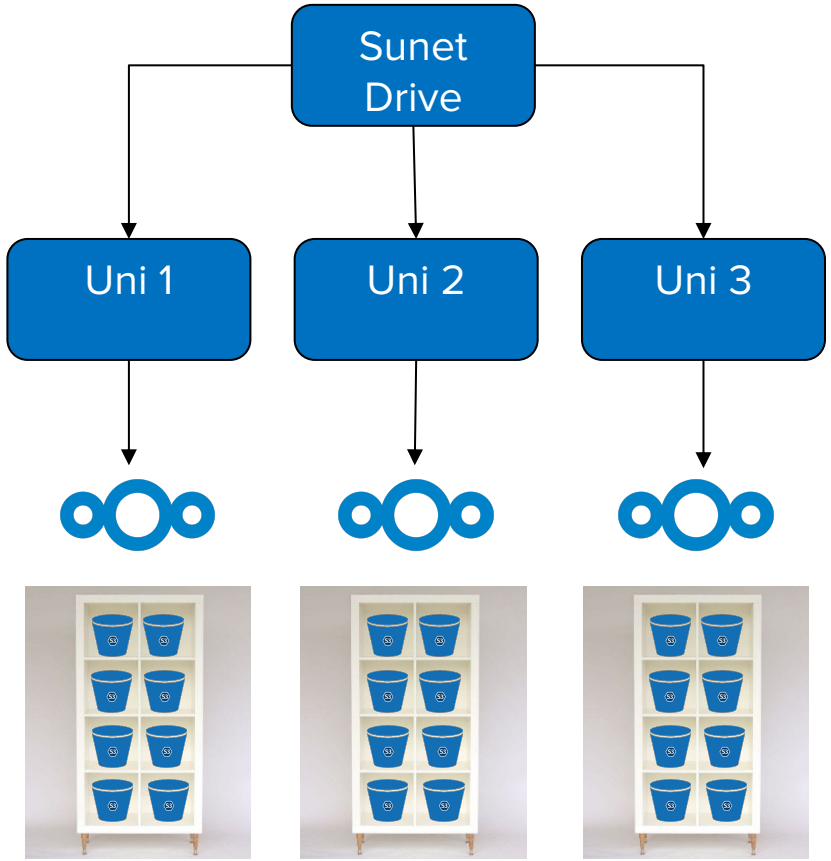
Nodes and Buckets

- Main storage unit is an S3-bucket
- A bucket is comparable to a virtual hard disk
- NextCloud is the access layer to the bucket
- Individual buckets represent ”storage entities” (research project, lab, institution)
- Core-data resides in the bucket independent from NextCloud
- Buckets are (relatively) agnostic and can be migrated easily

SUNET

Sunet Drive - Overview

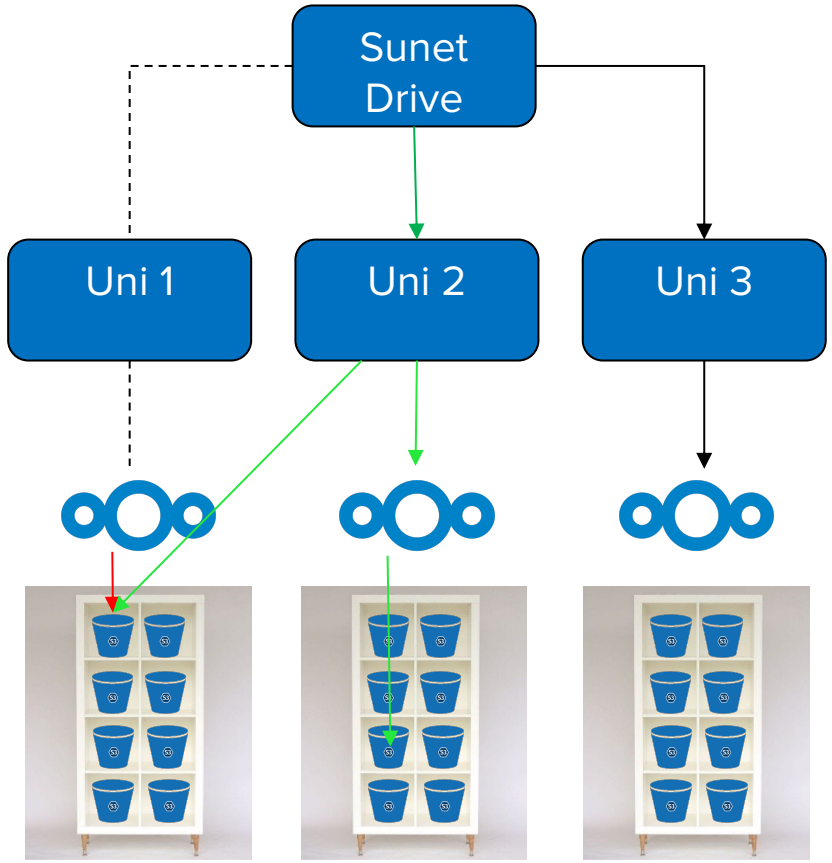
- Every university gets their own NextCloud node
- The node is part of a ”global scale” and can collaborate with other nodes
- The node is co-managed with Sunet
- **Separation of access vs ownership**



SUNET

Follow your data

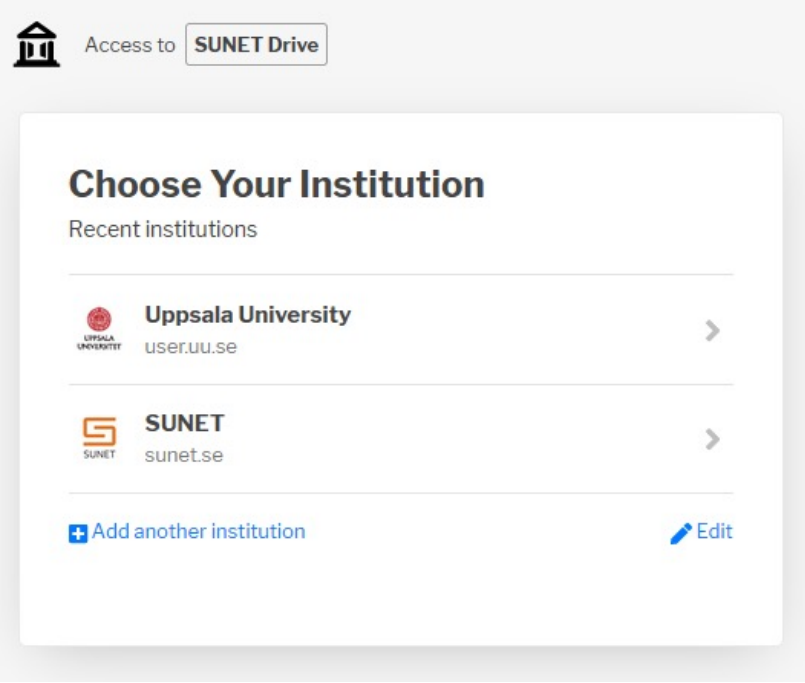
- Most researchers change institutions frequently
- Many researchers take their research data with them
- With NextCloud you can follow your data even from a new organisation



SUNET

SWAMID/eduGAIN/SeamlessAccess

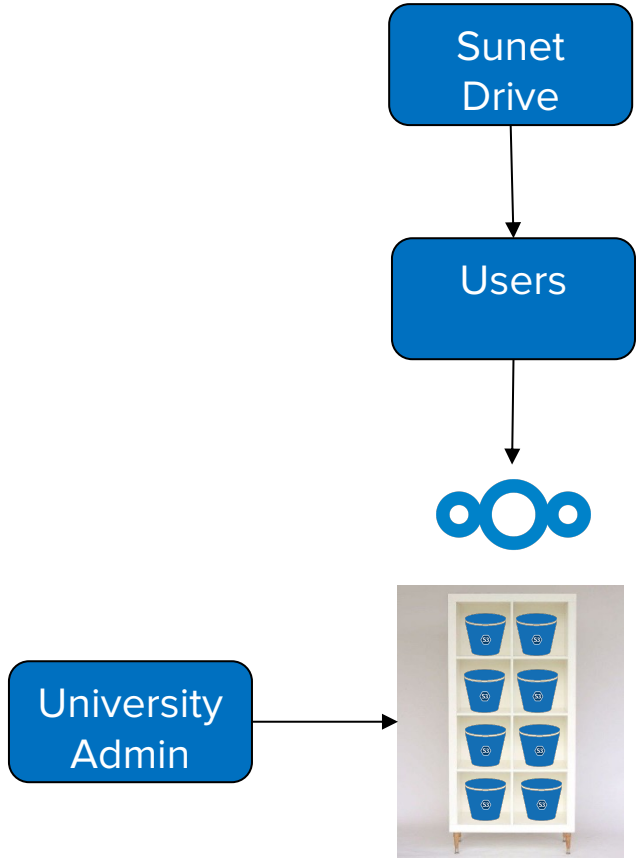
- Single Sign On using your organization or institution account
- Nextcloud lacks support for Discovery Service
- Use SaToSa for SAML2<->SAML2
- Users can log on using any federated provider using Seamless Access
- ~6000 IdP world wide
- Easy collaboration
- **Follow your data (as a researcher)**
- **Own your data (as an institution)**



SUNET

Own your data

- Universities administrate their own storage (buckets)
- Universities can create as many buckets for research projects or labs as needed and integrate them into NextCloud
- Existing data can be migrated into buckets and indexed by NextCloud
  (i.e. external hard disks or other ”Shadow-IT)



SUNET

Private Cloud Security

- Personnel control
  - Background checks
  - Data center staff
  - Trusted consultants
  - Third party vendors
- Data center security
  - Availability
  - Fire Protection
  - Facility Access Controls
  - Physical Intrusion Controls
- Network Access Controls
  - Public Network Access
  - Management Network Access
- Systems Management
- Auditing
- Process Description and Configuration Management
- Infrastructure & Storage as a Service
  - Iaas high level architecture
  - Network exposure management
  - IaaS/STaas – Auditing
  - Data Privacy
- Third Party Review

https://www.safespring.com/documents/sunet/safespring-sunet-private-cloud-security-controls.pdf

SUNET

MSBFS 2020:7 och Sunet Drive

Paragraf	MSBFS 2020:7 Text	Sunet	Kundens eget arbete	Referenser

Beskrivning av alla punkter i MSBFS 2020:7

- Ansvar av Sunet
- Kundens eget arbete
- Referenser

https://wiki.sunet.se/display/Drive/MSBFS+2020%3A7+och+Sunet+Drive

SUNET

Pricing

Post	Antal block om 500 anv. licenser	SEK/månad
Grundpris	---	24 800
Instans-pris (beroende på antalet användare i block om 500)	1	3 190
	2–3	6 380
	4–11	15 950
	12 -->	31 900
Per användarlicens (*) (minsta antal 500, köps i block om 500 st.)		2.00
Tillval:		
Lagring för kunder som **inte** har köpt tjänsten ”Sunet lagring i privat moln”. Minsta volym 5 TB, köps i block om 5 TB.		10 öre per GB/månad

SUNET

Pricing Examples

License blocks	Licenses	Base price	Instance price	License price	Total/Month	Per User/Month
1	500	24800	3190	1000	28990	58*

Small institution

License blocks	Licenses	Base price	Instance price	License price	Total/Month	Per User/Month
7	3500	24800	15950	7000	47750	14*

Medium institution

License blocks	Licenses	Base price	Instance price	License price	Total/Month	Per User/Month
13	6500	24800	31900	13000	69700	11*

Large institution

(*) plus storage

SUNET

https://drive.sunet.se

(*production ready, requires onboarding)

SUNET

Onboarding Process

Sunet Drive onboarding means that your institution has decided to move from a free minimal node to a full node. The onboarding process usually consists of three phases:

- Phase 1: Sunet will set up a dedicated node in the Sunet Drive Pilot and train the institution on using and administrating their own node
- Phase 2: The institution will test the dedicated pilot node with a limited number of users, including node-administration and administration of S3-buckets
- Phase 3: Sunet sets up a full node in the production environment, where the institution implements their specific processes and routines

Phase 1: Sunet Drive Pilot Node

A dedicated node will be set up for the institution, and the institution will be trained on how to administrate their own node. This mostly includes user-management, as well as management of the underlying S3-storage. During this phase, the institution has the possibility to decide on their level of integration with their own local IT-architecture. In its most simple manifestation, this could mean that users will only be managed through the institutions' identity provider, and S3-storage will be assigned through a simple application process (e.g, via mail). If a deeper integration and automation with the local IT-architecture is and automation is desired, it is recommended to be done during this phase.

Phase 2: Test with a limited number of users

A limited amount of users and administrators will get official access to the pilot node, in order to test, evaluate, and give feedback. This includes a limited amount of S3-buckets that can be used for data storage (and potentially be migrated later). Ideally, the institution selects a representative group of users that are open to being early adopters, and give constructive feedback on the implementation. During this step, the institution will also define their backup-requirements and go through the necessary processes.

Phase 3: Roll-out for all users

End-users will now have the possibility to apply for storage through their institution, based on the level of integration determined during the first two phases.

https://forum.sunet.se/s/sunet-drive/wiki/page/view?title=Sunet+Drive+onboarding+process

SUNET

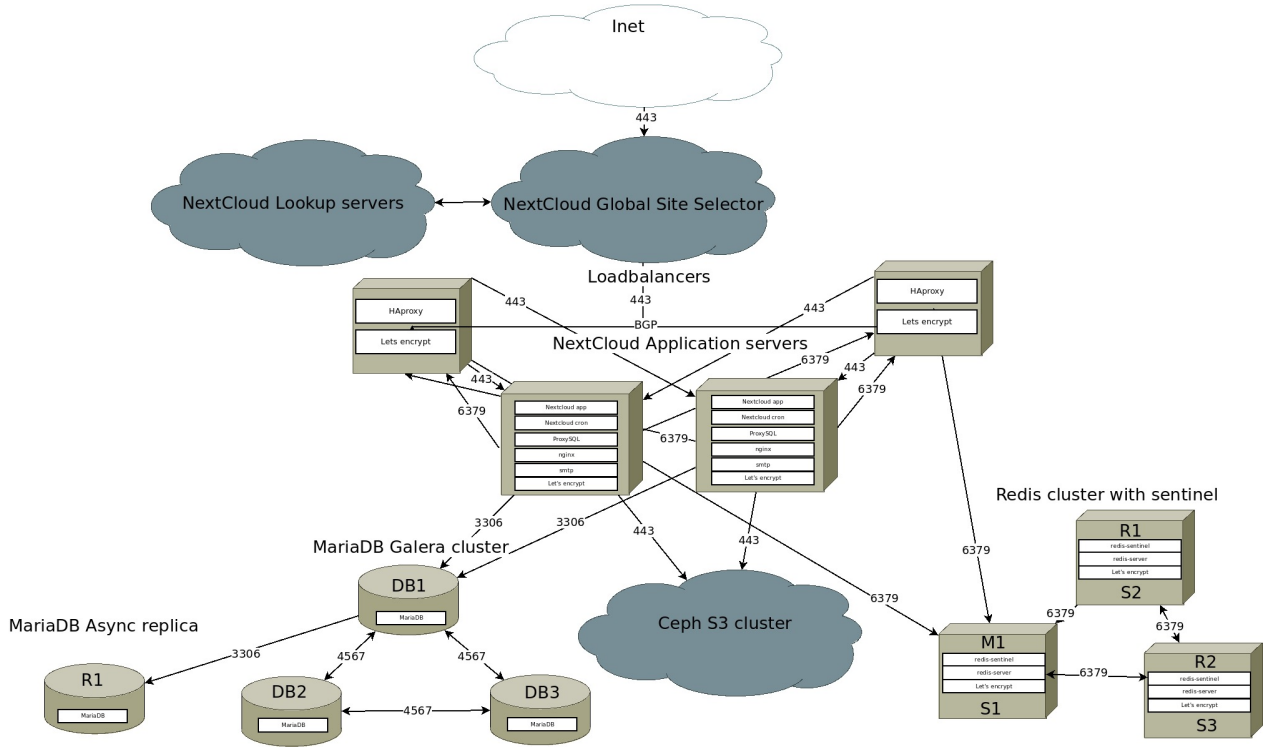
drive.sunet.se

Paying universities get an automatically provisioned node consisting of:

- 2 Load-balanced Sunet Drive Frontend Servers
- 3 Load-balanced clustered database servers
- Redis-caching for improved performance
- S3 storage integration
- Access to test environment
- Optional storage-mirroring
- Customer support

Non-Paying universities get:

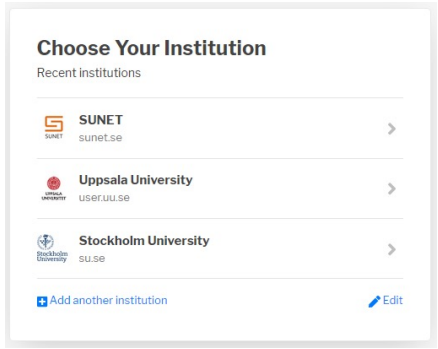
- Access via a shared node
- Provisioning of user accounts
- ⬇The option to upgrade to this⬇ ☺



SUNET

drive.sunet.se - nodes

SWAMID/eduGAIN/Seamlessaccess

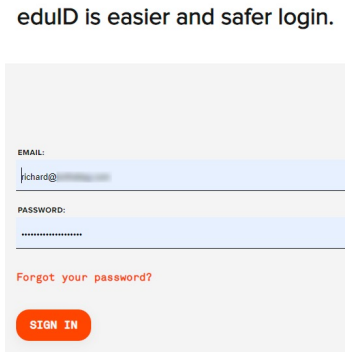


sunet.drive.sunet.se

uu.drive.sunet.se

su.drive.sunet.se

EduID



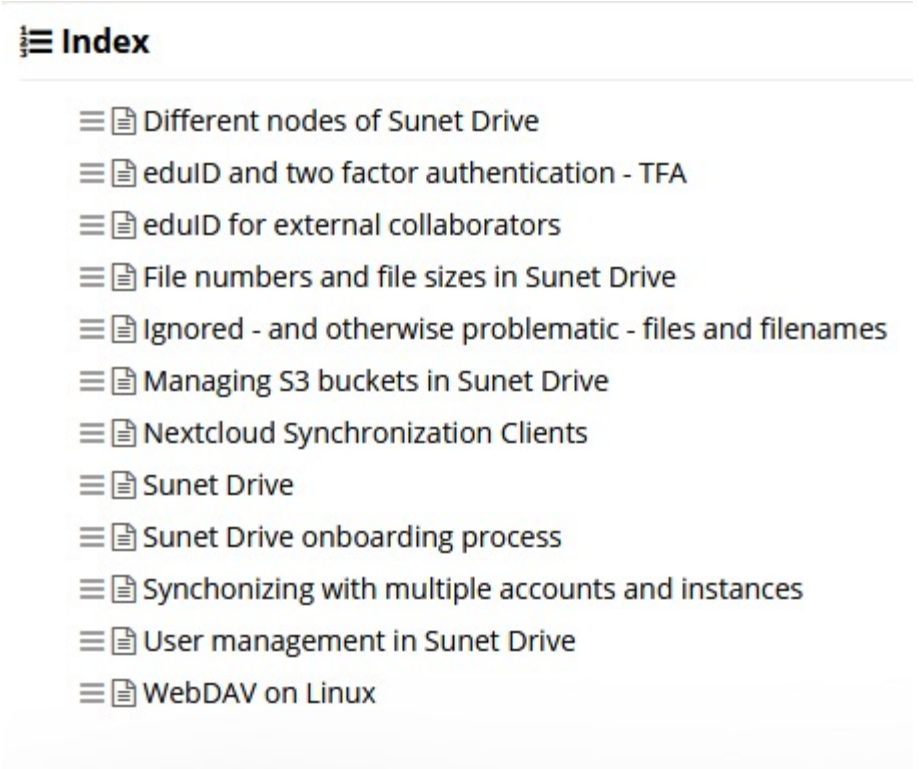
extern.drive.sunet.se

SUNET

Sunet Forum

- Central point of documentation for Sunet Drive
- Information Stream
- Files
- Wiki
- Information for users and administrators

https://forum.sunet.se/s/sunet-drive/



SUNET

Future

Beskrivning	Status	Prioritet (1 är högst)
Styrning av begärd assurance-nivå på objekt-nivå. Målet är att kunna (tex) ställa in att åtkomst till en katalog kräver (tex) AL3-inloggning och MFA.	Analys	1
Integration med SNDs Doris-portal för forskningsmetadata. Målet är att kunna underhålla metadata för forskningdata för publicering i Doris direkt i Sunet Drive.	Beslutad	2
Regelbaserad övervakning av åtkomstloggar. Målet är att kunden ska kunna definiera regler och triggers för övervakning av åtkomst till filer via nextcloud loggar. Den sannolika målbilden handlar om att etablera en loggtjänst i Sunet som kan erbjudas som komplement till Sunet Drive för de kunder som inte har egna loggtjänster med de tekniska förmågor som krävs för regelbaserad loggövervakning.	Analys	
Office-integration med någon av collabraoffice, onlyoffice och/eller ms office.	Beslutad	
Analysresurser - tex i form av k8s kluster eller jupyter notebooks. Målet skulle vara att erbjuda möjlighet till enklare/mindre resurskrävande data-analys på data lagrat i tjänsten. Ej att förväxla med vad SNIC gör där fokus är på resurskrävande beräkningar. Målet för denna funktion skulle snarare ligga på enklare och mindre resurskrävande analys.	Analys	
Integration med edusign. Målet är att kunna signera dokument direkt i Sunet Drive.	Inkommen	

Sunet Drive: Now and then – 27 oktober 2021

SUNET

Sunet Drive – Now and Then
Agenda – 2021-10-27

- Multifactor Authentication
  - Why it sounds easy, but is hard to implement
  - Nextcloud step-up authentication
- Integration with SND/Doris
  - How would an ideal integration look like?
  - Comparison with other services (Harvard Dataverse, Zenodo)
- Office-integration
- Other features
  - Monitoring
  - Compute-integration
  - edusign
- Storage incident
  - What happened during the incident?
  - Restoration of data
  - Disaster recovery
  - Backup, backup, backup

SUNET

Storage incident

Summary

On the morning of Thursday, the 17th of June, Sunet Drive reported a lot of gateway timeouts (504) for the object storage in Sto4. The problem was reported at 10:13 CET to Safespring via their support portal. Subsequent investigations led to the conclusion that another customer was unintentionally causing a larger than expected load on the object storage, which as a result had to be taken offline. A detailed technical description of the incident can be found in the document “2021-06-17 sto4 ceph cluster”. As a result, instances of Sunet Drive using Sto4 as their storage backend had to be taken offline and prioritized buckets/customers are at the time of writing waiting for the results of the restore initiative from Safespring.

A main reason for the severity of the incident was the decision to assign backup responsibilities to the customers and end-users of Sunet Drive. This led to the situation where the only copies of certain files resided solely in the affected object storage.

Before the incident, the cluster contained about 9 million objects. When the incident occurred, another customer had uploaded 680 million objects.

SUNET

If you are interested

- Contact anders@sunet.se or freitag@sunet.se
- Test-run using pilot.drive with S3-buckets and more data
- Provision nodes for test and production
- Regular feedback- and training-sessions

SUNET

Tack! Frågor?

Anders Nilsson

Kontaktansvarig för tjänsterna Backup, Lagring och Virtuella servrar

anders@sunet.se

Richard Freitag

Projektledare

freitag@sunet.se