

Federated scalable cloud storage with CERNBox

Giuseppe Lo Presti
On behalf of the CERNBox team

4 December 2025
16th SIG-CISS Meeting on Federated Storage, Bologna, Italy





Outline

- What is CERNBox
 - Main features from a user's perspective
 - A quick dive in the technology
- The Service at CERN:
Scale and diversity of usage
- The Community and the Federations





What is CERNBox?

CERN's cloud collaboration storage platform
providing 1 TB for every user, up to 100 TB on demand

Fully open-source

- Powered by Reva
- Storage: EOS and Ceph
- Sync clients: ownCloud
- Web: collaboration with ownCloud



Hub for **scientific collaboration**



Integrates with OSS components:
Jupyter, SAMBA, HTCondor, ... as
well as Office



Data sovereignty: **safely and
securely stored on premises**



At CERN, it supports **4.1B files, 12
PB of data from 27k people around
the globe**





Global namespace: what makes CERNBox different

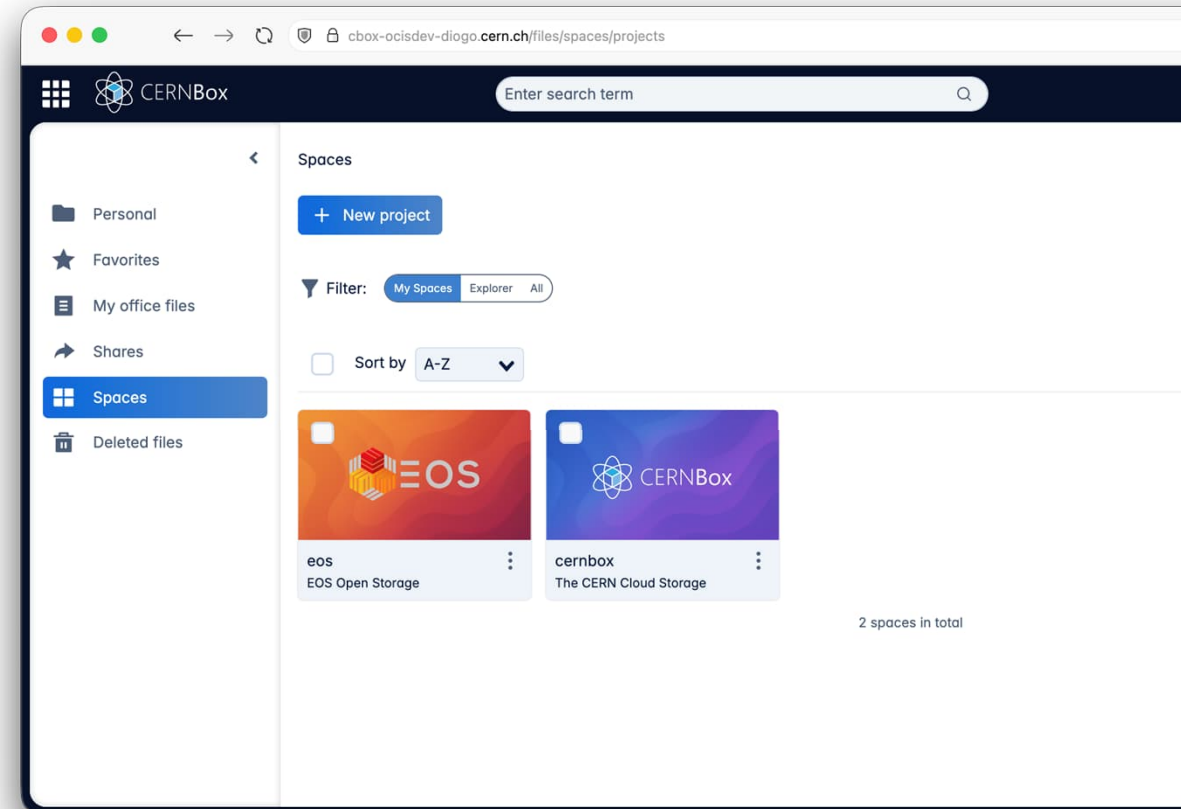
Permissions are set in the storage and respected independently of the protocol.
There's a consistent view no matter the access.





Private and Collaboration Spaces

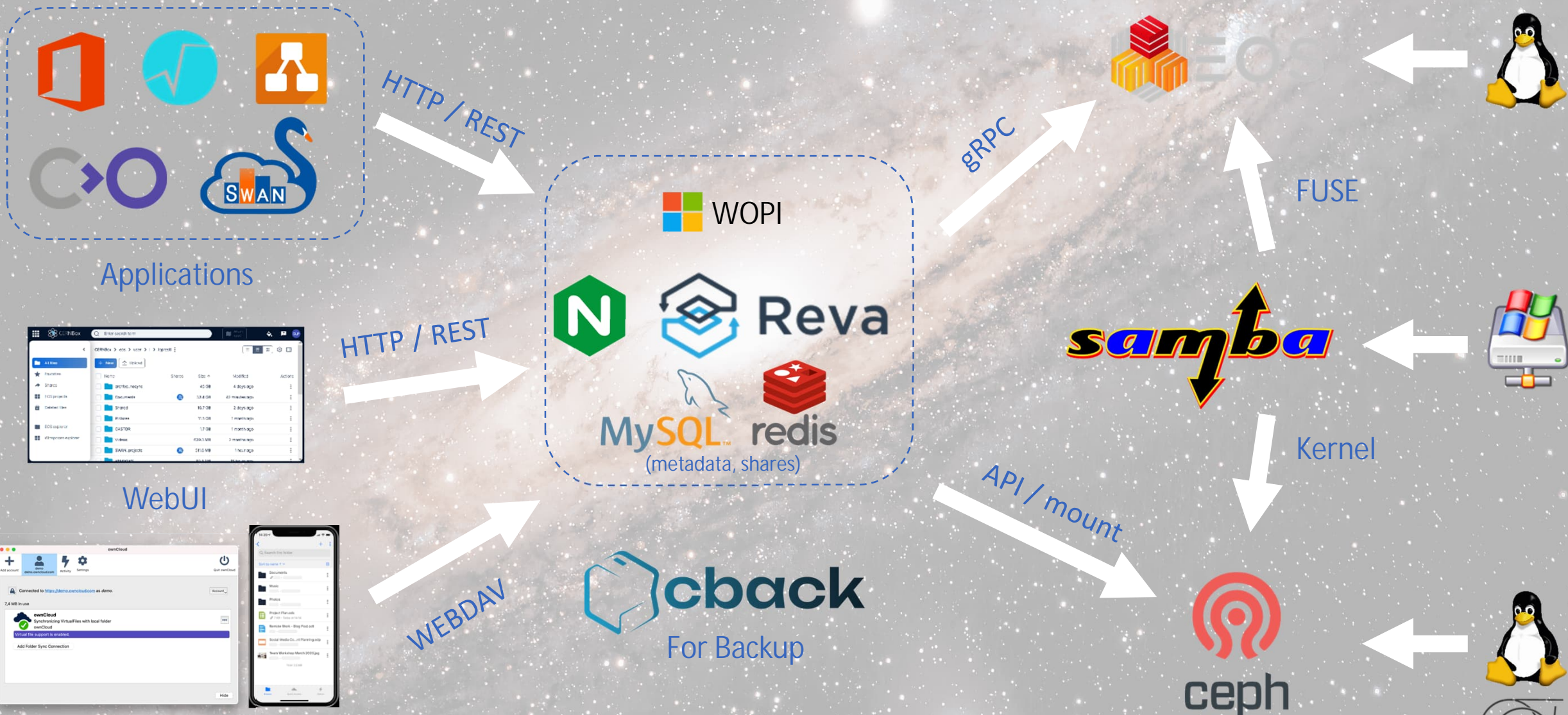
- **Private Space**
 - Your personal area, visible only to you.
- **Project Spaces**
 - For teams, with roles (Admins, Editors, Viewers).
 - Admins manage membership & sharing.
- Both spaces allow sharing content with others



The Technology and the Service at CERN



The CERNBox Galaxy



Sync Client / App





The Technology

- Architecture paradigm: **microservices**
- Multiple components and daemons, each responsible for (a simple) part of the service, easing horizontal scalability
 - Most parts are distributed, metadata is kept on the storage
 - A MySQL database and a Redis cache are used to optimize internal queries
 - Technologies of choice: [Golang](#) and [Python](#), [Protobuf](#)





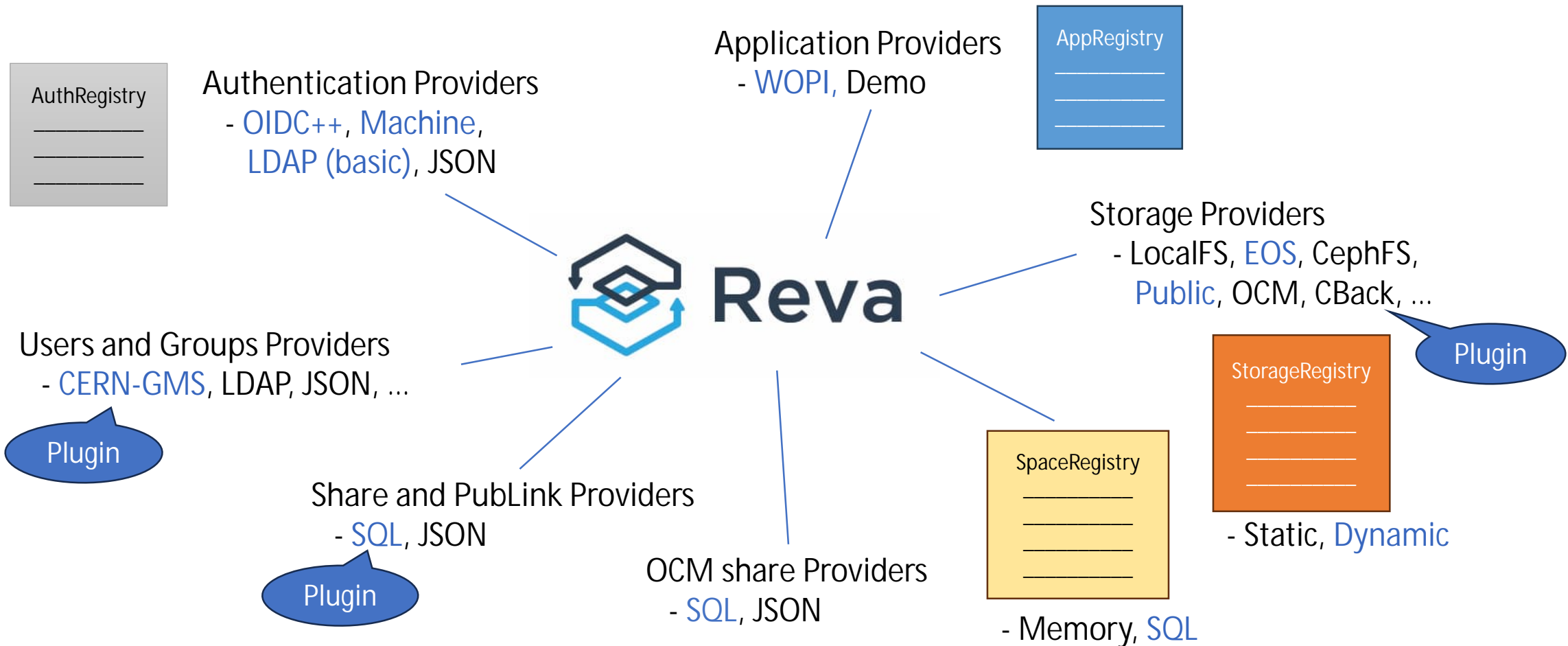
Reva: the CERNBox back-end

- Bridge between high-level clients and the underlying storage
 - Supports [EOS](#) (OSS storage system developed at CERN), [CephFS](#) and any POSIX local FS
- Exports WebDAV, LibreGraph, and a high-performance gRPC API (CS3APIs)
 - The [CS3APIs](#) are maintained by CERN in collaboration with ownCloud and OpenCloud
- Currently there are three independent implementations of [Reva](#)
 - Powering **CERNBox**, **oCIS** (ownCloud), and **OpenCloud**





Zooming into Reva



Blue = used in production





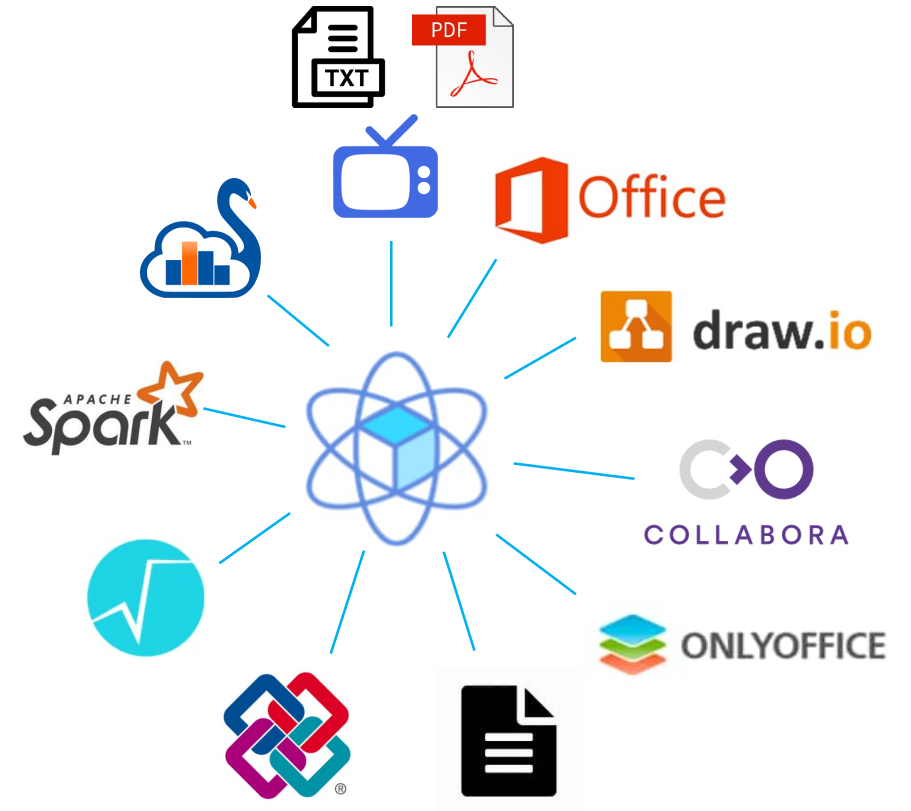
Applications

- Internal, frontend-only

- ROOT viewer
- Python notebook viewer
- 3D CAD renderer
- Text editor
- Draw.io
- Media viewer

- External

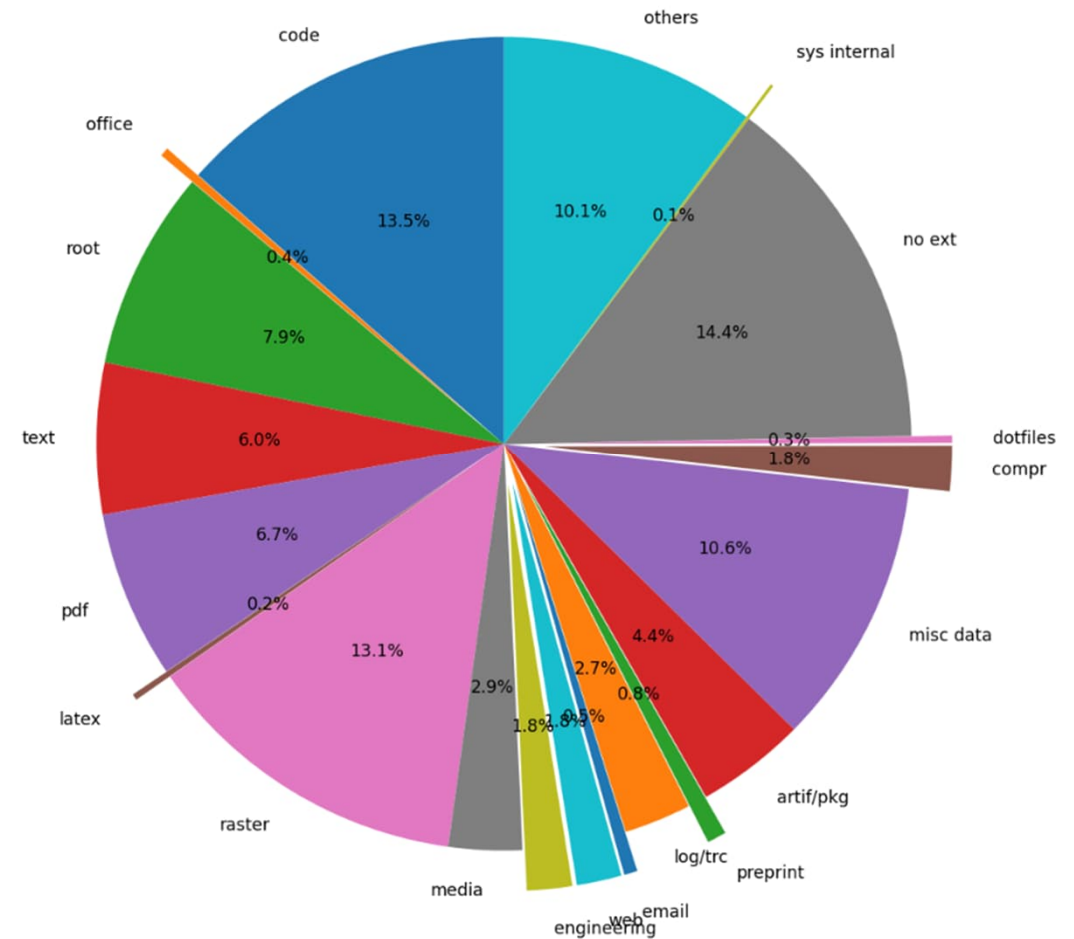
- MS Office via WOPI + proxy
- CodiMD
- Collabora
- Open in SWAN





A large variety of file types

- ~10K different file extensions
- ~12M (0.4%) Office files
- Most common extensions: ROOT, PDF, TXT
 - 280M ROOT files
- Lots of code repositories
- Vast majority is computer-generated data



The Community and the Federations







The EOSC Federation build-up

- Recent efforts driven by the EOSC Symposium 2025 culminated in a working mesh of sites
- More work to come, cf. OCM presentation later

EOSC Node | CERN Physical Sciences & Engineering


Linking bridge 

Linking to the File Sharing EOSC Federation


Sync & Share Services for Science
CERNBox is a collaboration space for **science and engineering** for the global High Energy Physics community, with federation capabilities based on **OpenCloudMesh** 


Contributing to Digital Sovereignty in EU
 CERN leads the **OpenCloudMesh standardization effort with the IETF**, in collaboration with other community actors (SUNET) and the EOSC Build-up Groups.


Significant impact on the Research & Education sector in Europe: the CS³ community
 The Cloud Sync & Share Storage community represents **hundreds** of installations in 30+ sites, **servicing 300K+ users** across Europe.


CERNBox 

EOSC Federation
 Data and File Sharing

EOSC Node 

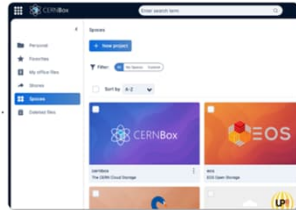
EOSC Node 

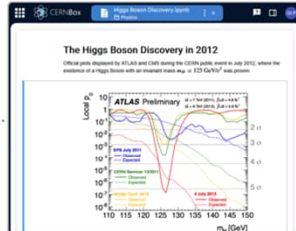
EOSC Node 



EOSC Node 

EOSC EU Node
 Your Gateway to Open Science

2025 Achievement: interconnected EOSC Nodes

Personal data management system 

Integrated in data-processing and institutional workflows 





The ScienceMesh

- An application integrated in the CERNBox web UI to manage *invites* and *contacts*
 - Leveraging the **OCM Invite Flow**: users exchange an invite, the underlying EFSSs exchange their identities
 - The same app is available in ownCloud and OpenCloud

The screenshot shows the CERNBox web interface. At the top, there's a search bar and a user profile 'GLP'. The main content is divided into three sections:

- Invite users**: Contains a '+ Generate invitation' button and a message 'You have no invitation links' with a person icon and an arrow.
- Accept invitations**: Contains an 'Enter invite token' input field, an 'Institution: -' dropdown, and a '+ Accept invitation' button.
- Federated connections**: A table with columns 'User', 'Email', 'Institution', and 'Actions'. It lists five users with their respective email addresses and institutions, each with a 'Delete' action.

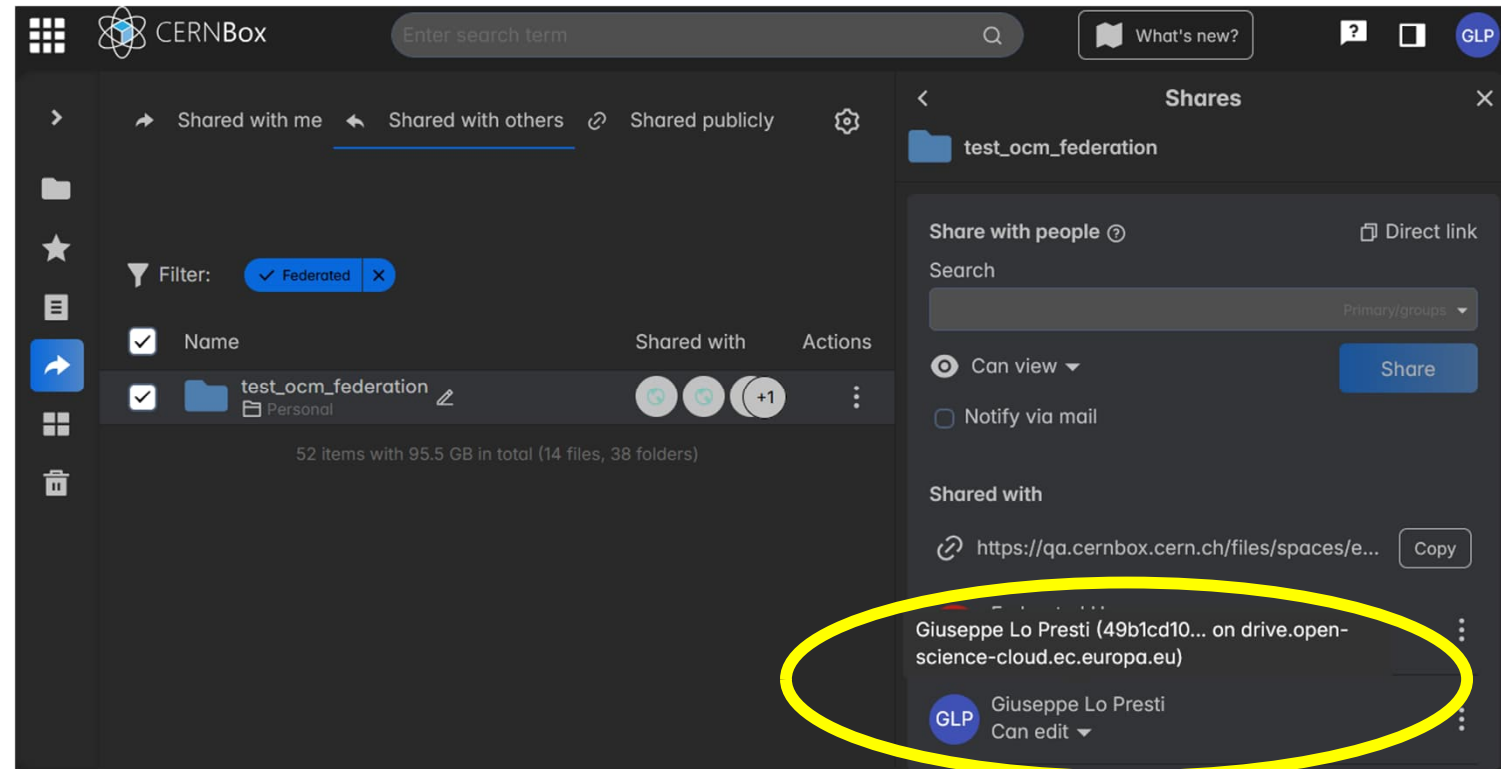
User	Email	Institution	Actions
Maurice Moss	moss@example.org	ocm2.eo.owncloud.works	Delete
Giuseppe Lo Presti	giuseppe.lopresti@cern.ch	drive.open-science-cloud.ec.europa.eu	Delete
lopresti	giuseppe.lopresti@cern.ch	sunet-eosc.drive.test.sunet.se	Delete
Micke Nordin	kano+micke@sunet.se	sunet-eosc.drive.test.sunet.se	Delete
Mikael Karlsson	mikael.karlsson@csc.fi	fip-86-50-168-150.kaj.poutavm.fi	Delete





The ScienceMesh

- Seamless sharing to a federated recipient after successful invitation
- Support for OCM v1.2
 - Bearer auth, with Basic auth for backwards compatibility
 - **Directory Service** and “Where-are-you-from” page
 - **Planned in 2026**: http-sig and exchange token flow
- RO-Crate sharing via OCM
 - Being developed for EOSC DC to integrate repo workflows





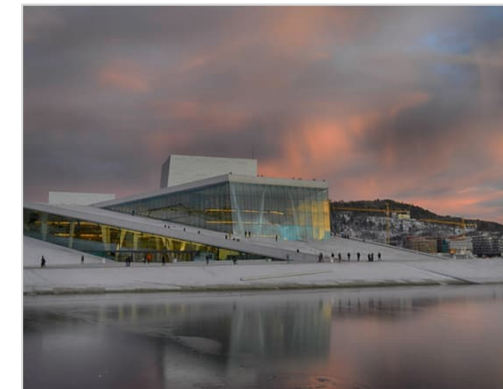
CERNBox outside CERN

- CERNBox has been deployed at Uni Paris-Saclay and Uni Vienna
- A test deployment is in progress at LBNL
- We are in the process of building up a community of interested site administrators, out of the broader Cloud Sync&Share Storage community
 - CERN organizes yearly workshops, next one in Oslo (March 2026): [Registration and call for abstracts is open](#)

université
PARIS-SACLAY



universität
wien



17-19 March 2026

Oslo, NO
Organized by University of Oslo





Want to know more?

- Our communication channels:
 - The CS3 Community: www.cs3community.org
 - The CS3 APIs: <https://github.com/cs3org/cs3apis>
 - Reva: <https://reva.link>, <https://github.com/cs3org/reva>
 - A Discourse site is being setup
 - Matrix chats are available in each repository
- Thank you for your attention, time for questions!

