Whitepaper

High Performance Computing

Easy access for the whole organisation



https://opencloud.eu

OpenCloud

Table of Content

OpenCloud: Easy access to HPC data for the whole organisation 3

Access to cluster data: two workarounds 4

The HPC add-on as an access layer 5

Create access, maintain control 6

Operational benefits 7

An overview of the HPC add-on 8

Sovereignty through open standards 9

OpenCloud: Easy access to HPC data for the whole organisation

Simulations carried out on computing clusters, sequencers, electron microscopes or other measuring instruments generate substantial amounts of data, usually amounting to several petabytes. Universities, research organisations and scientific institutes often store the resulting data on large, parallel file systems, as is typical for HPC (High Performance Computing) environments.

Anyone wishing to access this data normally requires their own user account. Everyone else is denied access: from the administrative office to partner institutions and external colleagues. To make the data accessible, many organisations copy it to a second, browser-based system or set up VPN and shell access. However, both approaches have significant drawbacks.

This is precisely where the HPC add-on for OpenCloud comes in: It acts as an access layer directly on top of the existing parallel file system and uses it as a storage backend. Newly created files appear immediately in OpenCloud and can be accessed via browsers, desktop clients, mobile apps and share links – without a VPN and without a command line. This means everyone works on a single dataset, from researchers at the cluster to external partners.

OpenCloud is based on a fork of the open source software ‘ownCloud Infinite Scale’ (OCIS), whose components were co-developed by developers from the science organisation CERN and other active contributors. OpenCloud is now being further developed by the Heinlein Group with new ideas and a clear focus on data protection, interoperability and sustainable digitalisation.

Access to cluster data: two workarounds

In practice, two approaches have become established. Both only appear to solve the access problem:

Cluster file system · Petabytes
CephFS · GPFS · Leil FS · accessible only via the cluster network

1 Copying data to a second system



Double storage required
The copy takes up the same amount of space as the original, including the backup.



Delays
Whilst the data is being copied, nobody can work productively with it.



Version chaos
Over time, the original and the copy inevitably drift apart.

2 Distribute VPN and shell access



High barrier to entry
Anyone unfamiliar with the command line will be left out.



Security risk
External partner organisations require VPN access to the internal network.



Shadow IT
If the process is too complicated, people often just use USB sticks or private cloud accounts.

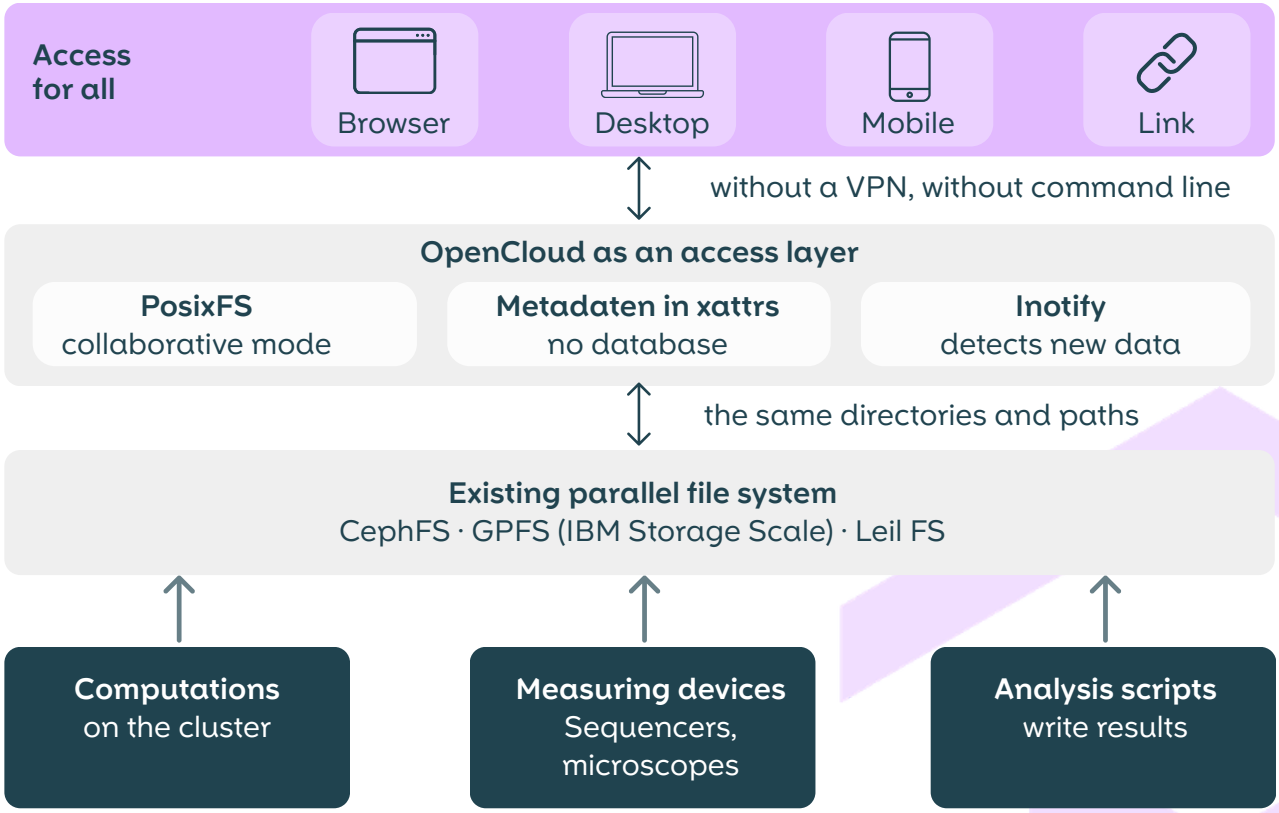


The result: uncontrolled data flows
Both approaches merely shift the actual problem rather than solving it.

The HPC add-on as an access layer

The add-on operates directly on the file system, using the PosixFS storage driver in collaborative mode. This mode allows other processes, as well as OpenCloud, to write to the same directory tree. OpenCloud therefore accesses exactly the same directories and paths as the computations, measuring devices and analysis scripts. This eliminates the need to copy or migrate the data.

The metadata also remains on the file system: OpenCloud stores it in the extended attributes (xattrs) of the respective files. OpenCloud detects newly created and modified files via Inotify, a Linux kernel interface that monitors changes to the file system and notifies programmes without any measurable delay. Results from computations and measurements are therefore immediately available in the browser or other graphical clients, without OpenCloud having to scan the file system.



Create access, maintain control

This allows the data to be visible – though it remains to be clarified who will have access to it. Users should be able to work with the data without the organisation having to open up its network. OpenCloud therefore clearly separates access paths, sharing settings and permissions from one another.

1 Access via familiar channels

Working with the cluster data requires neither shell access nor knowledge of the command line. Users work via the web interface or dedicated desktop clients for Windows, Linux and macOS, as well as apps for iOS and Android. Underpinning this is WebDAV, which handles file access and synchronisation. WOPI is also used for Office integration. Anyone wishing to process the data further can integrate it directly into their own operating system via WebDAV.

Approvals within and outside the organisation 2

Within the organisation, users share data with other registered user accounts. For everyone else, they create links that allow people without their own account to access files. This means it is no longer necessary to set up VPN or SSH access. Link sharing can be disabled if required.

Partner institutions with their own OpenCloud instances or compatible platforms such as Nextcloud or ownCloud benefit from the OCM (Open Cloud Mesh) protocol. This enables them to share files across different locations, even beyond organisational boundaries. Before the first file is shared, there is an invitation process that is restricted to trusted instances. The organisation itself specifies which these are in the configuration.

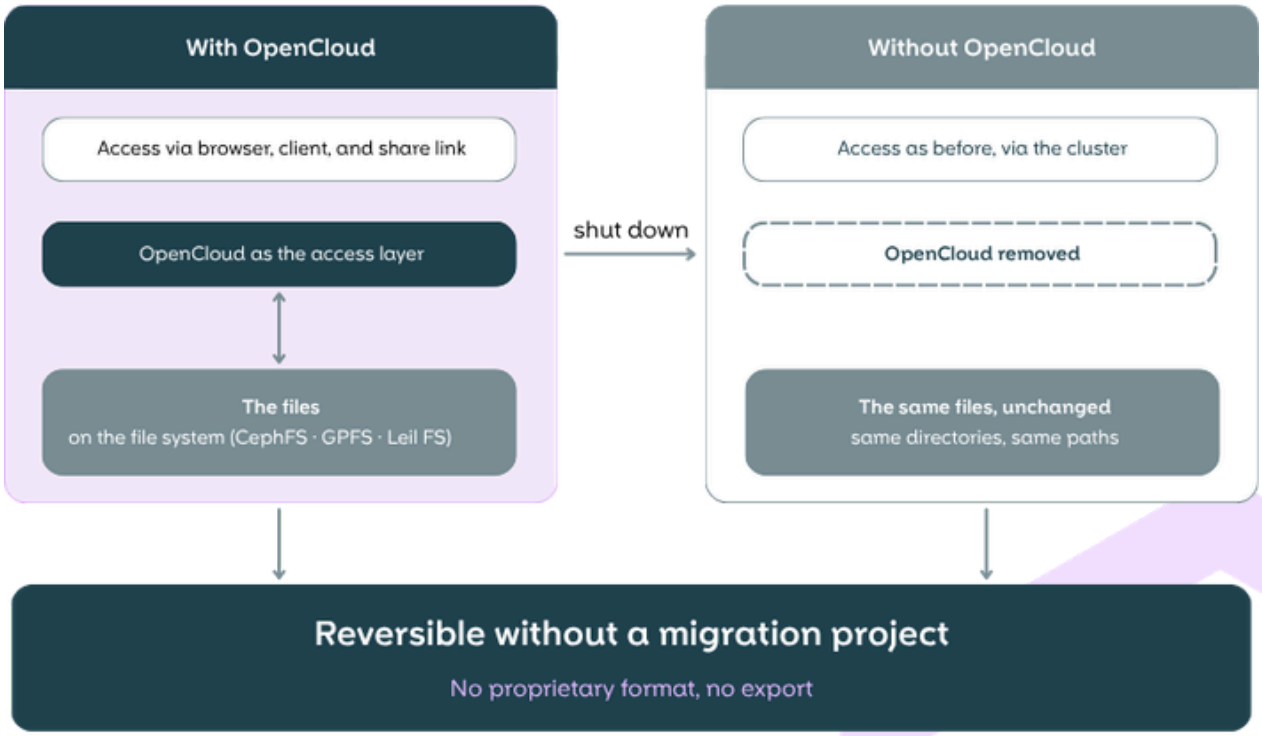
3 Registration and permissions management

Login is via OIDC and OAuth2, using the accounts and groups from the existing identity management system. Alternatively, OpenCloud's built-in user management can be used.
OpenCloud's rights management recognises three different roles. User roles apply system-wide and govern the management of accounts, groups and spaces. Within a space, space roles take precedence. Independently of these, sharing roles determine what recipients are permitted to do with shared files and folders, without their role within the system or within a space changing.

Operational benefits

The OpenCloud architecture simplifies day-to-day operations:

- **A single data set:** No copies, no additional backups, no waiting time due to copying processes. External researchers gain access more quickly.
- **Less shadow IT:** Access without a shell or VPN, and therefore no USB storage devices or private clouds. Data remains within the IT department's controlled environment.
- **Reversible without migration:** After deactivation, the data remains on the file system as normal files, in the same directories and with the same paths. No proprietary format, no export.



Supported file systems

The HPC add-on supports three parallel file systems that are well established in data centres. CephFS is widely used in the academic and research sectors and is deployed as a storage backend in many organisations. In larger HPC environments, GPFS – which IBM has been marketing under the name Storage Scale for several years – is the dominant choice. In addition, there is Leil FS (formerly SaunaFS), a distributed, POSIX-compatible file system licensed under the GPL 3.0. There is currently no integration for Lustre or BeeGFS.

An overview of the HPC add-on

Benefit	Description
Direct file system access	OpenCloud functions as an access layer for the same directories and paths as computations, measuring instruments and analysis scripts.
No second data set	No copying, no additional backups, no waiting time due to copying processes.
Instantly visible	OpenCloud detects newly written files via Inotify and displays them without the need for a separate search.
Access without a shell or VPN	Web interface, desktop clients for Windows, Linux and macOS, apps for iOS and Android, WebDAV.
External sharing	Links without the recipients having their own accounts, across locations via OCM with partner institutions.
Login and permissions	OIDC and OAuth2 using accounts from the existing identity management system, with separate user, space and sharing roles.
Established file systems	CephFS, GPFS (IBM Storage Scale) and Leil FS.
No Lock-in	Once the system is switched off, the data remains as normal files on the file system – no proprietary format, no export.

Sovereignty through open standards

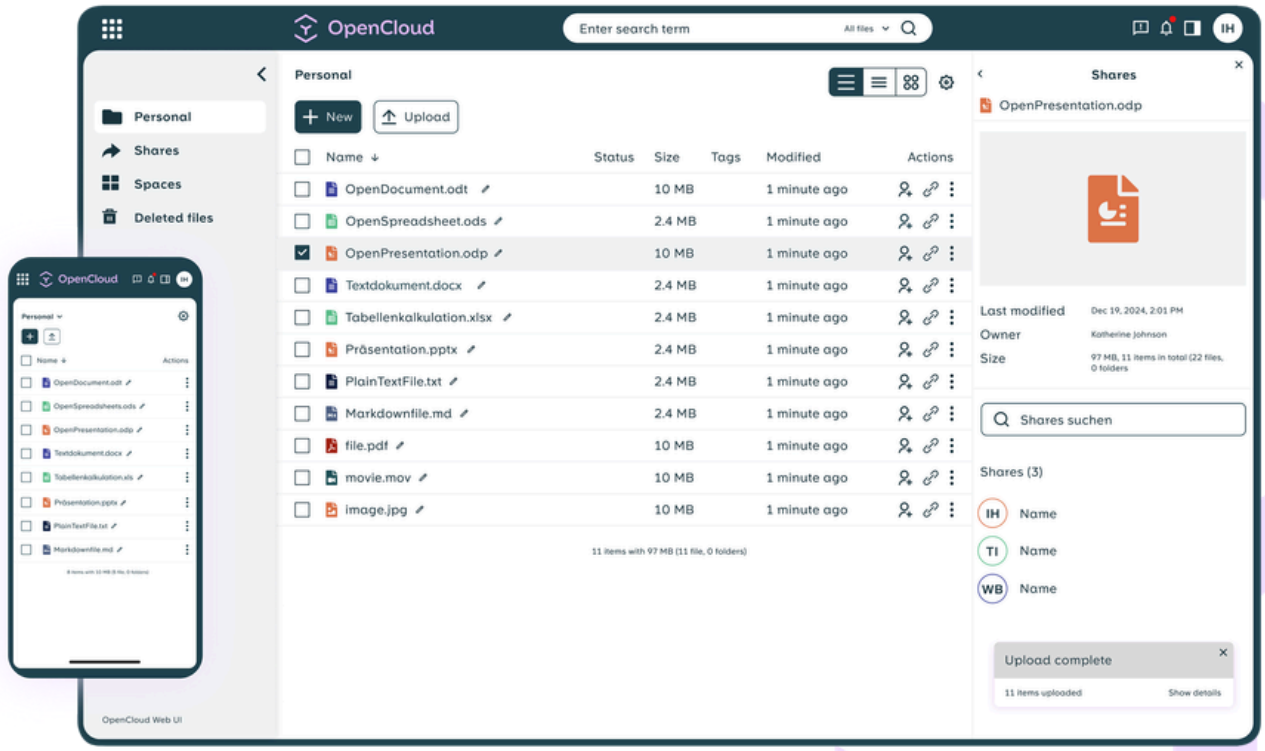
With the HPC add-on, the data stays where it is. OpenCloud does not alter its structure or format, but acts as an access layer on top of it – regardless of which file systems are in use. The organisation decides who is granted access to the cluster data.

Open standards make this possible: POSIX-compatible file systems, WebDAV for file access, OIDC and OAuth2 for authentication, and OCM for sharing across organisational boundaries. OpenCloud is open source, and all data flows are traceable.

It is also possible to revert at any time: if OpenCloud is switched off, only normal files remain on your own file system. A dependency that a migration project would need to resolve never arises in the first place.

Would you like to make your HPC data accessible to the whole organisation? We can support you with architecture, planning and implementation.

Please feel free to contact us at sales@opencloud.eu.
We look forward to hearing from you.



OpenCloud

OpenCloud is the powerful open source solution for file management, file sharing and content collaboration that meets the highest standards of user-friendliness, performance, stability and scalability. The software can be operated flexibly on-premise or securely procured via partners. With intuitive operation, seamless integration, person-independent file management ("Spaces") and secure versioning, OpenCloud is ideal for the public sector, educational institutions, research and large companies and providers.

Together with OpenTalk and mailbox, OpenCloud is one of the three strong pillars of the Heinlein Group, which provides secure and digitally sovereign IT infrastructures for Europe.



https://opencloud.eu