

# CS3 2026 - Cloud Storage Synchronization and Sharing

Tuesday, 17 March 2026 - Thursday, 19 March 2026

University of Oslo



## Book of Abstracts



# Contents

Building the Brazilian Sovereign Data Ecosystem for Research: The RAS Case Study on Federated Architecture and High-Performance Integration . . . . .	1
Fueling the AI Factory: The Central Role of the Metadata Catalog . . . . .	1
Status Update of the no-code platform SeaTable . . . . .	2
How to Run Your Own AI: Real-Life Experiences Operating a Self-Hosted Large Language Model for SeaTable Automations . . . . .	2
HDD Native Parallel Filesystem: Maximizing Throughput for Highest Capacity CMR and HM-SMR Drives . . . . .	3
Migrating DESY's Collaboration Platform from Mattermost to Nextcloud Talk . . . . .	3
Migration from AFS to Cephfs . . . . .	4
Your Infrastructure, Your Rules - Deploy and operate collaboration services the way you want . . . . .	4
Our 11 year fight for data sovereignty . . . . .	4
CBorg Studio: An AI-Ready Data Science Environment for Scientific Programming . . . . .	5
Empowering scientific collaboration: Latest advancements in document editing with ON-LYOFFICE . . . . .	5
Seafile 13, a new way to organize your files . . . . .	6
B2DROP: A Cross-Domain and Cross-Border EFSS for Research Collaboration in the EOSC Federation . . . . .	6
Data staging and caching challenges in the terabit/s era . . . . .	7
Implementing S3-fronted cold storage at CERN . . . . .	7
Segregate Military and Civilian Digital Infrastructure: Building the Future of Connectivity and Storage . . . . .	8
One Codebase, Two Worlds? Collaborating between Browser and Desktop . . . . .	9
From Primary Data to Computational Analysis: Bridging HPC Workflows and Research Documentation . . . . .	9
EFSS and federating storages of the EOSC nodes infrastructure . . . . .	10

Overview of software and infrastructure for research data at ETH Zurich . . . . .	10
Ongoing OCM work funded by SovereignTech and the status of the OCM test suite . . . . .	11
EXA4MIND: A Modular Extreme Data Analytics Platform . . . . .	12
OpenCloud Rising: Wins and Fails of the last 12 Month . . . . .	12
Data protection on a Cloud-Native Infrastructure: experiences over the implementation of AI-driven workflows . . . . .	13
Opencloud instance for PSDI . . . . .	13
Building Kiteworks' Open Source Program Office with ownCloud . . . . .	14
First Steps Towards Building a Pan-European Object Storage . . . . .	14
Sunet Drive Community Report 2026 (Lightning talk) . . . . .	15
Performance Benchmarks of S3 Storage Middleware . . . . .	15
Storage aspects consumed by OpenCloud . . . . .	16
Desktop Client Innovations . . . . .	17
SCIERA: The SCION Education, Research and Academic Infrastructure . . . . .	17
Object Storage at Scale: Real-world Operational Insights from RGW Deployments . . . . .	18
OpenCloudMesh towards IETF Standardization . . . . .	18
Data Commons and Jupyter integration in CERNBox . . . . .	19
CERNBox service evolution towards a federation-enabled EFSS . . . . .	19
Nextcloud. State of the nation . . . . .	20
What's Next in AI for Business? Pushing the analog digital interface boundary. . . . .	20
Expect the (un...-)-expected" - IT business continuity in an increasingly uncertain world . . . . .	20
Feasibility study trans-national replicated object storage using Ceph . . . . .	21
Onedata as Unified Data Namespace for the Polish EOSC Node . . . . .	22
TSD - a part of Norwegian Trusted Invironment Consortium . . . . .	23
Local AI-models at University of Oslo . . . . .	23
Nextcloud deployment at the University of Oslo . . . . .	24
Interoperable Infrastructure for Universities, Enterprises and the Public Sector . . . . .	24
Data Access Broker: machine-actionable workflows to access sensitive data . . . . .	24
Nextcloud Enterprise deployment for Kubernetes, OpenShift and Podman . . . . .	25
Sync&Share Storage since 2013 . . . . .	26

The Czech Path to a Community-Driven Research Data Infrastructure . . . . .	26
Distributed data sharing infrastructure for scientific institute based on Onedata . . . . .	26
Bringing legacy (filesystem) storage systems to the S3 cloudy world . . . . .	27
First experience with OpenCloud in PSNC . . . . .	28
DESY Sync & Share: A dCache-Based Federated Storage Service with Multi-Protocol Access, AI Assistance, and Full-Text Search . . . . .	29
EFSS as Infrastructure: Foundational Layer of Europe’s Sovereign Data Stack . . . . .	29
Digital Sovereignty or Digital Dependency: Europe’s Tech Moment of Truth . . . . .	30
The HPC-Cloud Pipeline: Data Provisioning for Destination Earth and the AI Revolution in Weather Forecasting . . . . .	30
Clinical research and data in the AI era: Lessons and future directions from the AI-Mind project . . . . .	31
Panel . . . . .	31
A Decade of Keeper: 10 Years of Secure Sync, Share & Long-Term Archiving in the Max Planck Ecosystem (2016–2026) . . . . .	31



**Digital Sovereignty / 1****Building the Brazilian Sovereign Data Ecosystem for Research: The RAS Case Study on Federated Architecture and High-Performance Integration**

**Authors:** Lucas Ma<sup>None</sup>; Luiz Eduardo de Souza Coelho<sup>None</sup>; Mateus Rodrigues Oliveira<sup>None</sup>; Rodrigo Moreira de Azevedo<sup>None</sup>

We will present the Rede de Armazenamento Seguro (RAS), the Secure Storage Network of RNP (Brazil's National Research and Education Network), as the core pillar of a sovereign, scalable, and high-performance data ecosystem in Brazil. Our initial strategy leverages an "On-Premise as a Service" (OPaaS) approach through a strategic OPEX partnership, allowing rapid deployment with low capital expenditure (low-CAPEX) in National Data Centers (CNDs) of TIER III standard. This business model is crucial for delivering S3-compatible Storage at a predictable, transparent, and significantly more accessible cost than global hyperscalers, solving the vendor lock-in challenge for the academic community. Furthermore, we will highlight our 2026 expansion roadmap toward a fully open-source approach, including the Ceph platform over Open Compute Project (OCP) hardware, aiming for complete control over the value chain, increased resilience, and the consolidation of national expertise in open infrastructure.

From a technical and scalability perspective, the primary differentiator of RAS is its native integration with the RNP's "Rede Ipê" and e-Science network. This high-performance connectivity (Tb/s backbone) is a non-negotiable requirement for modern science, enabling the efficient, large-scale transfer of petabytes of data generated by global HPC/AI and research projects. The presentation will focus on how the S3 API compliance of RAS facilitates integration with scientific data workflows, and our next steps toward federated interoperability, which directly aligns with the standards and APIs of the CS3 and OCM communities. The system's technical robustness, supported by backends like Ceph, meets the high durability and security requirements needed for critical research data.

By sharing the RAS business strategy, technical roadmap, and governance model, our goal is to offer the CS3 community a replicable blueprint for a collective and federated data infrastructure in the Global South. This case study assertively demonstrates how a National Research and Education Network (NREN) can strategically drive the creation of a local storage services market, forging public-private partnerships to sustain critical infrastructure. We will conclude with an invitation for international collaboration on OCM interoperability testing and the development of connectors for scientific repositories and data lakes, reinforcing our mission to expand the global network of open and trustworthy research infrastructures.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**AI and Storage / 2****Fueling the AI Factory: The Central Role of the Metadata Catalog**

**Author:** Jean-Thomas Acquaviva<sup>1</sup>

<sup>1</sup> *DDN Storage*

The rise of industrial "AI Factories," exposes a critical bottleneck that transcends raw storage performance: managing the data itself.

As AI/ML pipelines ingest and process vast, heterogeneous datasets, the complexity of data discovery, lineage, and governance becomes a primary inhibitor to scaling operations. Traditional storage

systems fail to answer vital questions: “Where is the verified, compliant dataset for training?”, “What is the exact data lineage of this deployed model?”, and “How do we optimize data placement across a distributed infrastructure?”

This presentation details the design and implementation of the intelligent metadata catalog at the heart of the European project DaFAB. We demonstrate that a metadata-driven approach is the key to unlocking efficient, reproducible, and sovereign AI. We will cover: (1) The use of semantic search and active metadata for federated data discovery; (2) Automated data lineage and versioning to ensure model reproducibility and compliance; and (3) How the catalog integrates with the storage layer to orchestrate data movement, respecting data gravity to optimize for performance and cost.

By treating metadata as a primary asset, the DaFAB catalog transforms the storage infrastructure from a passive repository into an active, intelligent component of the modern AI factory. The metadata catalog underlying technology used in DaFab, is the Rucio open-source technology originated from CERN. We will conclude by drawing some perspective with industrial solutions.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Collaborative Applications / 3**

## Status Update of the no-code platform SeaTable

**Author:** Christoph Dyllick-Brenzinger<sup>None</sup>

SeaTable is the world’s leading self-hosted no-code platform. It empowers teams to design, build, and automate business processes in record time—all without writing a single line of code. Users can easily model their data, manage permissions for internal and external collaborators, and visualize information through a variety of chart types. Built-in automations help streamline workflows and accelerate digital transformation across organizations of any size.

With version 6.0, SeaTable introduced AI-powered automations that combine the flexibility of its no-code environment with the rapidly advancing capabilities of artificial intelligence. SeaTable integrates seamlessly with providers such as OpenAI and also supports self-hosted large language models.

In this presentation, I will highlight the major innovations introduced in SeaTable over the past year and demonstrate how they unlock new levels of efficiency and customization for data-driven teams.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**AI and Storage / 4**

## How to Run Your Own AI: Real-Life Experiences Operating a Self-Hosted Large Language Model for SeaTable Automations

**Author:** Christoph Dyllick-Brenzinger<sup>None</sup>

Self-hosting a large language model (LLM) to power AI-driven automations is no longer just a theoretical possibility—it’s a practical reality. In this talk, discover how SeaTable 6.0 integrates AI-powered automations by running a self-hosted LLM on a GPU server at Hetzner using the vLLM framework. Learn about the technical challenges, infrastructure requirements, and operational insights gained from deploying and managing your own AI model in a production SaaS environment.

This session offers a unique, hands-on perspective for organizations looking to leverage AI while maintaining full control over their data and customization. Whether you're exploring AI adoption or seeking alternatives to cloud-only providers, this presentation shares actionable lessons and best practices from real life.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Scalable Storage Backends and Integration with Data Processing / 6**

## **HDD Native Parallel Filesystem: Maximizing Throughput for Highest Capacity CMR and HM-SMR Drives**

**Author:** David Gerstein<sup>1</sup>

**Co-author:** Piotr Modrzyk<sup>1</sup>

<sup>1</sup> *Leil Storage*

As HDD capacities reach tens of terabytes while per drive IOPS remain largely unchanged, conventional SDS designs often hit IOPS limits before achieving full sequential bandwidth. This imbalance frequently leads operators to choose smaller disks in an attempt to balance capacity and performance. Leil Storage addresses this challenge with a sequential access model that aligns I/O patterns with the natural behavior of HDDs, making the approach truly HDD native. This enables sustained maximum throughput on high capacity CMR and HM-SMR drives alike. Using SaunaFS as a reference, we describe the architectural changes required to meet HM-SMR constraints, including zone appends, zone cleaning, and the elimination of in place rewrites, while preserving both performance and reliability.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Lightning talks / 7**

## **Migrating DESY's Collaboration Platform from Mattermost to Nextcloud Talk**

**Author:** Ingo Ebel<sup>1</sup>

<sup>1</sup> *Deutsches Elektronen-Synchrotron (DE)*

DESY is transitioning its long-running Mattermost service to Nextcloud Talk to create a more integrated, secure, and user-friendly communication environment. This migration strengthens alignment with DESY's digital sovereignty strategy and unifies messaging, file sharing, and collaborative editing within a single platform.

The presentation highlights key technical and organizational aspects of the move: architecture and scaling, federation, compliance and integration with existing DESY infrastructure. We share lessons learned from pilots, operational evaluations, and user feedback, offering practical insights for research institutions modernizing their real-time communication services.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

## Scalable Storage Backends and Integration with Data Processing / 8

### Migration from AFS to Cephfs

**Author:** Ingo Ebel<sup>1</sup>

<sup>1</sup> *Deutsches Elektronen-Synchrotron (DE)*

DESY completed a migration of user home directories from AFS to CephFS to address performance limitations, aging infrastructure, and operational complexity. The project involved deploying a high-availability Ceph cluster with NVME-backed metadata pools, erasure-coded data pools and quota management - demonstrating CephFS as a scalable, production-grade replacement for legacy AFS deployments in scientific research environments.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

## Access to data / 9

### Your Infrastructure, Your Rules - Deploy and operate collaboration services the way you want

**Author:** Björn Schießle<sup>1</sup>

<sup>1</sup> *Nextcloud GmbH*

Operating an on-premise solution differs fundamentally from providing a typical Software-as-a-Service product —especially when it comes to deployment and operation. While SaaS providers control their runtime environment end to end, an on-premise solution must empower the organization itself to operate the service efficiently within its own technical and organizational context. This requires flexibility to adapt to diverse infrastructures, existing expertise, and established workflows.

Nextcloud, as the leading open source on-premise collaboration and file exchange platform, has extensive experience in achieving this balance. The talk will present practical examples from real-world deployments illustrating how Nextcloud integrates seamlessly with existing systems without disrupting established processes or requiring fundamental architectural or organisational changes.

Beyond deployment, flexibility also extends to user experience and customization: different user groups often have distinct workflows and needs. The second part of the talk will discuss how Nextcloud's modular design and open architecture enable organizations to tailor the platform precisely to their requirements while maintaining a consistent user experience and operational reliability.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**CS3 Community Site Reports / 10****Our 11 year fight for data sovereignty****Author:** Lilian Emming<sup>None</sup>**Co-author:** Ron Trompert

In 2014 we started SURFdrive since universities and higher education in the Netherlands did not want their data to end up at commercial cloud services. 11 years have gone by where the number of users increased, the amount of stored data increased and also the number of sync-and-share installations which went up from 1 to more than 40 serving about 80 institutes in total. Also sync-and-share we offer in two flavours, one for personal data but also one especially tailored for research purposes. The last few years however, we increasingly felt the competition from Microsoft in the form of a reducing number of users. Recently, the geopolitical situation has changed the landscape drastically. Next year we are going to start a large pilot project with more than 30 organisations where we want to show that there are viable alternatives to Microsoft and that they can make a different choice.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**AI and Storage / 11****CBorg Studio: An AI-Ready Data Science Environment for Scientific Programming****Authors:** Andrew Schmeder<sup>1</sup>; Timothy Fong<sup>2</sup><sup>1</sup> *Lawrence Berkeley National Laboratory*<sup>2</sup> *Berkeley Lab*

This presentation introduces CBorg Studio, an integrated AI-ready data science environment developed by the Science IT Consulting Group at Lawrence Berkeley National Laboratory.

CBorg Studio is designed to help scientists leverage the newest AI-powered software engineering tools and research data management systems via an easy-to-use JupyterHub-based computing platform. At launch, the CBorg Studio container is pre-populated with an ephemeral API key connecting the user to the CBorg AI inference server, enabling the user to immediately use pre-configured AI developer tools with on-prem and cloud-based models including coding agents and editor extensions. Additionally, CBorg Studio operates as an experimental platform for development, testing and training users on new tools for AI-first scientific programming, such as prompt logging & introspection, agentic memory, data version control, and agentic workflow orchestration.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Collaborative Applications / 12****Empowering scientific collaboration: Latest advancements in document editing with ONLYOFFICE**

**Authors:** Galina Goduhina<sup>None</sup>; Richard Ruf<sup>None</sup>

In the fast-paced world of scientific research, efficient document processing and seamless collaboration are critical. This session explores the major evolutions within the ONLYOFFICE suite over the past year, designed specifically to support the complex needs of research and academic environments.

We will demonstrate how deep AI integration now accelerates drafting and analysis, allowing researchers to automate macro creation and process data more intelligently. Attendees will discover the redesigned interface that modernizes the user experience across all editors, alongside new specialized tools like the Diagram Viewer and enhanced support for scientific formats including Markdown.

The talk will also cover crucial updates for handling sensitive research data, featuring advanced PDF capabilities such as redaction and stamping. Furthermore, we will highlight improvements in accessibility and customization, from customizable keyboard shortcuts to robust RTL support. Join us to see how these innovations create a more flexible, secure, and efficient workspace for the scientific community.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**File Sync & Share Solutions and Requirement from the Community / 13**

## **Seafile 13, a new way to organize your files**

**Author:** Jonathan Xu<sup>None</sup>

We're thrilled to introduce Seafile 13, the latest version of our open-source file management and collaboration platform.

Since 2012, we've been dedicated to providing teams with a secure, reliable, and fast file sync & share solution. While people work with files and documents every day, the way they do so have remained largely unchanged for over a decade. With advancements in technology and AI, we want to offer people a fresh approach to organizing their files.

Seafile 13 introduces several innovative features designed to help you organize your files more efficiently.

- Extensible File Properties
- Flexible File Views
- Multi-level File Tagging
- Organize your files with AI

In addition to smart file organization, Seafile also features handy add-ons, including an AI-powered, Notion-like doc editor, an integrated whiteboard, and knowledge bases. These tools help you create documents in a smarter way.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**CS3 Community Site Reports / 14**

## B2DROP: A Cross-Domain and Cross-Border EFSS for Research Collaboration in the EOSC Federation

**Authors:** Marvin Winkens<sup>1</sup>; Sander Apweiler<sup>2</sup>

<sup>1</sup> JSC

<sup>2</sup> Forschungszentrum Jülich

B2DROP is EUDAT's Enterprise File Synchronization and Sharing (EFSS) service, designed to provide researchers across Europe with a collaborative platform for seamless data exchange, sharing, and real-time collaboration—regardless of research domain or national borders.

As part of EUDAT's B2 service suite, B2DROP streamlines data publication and long-term archiving after project completion while integrating with domain-specific services, such as the CLARIN Switchboard for the humanities and social sciences. The premium version of B2DROP offers dedicated storage solutions for research groups or projects, with customizable extensions to meet community-specific requirements.

As a core component of the European Open Science Cloud (EOSC), B2DROP supports cross-node use cases, enabling interoperability and collaboration within the EOSC federation. EUDAT, as part of the first wave of EOSC nodes, ensures that B2DROP aligns with EOSC's goals for open and FAIR (Findable, Accessible, Interoperable, Reusable) data management.

To simplify access control and group management, B2DROP leverages group membership information from upstream Authentication and Authorization Infrastructures (AAIs). Researchers can manage groups via B2ACCESS (EUDAT's AAI) or any other EOSC-compatible AAI, ensuring flexible and secure collaboration across the European research ecosystem.

### **Suggested Contribution Type:**

Regular Talk (15-30 min)

### **Access to data / 15**

## Data staging and caching challenges in the terabit/s era

**Authors:** Maiken Pedersen<sup>None</sup>; Mattias Wadenstein<sup>1</sup>

<sup>1</sup> University of Umeå (SE)

The Nordic Tier-1 site for LHC data and computing is a distributed one, where computing and storage is spread over several countries. In order to do efficient computing on the data, a local low-latency cache is needed close to each computing cluster.

This talk will give an overview on how this works in production today, with data throughputs in the 100Gbit/s range, and then talk about the challenges ahead where scaling of a factor of 10-20 on top of that will be needed to keep up with the increased data rates from the High-Luminosity LHC upgrade that comes online in 2030.

The challenges of staging data at 1 Tbit/s will be discussed from networking, hardware, software, and operational points of view.

### **Suggested Contribution Type:**

Regular Talk (15-30 min)

**Scalable Storage Backends and Integration with Data Processing / 16****Implementing S3-fronted cold storage at CERN**

**Author:** Mario Vitale<sup>1</sup>

<sup>1</sup> *CERN*

The CERN Tape Archive (CTA) is the open source solution developed at CERN to store more than 1 Exabyte of data from CERN's experimental programmes. CTA interfaces with two disk systems widely used by the High-Energy Physics (HEP) community, EOS and dCache. However, until now there has been no integration with systems used outside of HEP.

Looking at current industry standards, the leading interface for object storage is S3, which includes cold storage extensions for data archival. The CTA team is investigating whether CTA can be fronted by an S3 API. During this talk, we'll review a proof-of-concept implementation, and look at alternative solutions to explore along with their respective trade-offs.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Digital Sovereignty / 17****Segregate Military and Civilian Digital Infrastructure: Building the Future of Connectivity and Storage**

**Author:** Mauro Vignati<sup>1</sup>

<sup>1</sup> *ICRC*

Digital infrastructure and services—including terrestrial and satellite networks, cloud computing, undersea cables, and cybersecurity—are often shared by civilian and military users. Civilians in situations of armed conflict rely on stable and secure internet access and cloud storage to locate aid stations and hospitals, receive real-time safety warnings, maintain or restore family links, and access sensitive information. Militaries engaged in hostilities may use the same infrastructure or services for offensive and defensive operations, leveraging them for reconnaissance, exploitation, command and control (C2), cybersecurity, or data analytics. This overlap creates risks during armed conflict: civilians risk losing access to essential services, and even physical harm, as military components of digital infrastructure are targeted. Growing digital public-private partnerships and increasing integration of civilian and military digital infrastructure and services only increase shared use.

Rather than accept that civilians will suffer from collateral damage, the ICRC is working to avoid, or at least minimize, their exposure to the dangers of warfare. The goal is to improve civilian protection in armed conflict by exploring ways to separate civilian and military digital infrastructure and services or otherwise protect civilian and humanitarian digital uses. We consider physical, technical, standards, policy, and strategic approaches. We do so by bringing together technologists; industry, policy, and legal experts; humanitarian actors; and civil society. The project will identify achievable and forward-looking solutions that anticipate future technological and geopolitical challenges.

Attendees will discover the problems associated with the shared use of the same digital infrastructure by civilians and the military, especially cloud and connectivity, and how in the future these infrastructures will have to be separated and segregated to reduce the impact on civilians.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

### Collaborative Applications / 18

## One Codebase, Two Worlds? Collaborating between Browser and Desktop

**Author:** Michael Meeks<sup>None</sup>

As we bring our Online product to the native desktop, this allows all sorts of interesting new interactions for on-line & off-line and negotiating transitions between co-editing, sharing, on-line and off-line.

Come and hear how we brought the refreshed Collabora Online experience to the desktop, and some of the major improvements in ergonomics and usability we've investigated—as well as some of the statistical data driving this work.

Learn about our developing protocol extensions for integrating EFSS into our desktop app, as well as a general update on new functionality and interoperability work.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

### Data Science Environments & HPC integration / 20

## From Primary Data to Computational Analysis: Bridging HPC Workflows and Research Documentation

**Authors:** Mohamad Hayek<sup>1</sup>; Ramon D'Agosta<sup>2</sup>; Rory Macneil<sup>2</sup>; Tilo Mathes<sup>None</sup>

<sup>1</sup> LRZ

<sup>2</sup> ResearchSpace

Research institutions face a critical challenge in managing large-scale data, such as sequencing data, across their complete lifecycle—from generation of the primary data with an instrument through HPC analysis to long-term archival storage—while maintaining data provenance and FAIR compliance. Researchers currently lack integrated tools to coordinate data movement across storage tiers, forcing manual action that creates compliance risks, inflate storage costs due to duplication of data, and break connections between primary data, experimental context, and computational results. As computational methods become increasingly intrinsic to research, the integration between active research documentation and compute infrastructure becomes critical for efficient workflows to create FAIR data.

Building on our approach to vertical interoperability presented at CS3 2025, where we demonstrated how RSpace, an open-source research data management platform including an electronic laboratory notebook, can provide a user-friendly frontend for institutional file sync and share solutions like iRODS, we are now extending this approach to manage data from its primary origin to being used in HPC workflows and the outputs created there. RSpace's S3 integration enables researchers to seamlessly manage sequencing data across distributed storage locations through an interface they already use daily for experiment documentation and sample management. The solution provides intuitive file operations between S3 buckets and other RSpace-supported storage systems, links rich contextual metadata from experimental documentation and sample records to files in S3 storage, and robustly connects HPC workflow outputs back to originating experiments with relevant run metadata.

A typical workflow illustrates the value: researchers document experiments in RSpace with linked samples and protocols, connect raw sequencing data (regardless of storage location) to their documentation with contextual metadata, transfer data with metadata collected in RSpace to an HPC-environment via RSpace's unified interface, and after HPC analysis, link result files back to original experiments, preserving complete lineage.

This approach, developed in collaboration with the Leibniz Supercomputing Center and the University of Göttingen, addresses FAIR adoption barriers by providing seamless access to data regardless of storage backend, ensuring experimental context travels with data for discoverability and lineage tracking, and enabling cost optimization through efficient data lifecycle management across storage tiers. By integrating institutional storage infrastructure directly into researchers' daily workflows, we reduce usability barriers while improving FAIR compliance—making data management practices easier to adopt and sustain.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Federations for EOSC and eResearch infrastructures. / 21**

## **EFSS and federating storages of the EOSC nodes infrastructure**

**Author:** Ron Trompert<sup>None</sup>

In the spring of 2025 14 EOSC candidate nodes started their journey to add institutional, national and thematic service platforms to EOSC. These nodes are federated through EOSC AAI where users have a single login to multiple services. An optional component of EOSC are Enterprise File Sync and Share services. Independent of a federated EOSC AAI, they provide their own mechanism for federation based on trust relations between groups or individuals allowing for a fine-grained access to data. Currently the so-called federation build-up group is working on establishing the federation and "File sync-and-share" is a subgroup of the build-up group.

In this talk the work of the EFSS subgroup is discussed as well as their vision and future plans.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Data Science Environments & HPC integration / 22**

## **Overview of software and infrastructure for research data at ETH Zurich**

**Authors:** Caterina Barillari<sup>None</sup>; Jarunan Panyasantisuk<sup>1</sup>

<sup>1</sup> *ETH Zurich*

In this talk we will present the options available to researchers at ETH Zurich for managing, storing, analyzing and sharing research data.

Since 2022, guidelines for research data management are in place at ETH Zurich. These state that all processes around research data must be documented according to the FAIR data principles and recommend the use of Electronic Laboratory Notebooks (ELN).

At ETH Zurich we develop a combined data management, inventory management system and digital notebook for experimental scientists across several disciplines (openBIS). This is provided as a Research Data Management (RDM) service inside ETH Zurich, to all interested groups. The software runs on ETH infrastructure and relies on Network Attached Storage (NAS), Cost Defined Storage

(CDS) and Long Term Storage (LTS) for data storage. Moreover, data can be easily directly transferred to the Zenodo and ETH Research Collections data repositories for data publication and sharing.

In addition to this resource for data management, we provide solutions for collaborative work within ETH Zurich and with external partners. These include Microsoft Teams, Sharepoint and OneDrive, Google Workspace and polybox (based on Nextcloud). For data processing and analysis, we maintain 2 cluster infrastructures: Euler (for research data with no specific restrictions) and Leonhard Med (for strictly confidential research data). In addition, we also provide access to external cloud resources such as Azure, GCP, AWS and CSCS. A cloud assessment process is in place.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Interoperability: protocols, APIs, OpenCloudMesh (OCM) / 23**

## **Ongoing OCM work funded by SovereignTech and the status of the OCM test suite**

**Author:** Mahdi Baghbani<sup>None</sup>

### **Background**

The Open Cloud Mesh (OCM) initiative aims to enable secure, interoperable data and application sharing across independently operated cloud platforms. Backed by funding from the Sovereign Tech Agency (STA), ongoing work within the CS3 ecosystem is advancing both practical interoperability and the maturation of OCM as an Internet standard.

### **Scope of the STA-funded work**

This talk presents an overview of the current STA-funded work program, structured around concrete milestones spanning implementation, validation, and standardization. Key efforts include deeper integration of directory services and Where Are You From (WAYF) user interfaces in CERNBox, aligned with developments in Nextcloud, to support realistic cross-instance sharing scenarios. In parallel, the OCM Test Suite is being extended and integrated into the CI pipelines of Nextcloud, Reva, and other services, enabling continuous, automated conformance testing across vendors and OCM protocol versions.

### **Protocol validation and reference implementation**

A central aspect of the work is strengthening protocol validation. This includes incorporating CERN-Box share flows into the test suite, implementing and validating OAuth 2.0-based code flows between Nextcloud and CERNBox, and enhancing the test infrastructure with structured, machine-readable logging and CI artifact reporting. The test suite is also being upgraded to cover newer OCM protocol features such as invite-first flows, .well-known/ocm-based discovery, and bearer-token-based access. To support implementers, the OCM stub is being rewritten as a compiled, production-grade reference implementation, with capability discovery, automated TLS via ACME, improved configurability, and containerized deployment.

### **Takeaways for the CS3 community**

The presentation provides a status update on these milestones, discusses lessons learned from cross-vendor testing, and outlines how the CS3 community can use the OCM Test Suite and reference components to improve interoperability and move the protocol toward standard maturity.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Data Science Environments & HPC integration / 24****EXA4MIND: A Modular Extreme Data Analytics Platform**

**Authors:** Ismail Hakki Toroslu<sup>1</sup>; Jan Martinović<sup>2</sup>; Martin Golasowski<sup>2</sup>; Mohamad Hayek<sup>3</sup>; Pinar Karagoz<sup>1</sup>; Stephan Hachinger<sup>4</sup>; Viktoria Pauw<sup>3</sup>

<sup>1</sup> *Middle East Technical University*

<sup>2</sup> *IT4Innovations, VSB - Technical University of Ostrava*

<sup>3</sup> *Leibniz Supercomputing Centre*

<sup>4</sup> *Leibniz Supercomputing Centre (LRZ) of the BAdW*

In this talk, we present the project EXA4MIND (Extreme Analytics for Mining Data spaces), which establishes a modular, flexibly-deployable Extreme Data analytics platform. The platform toolkit enables the user to combine different data storage, database systems and powerful computing infrastructures to enable advanced data analytics, machine learning and AI on heterogeneous, large datasets coming from science, industry, and SMEs. Our talk will focus on the design and implementation of distributed data management within the project, where functionalities from data storage and management, over data staging and caching, to FAIR-compliant publication services are provided. For data processing, EXA4MIND provides an “Advanced Query and Indexing System” (AQIS), where analytics workflow orchestration is facilitated leveraging Apache Airflow and DASK. The platform thus helps to automate data ingestion, transfer, caching, querying and computing integration across cloud and supercomputers (HPC clusters), enabling seamless data processing across diverse backends. The system supports persistent identifier assignment and metadata management, in particular for FAIR compliance of research data, and aims at an integration with European data ecosystems (e.g. European Data Spaces, EUDAT, EOSC). Co-designing our systems with four application cases from the molecular-dynamics, automotive, smart viticulture and health sectors, we contribute to the technical groundwork for enabling next-generation data analytics in Europe.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**File Sync & Share Solutions and Requirement from the Community / 25****OpenCloud Rising: Wins and Fails of the last 12 Month**

**Author:** Tobias Baader<sup>1</sup>

<sup>1</sup> *OpenCloud*

Over the past 12 months, OpenCloud has advanced its architecture, scalability and integration capabilities while uncovering critical operational lessons. This talk reviews key wins including multi-tenancy, Kubernetes deployment, OpenSearch integration alongside the failures and challenges that shaped our approach to performance, reliability and data sovereignty. Attendees will gain a high-level technical perspective on what worked, what didn't and how these insights guide our roadmap.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Applications handling sensitive data, data classification, data security and privacy / 26****Data protection on a Cloud-Native Infrastructure: experiences over the implementation of AI-driven workflows**

**Authors:** Alessandro Costantini<sup>1</sup>; Andreas Turk<sup>2</sup>; Barbara Martelli<sup>3</sup>; Francesco Sinisi<sup>3</sup>; Giusy Sergi<sup>None</sup>; Jacopo Gasparetto<sup>4</sup>; Jiri Horak<sup>2</sup>; Lukasz Hejtmank<sup>None</sup>; Petr Holub<sup>None</sup>

<sup>1</sup> *INFN-CNAF*

<sup>2</sup> *bbmri-eric*

<sup>3</sup> *INFN CNAF*

<sup>4</sup> *CNAF*

The convergence of artificial intelligence (AI) and biomedical research is reshaping the way complex datasets are analyzed, particularly in histopathology, where high-resolution imaging plays a critical role in diagnosis and treatment planning. As research institutions increasingly adopt AI-driven workflows, the challenge lies not only in achieving computational scalability and performance but also in ensuring compliance with stringent security and privacy requirements.

This tension becomes especially pronounced when operating within an Information Security Management System (ISMS), where every technological choice must align with established governance frameworks.

In this context, the collaboration between Italian EOSC and BBMRI-ERIC nodes, supported by INFN and Masaryk University's CERIT-SC team, represents a significant step toward building secure, cloud-native infrastructures for medical data processing. The initiative leverages advanced technologies - such as Kubernetes clusters hardened with CIS benchmarks, federated identity management via Keycloak, and containerized AI pipelines - to enable multi-institutional research while safeguarding sensitive health data. These components collectively form a platform capable of hosting deep learning workflows for histopathological image analysis, ensuring reproducibility, traceability, and compliance with security standards.

However, the integration of these technologies within an ISMS environment has revealed operational complexities. Balancing elasticity and modularity with strict security policies requires continuous negotiation between innovation and governance.

This contribution explores the practical experience of achieving a partial equilibrium between these competing demands, highlighting the architectural decisions, security controls, and collaborative strategies that enabled the deployment of an AI-ready infrastructure without compromising data integrity or regulatory compliance. By sharing these insights, we aim to contribute to the broader discourse on secure AI adoption in health and life sciences, demonstrating that technological advancement and policy adherence can coexist, albeit through deliberate design and iterative refinement.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**CS3 Community Site Reports / 27****Opencloud instance for PSDI**

**Authors:** Alexander Belozarov<sup>1</sup>; Amali Pawula Hewage<sup>1</sup>; Stuart Meacham<sup>1</sup>; Tom Underwood<sup>1</sup>; Vasily Bunakov<sup>1</sup>

<sup>1</sup> *STFC UKRI*

PSDI (Physical Sciences Data Infrastructure) develops guidance, training and technology to address the needs of the UK “bench science” community who currently do not have a developed common digital infrastructure, in contrast to “big science” such as particle physics or astronomy who have quality international infrastructures of their own and may not need another one on a national scale.

Data sharing solutions and their integration with other services have been of interest to PSDI for some time, with OCIS and Nextcloud tried earlier and with a focus now moving towards the adoption of Opencloud. We are reporting on the experience of Opencloud setup in OpenStack environment, and on the use cases for its exploitation by PSDI partners.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**File Sync & Share Solutions and Requirement from the Community / 28**

## **Building Kiteworks’ Open Source Program Office with ownCloud**

**Author:** David Walter<sup>None</sup>

Open source success is no longer just about publishing code —it’s about stewardship, governance, and trust. This talk introduces Kiteworks’ Open Source Program Office (OSPO) with ownCloud as its open-source centerpiece, sharing real-world lessons on deciding what must be open, aligning community and enterprise expectations, and building credible open source governance inside a security-focused company. Practical insights, no marketing, and a few hard-earned lessons included.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Digital Sovereignty / 29**

## **First Steps Towards Building a Pan-European Object Storage**

**Authors:** Bo Nygaard Bai<sup>None</sup>; Kalle Happonen<sup>1</sup>

<sup>1</sup> CSC - IT Center for Science Ltd.

Object storage –specifically the S3 protocol –has become a de-facto standard for storage integration. More and more of us are running S3 services, but they are still disjointed, non-interoperable, and scattered from both admin and user perspective.

At the same time we are faced with more and more challenges when working with research data, from data sizes to legal and compliance requirements. Growing trends toward data sovereignty and in-house or community services also push the scaling and quality requirements of the object storage services to their limits.

Within the GÉANT GN5-2 project, we are developing a blueprint for a trusted pan-European storage infrastructure layer that provides a common, resilient foundation for research services. The goal is not to address higher-level data management directly, but to enable institutional and domain-specific services to build on an trusted object storage layer with consistent capabilities for metadata, policy enforcement, compliance, and sovereignty across borders.

A small group of sites who have been providing Ceph based S3 services for a long time has already started to work towards a more common approach of providing object storage services. We're in early days of the collaboration, but initial focus is on a operational convergence and uniformity. This includes the development of architectural blueprints for hardware and deployment that makes it possible to develop shared tooling. Achieving more uniform object storage services is the first step in the longer term goal of a Pan-European object storage service for the research community in Europe.

The Pan-European service will significantly raise the abstraction level of research storage. This stable foundational storage layer will be a building block for higher level data services and research collaborations. EFSS will likely be a pivotal services that institutions will want to deploy on top of this.

In this talk, we will go through the first steps towards this vision, and present a rough roadmap how we can achieve this goal. We will also go through what benefits adopting the blueprint would bring to sites, and how we see it developing in the coming years.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**CS3 Community Site Reports / 30**

## **Sunet Drive Community Report 2026 (Lightning talk)**

**Author:** Lars Delhage<sup>None</sup>

**Co-authors:** Micke Nordin<sup>1</sup>; Richard Freitag ; Rikard Danielsson<sup>1</sup>

<sup>1</sup> *SUNET*

Sunet Drive is Sweden's national data storage solution and an active member of the Open Cloud Mesh standardization process. It is a federated solution, consisting of 54 nodes, one for every Swedish institution, including one node for external users. We will give an up-to-date overview of of Sunet Drive, including

- User and storage development
- New customer on-boarding and customizations
- Updates and incidents
- OCM integration for EOSC
- Multi-factor and step-up authentication

Special focus of the community report will lie on the development of Sunet Drive into a mature and sovereign solution, while remaining a "purely federated" deployment.

**Suggested Contribution Type:**

Lightning Talk (5 min)

**Scalable Storage Backends and Integration with Data Processing / 31**

## **Performance Benchmarks of S3 Storage Middleware**

**Author:** Huseyn Gurbanov<sup>1</sup>

**Co-authors:** Martin Golasowski<sup>2</sup>; Mohamad Hayek<sup>1</sup>; Pinar Karagöz<sup>3</sup>; İsmail Hakkı Toroslu; Stephan Hachinger<sup>1</sup>; Jan Martinović<sup>2</sup>

<sup>1</sup> *Leibniz Supercomputing Centre (LRZ) of the BAdW*

<sup>2</sup> *IT4Innovations National Supercomputing Center, VSB - Technical University of Ostrava*

<sup>3</sup> *Middle East Technical University (METU), Ankara, Turkey*

Amazon's Simple Storage Service (S3) protocol has become a de-facto standard for object storage, and academic supercomputing centres are increasingly interested in offering S3-compatible storage solutions. This is on the one hand due to more (on-premises) cloud computing at such centres, and on the other hand due to the increasing demands of flexible data access and sharing across computing sites. Certainly, the movement towards FAIR (Findable, Accessible, Interoperable, Reusable) Research Data Management plays a significant role here.

The EXA4MIND Horizon Europe project aims at offering optimised storage back-ends, including object storage and databases, for Extreme-Data applications at European supercomputing centres. As part of this effort, we have conducted performance tests of different open-source middleware alternatives to implement S3 storage. We have performed these benchmarking experiments on bare-metal, virtual-machine and High-Performance Computing (HPC) systems.

We show first results using widely-used open-source S3 middleware such as MinIO, S3Proxy, and VarsityGW. It appears that a combination of such middleware and optimised client software, making use of multiple parallel streams, is able to saturate network bandwidths up to several GB/s, possibly enabling direct object-store usage from HPC codes. Typical performance-limiting factors include I/O based on small files in the 100 MB range or below, and a too strongly restricted number of streams.

**Acknowledgement:** This work received support from the EXA4MIND project ("EXtreme Analytics for MINing Data spaces"), funded by the European Union's Horizon Europe Research and Innovation Programme, under Grant Agreement N° 101092944. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Scalable Storage Backends and Integration with Data Processing / 32**

## **Storage aspects consumed by OpenCloud**

**Author:** Jörn Dreyer<sup>1</sup>

<sup>1</sup> *OpenCloud GmbH*

OpenCloud has the design goal to not use a relational database. This requires a deeper integration with the underlying storage system, ie. through extensive use of extended file attributes. Since features like file revisions, trash and shares are inevitable nowadays, OpenCloud makes use of SDS native supported storage aspects to build these advanced features in an efficient way.

In this talk we will give an overview of the storage aspects that are relevant from OpenClouds perspective, the integrations that we currently support as well as ongoing research topics.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

Access to data / 33

## Desktop Client Innovations

Author: Klaas Freitag<sup>1</sup>

<sup>1</sup> *OpenCloud GmbH*

OpenCloud is an innovator in the area of data synchronization with the desktop client to keep data in sync between workstations and the server.

This talk will highlight a few new architectural aspects of OpenCloud and its desktop client that are used for reliable and performant synchronisation of data. Also it will talk about a few central APIs that are used between client and server.

With some news around the so called Virtual File System we have improved the user experience a lot. Together with other decisions we aim to improve the user and deployment experience in large managed environments.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

Access to data / 34

## SCIERA: The SCION Education, Research and Academic Infrastructure

Author: Francois Wirz<sup>1</sup>

<sup>1</sup> *ETHZ*

SCIERA is the SCION Education, Research and Academic Infrastructure and leverages the layer 2 connectivity fabric of RRENs, NRENs and other academic networks to enable a global deployment of the SCION path-aware network architecture, spanning over 5 regions, covering institutions in the EU, Asia, North and South America, as well as Africa.

After a short review of the benefits and trade-offs of such an architecture, we describe how we overcame the initial deployment hurdles for SCIERA, by leveraging existing connectivity within the participating networks, while staying entirely overlay-free between networks, and reducing costs.

Following an overview of the state of the deployment of SCION in other networks, spanning from financial settlement networks<sup>[4]</sup> to critical infrastructure providers, and their reliance on IP-to-SCION-to-IP gateways, we show how we increased the number of *native* SCION use cases in SCIERA, where applications are fully SCION-aware and optimize communication across all path choices offered by the network.

We then present the lessons learned while deploying SCIERA for bootstrapping user adoption of native SCION, enabling application developers, hardware deployment for the border routers, maintenance and observability and community support.

To conclude the talk, we give an overview of the ecosystem of cloud providers and NSPs offering SCION and present some use cases in the file sharing space and beyond enabled by the SCIERA

infrastructure as well as the open source application development libraries available for SCION [7][8][9].

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Scalable Storage Backends and Integration with Data Processing / 36**

## **Object Storage at Scale: Real-world Operational Insights from RGW Deployments**

**Author:** kritik sachdeva<sup>1</sup>

<sup>1</sup> IBM

Operating Ceph RADOS Gateway (RGW) as a backend for sync & share services presents unique challenges that differ from typical S3 workloads. Drawing from hands-on experience supporting RGW deployments across diverse production environments, this presentation shares practical operational insights that service operators can immediately apply.

We examine the most common issues encountered in RGW deployments backing sync & share platforms: from bucket index sharding misconfiguration and multisite sync lag to performance bottlenecks during metadata-heavy operations typical of EFSS workloads. Each issue is presented with its symptoms, root cause analysis approach, and resolution pattern.

The presentation introduces a systematic troubleshooting framework developed through supporting multiple deployments, covering diagnostic tools (radosgw-admin, ceph health, log analysis), key metrics to monitor, and decision trees for common failure scenarios. We discuss preventive measures including capacity planning considerations, configuration best practices for sync & share workloads, and monitoring strategies that provide early warning of emerging issues.

Attendees will leave with actionable knowledge to improve the reliability and performance of their RGW-backed storage infrastructure, whether they are planning a new deployment or operating an existing one.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Interoperability: protocols, APIs, OpenCloudMesh (OCM) / 37**

## **OpenCloudMesh towards IETF Standardization**

**Author:** Giuseppe Lo Presti<sup>1</sup>

<sup>1</sup> CERN

Open Cloud Mesh (OCM) is the protocol for federating storage adopted by the EOSC Federation.

- Past year activity, IETF Dispatch and WG
- Current status (developments likely covered by Mahdi)
- Plans - campfire discussion to be prepared with the vendors, to get their engagement going forward

**Suggested Contribution Type:**

Panel discussion

## Data Science Environments & HPC integration / 38

### Data Commons and Jupyter integration in CERNBox

**Authors:** Diogo Castro<sup>1</sup>; Rasmus Oscar Welander<sup>None</sup>

<sup>1</sup> CERN

CERNBox is CERN's open-source Enterprise File Sync&Share service and uses the Open Cloud Mesh (OCM) protocol to support federated sharing across independently operated sites.

EOSC Data Commons will provide building blocks that connect discovery, selection, and execution of research-data workflows, including services such as the EOSC Matchmaker and the EOSC Data Player. Through the project, such services are expected to enable the researcher to select a dataset (or set of datasets) and deliver it to a working environment for further processing: in this context, CERNBox plays the role of the working environment for researchers, leveraging its storage such as CephFS and its integration with JupyterLab.

In this talk, we present the integration work being done in CERNBox, where discovery results are turned into a package for processing and delivered into a user-accessible workspace via OCM-based federation mechanisms. The objective is to make it straightforward for researchers to bring selected datasets into CERNBox, run analysis in Jupyter-based environments and HPC resources close to the data, and then publish resulting outputs to repositories with minimal manual data handling.

#### Suggested Contribution Type:

Regular Talk (15-30 min)

## CS3 Community Site Reports / 39

### CERNBox service evolution towards a federation-enabled EFSS

**Authors:** Diogo Castro<sup>1</sup>; Jesse Geens<sup>None</sup>

<sup>1</sup> CERN

CERNBox is CERN's Enterprise File Sync&Share service, serving over 12 petabytes (spread over 4 billion files) to more than 16 000 monthly users. CERNBox is based on Reva, an interoperability server for cloud storages. In this contribution we will touch on the improvements that have been made to Reva over the last year, and on our operational experience running Reva as part of CERNBox.

Specifically, we will talk about:

- Full support for Spaces, DAV-compatible
- A full end-to-end test suite, based on Playwright
- A new Ceph driver
- A reva discourse server, for support and sharing configurations
- Outlook for next year

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**File Sync & Share Solutions and Requirement from the Community / 40****Nextcloud. State of the nation****Author:** Frank Karlitschek<sup>None</sup>

This talk will give an overview of the Nextcloud developments and improvements in the last 12 month. Several noteworthy things happened in the last Nextcloud releases. From architectural improvements to changes on APIs and the sync engine, to useability and functionality. This Talk will give a full overview.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**AI and Storage / 41****What's Next in AI for Business? Pushing the analog|digital interface boundary.****Author:** Axel Koester<sup>1</sup><sup>1</sup> *IBM Technologist*

What direction is "AI for business" heading, e.g. for project management, automated software deployment, automated quality control, and resource management?

How can agentic AI transform our work? What are "agentic skills"?  
Whose jobs will be impacted?

From the big headlines down to Ceph's MCP interface (*what's that??*), we will attempt to predict the impact of AI on multiple levels and see what's next.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Applications handling sensitive data, data classification, data security and privacy / 42****Expect the (un...-)expected" - IT business continuity in an increasingly uncertain world****Author:** Tilo Uwe Steiger<sup>1</sup><sup>1</sup> *ETH Zürich*

We are seeing cybercriminals becoming increasingly sophisticated. This is to be expected, as there is a lot of money to be made for criminals from imminent cyber incidents. Unexpected downtime or corrupted business processes cause significant damage to organizations and a loss of data integrity for society.

Reducing the imbalance between digital threats and the resilience of IT-based business processes, and ensuring that these processes can be quickly restarted, therefore remains a major challenge.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Scalable Storage Backends and Integration with Data Processing / 44**

## **Feasibility study trans-national replicated object storage using Ceph**

**Authors:** Adam Prycki<sup>None</sup>; Joni Junni<sup>None</sup>

**Co-authors:** Bo Bai ; Eugeniusz Pokora ; Jan Meijer ; Jean-Marie de Boer ; Kalle Happonen ; Maciej Brzeźniak ; Morten Kjeldgaard ; Ron Trompert

**Title:** Feasibility study trans-national replicated object storage using Ceph

**Subtitle:** Testing key technical components of a potential resilient pan-European object storage infrastructure

**Abstract:** The European NREN community is working on a proposal for a large scale sovereign object storage infrastructure on European scale. To test key technical aspects, four NRENs are currently conducting a feasibility study. In this talk I will share our experience from testing Ceph S3 object storage in a multi-domain scenario with cross-border replication between PSNC (Poland), CSC (Finland) and SURF (Netherlands)

**Audience:** those interested in technical scenarios for large scale pan-European sovereign object storage infrastructure.

**Proposal:** Currently the European NREN community does not have large scale, sovereign, object storage. In recent years we have seen that data sovereignty becomes more and more important with new data protection laws and funding decisions on the other side of Atlantic causing hastily efforts to copy data to Europe. In GÉANT GN5-2 Work Package 4, Task 3 we have been working on testing the feasibility of using Ceph for creating a coherent pan-European object storage infrastructure spanning different administrative domains which could provide us robust storage for scientific data.

Ceph RadosGW object storage has been “on the market” for a long time and has seen many features added to it over the years. It is currently the most feature rich and scalable open source s3 implementation with advanced asynchronous data replication. Standalone ceph s3 clusters are quite common and many NREN institutions have them on premises. Replicated RadosGW deployments are not as popular and challenges of such systems are not well documented. We know of at least one large scale big large scale Ceph s3 multisite deployment (Over 70PB of raw HDD space) at American company Bloomberg which shows that it can be done with enough patience and time. We want to build something similar.

For the past few months we have been working on creating proof of concept for similar large scale object storage system which would replicate data between our 3 institutions. In this talk I would want to share our experiences from these tests and what kind of technical management issues does one encounter when building such multi-organisation systems.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

Federations for EOSC and eResearch infrastructures. / 45

## Onedata as Unified Data Namespace for the Polish EOSC Node

**Authors:** Michał Orzechowski<sup>1</sup>; Łukasz Opiola<sup>2</sup>; Łukasz Dutka<sup>3</sup>

<sup>1</sup> AGH University of Science and Technology, Academic Computer Centre Cyfronet AGH, Krakow, Poland

<sup>2</sup> Academic Computer Centre Cyfronet AGH

<sup>3</sup> Academic Computer Centre Cyfronet AGH,

Onedata 1 is a high-performance, distributed data management system designed for global infrastructures. It provides seamless access to heterogeneous storage resources and supports diverse use cases, ranging from personal data management to large-scale scientific computations. By leveraging a fully distributed architecture, Onedata facilitates the creation of hybrid cloud environments that integrate private and public cloud resources. The system enables users to collaborate, share, and publish data while supporting high-performance computations on distributed datasets via various interfaces, including POSIX-compliant native mounts, PyFS (Python filesystem) plugins, REST/CDMI APIs, and S3 protocol.

Recent advancements in Onedata include improved data publishing features and added support for mainstream metadata formats (such as DataCite or OpenAIRE), maturing of S3 protocol support that now covers automated deployment of the S3 endpoints across a Oneprovider cluster, and optimizations to handle large numbers of data collections (spaces). Moreover, the ongoing EOSC DataCommons project is improving Onedata's capacity to serve as data and metadata integration middleware that bridges various data repositories and sources with execution engines and VREs, providing interoperability and FAIR compliance.

Within EOSC Node | Poland 2, Onedata serves as the common data layer that brings together datasets from different repositories, providing a unified data platform and workspace for researchers. EOSC-PL leverages Onedata's standardized interfaces to enable seamless data flows between distributed European research infrastructures. A key capability is the support for importing data by reference rather than copying, allowing datasets from heterogeneous sources such as EEA 3 or the Polish eCUDO platform to be served on demand directly from their origin servers, while being logically integrated in Onedata unified namespace. Having organized data from different sources into a Onedata space, users may run analysis on it using any compatible environment. An example showcasing collaboration with different EOSC Nodes is running workflows using Galaxy EU 4 that seamlessly access input data from the Onedata namespace, thanks to native integration between the two platforms. It supports deferred dataset resolution during data import, minimizing redundant transfers and storage quota consumption. After results are obtained, Onedata enables researchers to publish them with automatically minted PIDs/DOIs, which are regularly harvested via the OAI-PMH protocol and aggregated in the EOSC-PL catalogue. This closes the research data lifecycle—from discovery through processing to publication—while supporting provenance records using RO-Crate 5.

In addition to EOSC-PL and EOSC DataCommons, Onedata is currently deployed in several European projects such as DOME 6 and SPICE 7. There, it provides a data transparency layer for managing large, distributed datasets in dynamic, hybrid cloud environments with containerized deployments.

**Acknowledgements:** This work is co-financed by the Polish Ministry of Education and Science under the program entitled International Co-financed Projects (5399/DIGITAL/2023/2) and co-funded by the Government Office of the Slovak Republic within the European Union NextGenerationEU, Recovery and Resilience Plan (project no. 09I02-03-V01-00012).

We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grants no. PLG/2025/018992, PLG/2025/018994, PLG/2026/019224.

## REFERENCE

- 1 Onedata. <https://onedata.org>.
- 2 EOSC Node | Poland, Polish Open Science Platform. <https://eosc.pl>.
- 3 European Environment Agency - SDI - geospatial data catalogue. <https://sdi.eea.europa.eu>.
- 4 Creating Workflows and Advanced Workflow Options. <https://galaxyproject.org>.
- 5 Research Object Crate. <https://www.researchobject.org/ro-crate>.
- 6 DOME: A Distributed Open Marketplace for Europe Cloud and Edge Services. <https://dome-marketplace.eu>.
- 7 Smart data Pipelines for the Cognitive Compute Continuum. <https://spice-platform.eu>.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Applications handling sensitive data, data classification, data security and privacy / 46****TSD - a part of Norwegian Trusted Environment Consortium**

**Author:** Dagfinn Bergsager<sup>None</sup>

**Co-author:** Anne Bergsaker<sup>1</sup>

<sup>1</sup> *University of Oslo*

NORTRE is a collaboration between three main institutional research infrastructures for sensitive data in Norway. We share knowledge and expertise so scientists and data controllers from Norway and around the world can collect, analyze, store, share and collaborate on sensitive data in an optimized and trustworthy manner.

As the largest TRE in Europe, TSD will in this presentation give a brief introduction how to maintain and develop an infrastructure with over 12000 active users, 2500 projects and 16 pb of data.

There will also be demonstrations of how TSD now uses AI on sensitive data.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Lightning talks / 47****Local AI-models at University of Oslo**

**Authors:** Dagfinn Bergsager<sup>None</sup>; Maiken Pedersen<sup>None</sup>

UiO offers students and employees a set of AI-services hosted locally here at Campus Blindern.

This is a collaboration between other universities in Norway that use supercomputers for research, and we offer AI-services via GUI and API. Local AI models provide much flexibility and higher privacy and security and this presentation will demonstrate how.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Lightning talks / 48****Nextcloud deployment at the University of Oslo****Author:** Hagen Echzell<sup>None</sup>

UiO is currently running a pilot project with our own Nextcloud deployment. We'll give a brief overview over the current status, our future plans and discuss our container-based deployment architecture.

**Suggested Contribution Type:**

Lightning Talk (5 min)

**Federations for EOSC and eResearch infrastructures. / 49****Interoperable Infrastructure for Universities, Enterprises and the Public Sector****Authors:** Hagen Echzell<sup>None</sup>; Terje Sylvarnes<sup>None</sup>

Many organizational processes cross the boundaries of individual organizations. This is especially true for the academic sector where mobility programs like "Erasmus+" enable students to be enrolled with different universities throughout their studies. However, the technical means to enable the necessary collaboration between different institutions and their administrations are often not adequate. Many tools for group messaging, file exchange, or domain-specific applications do not work or lack functionality as soon as users or data from outside an organization are involved, such that E-Mail is a widely used fall-back.

Based on terminology from the recent Interoperable Europe Act - (IEA) we will present technical and semantic interoperability solutions to address these issues. On the technical side, we present two existing protocols: OpenCloudMesh is an open file sharing protocol implemented by open-source solutions like Nextcloud and ownCloud. The Matrix protocol enables secure, decentralized group and instant messaging, but might also support further use cases like video conferences in the future. On the semantic side, we present a case study where we, based on existing vocabularies, model a university course catalog in the Resource Description Framework (RDF) serialization format JSON-LD. The RDF representation of the course data introduces a semantic context to the course data, making the exchange of the catalog data independent of specific APIs, thus enabling interoperability across heterogeneous data structures and university-specific systems.

The introduced technical and semantic interoperability solutions enable public sector bodies like universities and enterprises to improve their cross-organizational collaboration without giving up their digital sovereignty to a centralized, commercial actor.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Lightning talks / 50**

## Data Access Broker: machine-actionable workflows to access sensitive data

**Authors:** Ahmad Hesam<sup>1</sup>; Claudio Cacciari<sup>None</sup>

<sup>1</sup> SURF

The use of data that is sensitive of nature is becoming increasingly more prominent. In research fields, such as life sciences, social sciences and humanities, the most valuable insights are obtained by analyzing sensitive data. The access to sensitive data however is extremely tedious and has shown to slow down or even block research projects.

At SURF, the national infrastructure provider for research and education in the Netherlands, we are working on a prototype service that aims to simplify and speed up the access to sensitive data. In this talk we wish to shed light on the Data Access Broker, which standardizes the process of accessing data stored in various storage systems and make them available in (cloud-based) Trusted Research Environments (TREs). TREs are controlled remote compute environments that allow for the data provider to remain in full control of the data while still allowing the data consumer to analyze the data. The Data Access Broker fills the crucial role of bringing data to compute environments which are currently out of bounds because of their sensitive nature.

Part of this talk will be a live demonstration of the end-to-end workflow of the Data Access Broker, both from the perspective of the data provider as well as the data consumer.

(This abstract is submitted for a Regular Talk, but can be adapted to a Lightning Talk if needed)

### Suggested Contribution Type:

Regular Talk (15-30 min)

### Lightning talks / 51

## Nextcloud Enterprise deployment for Kubernetes, OpenShift and Podman

**Author:** Sebastian Möbus<sup>None</sup>

Back in the early days apps and services were deployed manually. This led to errors and misconfiguration and flaky apps.

This changed in the early 2000 years with the rise of virtualization and hypervisors. Shared resources and better load spread on one hand at the cost of overhead and complexity though.

Since 2010 cloud native approaches are dominating the deployment with orchestrated containers/pods and CD/CI pipelines for automated and scalable deployments.

The ongoing requests of our prospects and customers for this kind of deployment was heard from Nextcloud and therefore we decided to create our own, fully supported and cloud native Nextcloud Enterprise deployment called Enterprise-AIO which is built around Kubernetes for big instances that need to scale.

It's your choice whether to go for a ready-to-use full flavored Nextcloud Hub with all components or to connect to your own services which might be already in place. Either way Enterprise-AIO allows you to choose your platform, your apps and services, airgapped scenarios and much more.

On top of that we also provide a well approved paths for migrating from other platforms to Nextcloud together with the chance of pushing your deployment to the next level.

### Suggested Contribution Type:

Lightning Talk (5 min)

## CS3 Community Site Reports / 52

### Sync&Share Storage since 2013

**Author:** Mihajlo Gajic<sup>None</sup>

**Co-author:** Tilo Uwe Steiger

ETH Zürich starts offering their Sync&Share Storage in 2013 –until now it is still growing and very popular among researchers, lecturers and the administrative staff.

In this talk we will show some technical aspects, numbers and university use cases being enabled by the Sync&Share Storage.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

## Federations for EOSC and eResearch infrastructures. / 53

### The Czech Path to a Community-Driven Research Data Infrastructure

**Author:** Matej Antol<sup>1</sup>

<sup>1</sup> *Masaryk University*

The Czech Republic is developing a national implementation of the European Open Science Cloud aimed at establishing a sustainable National Data Infrastructure (NDI) for FAIR research data. The initiative builds on the long-standing national e-infrastructure e-INFRA CZ, as well as domain-specific infrastructures, and actively involves scientific communities through twelve EOSC CZ working groups, ensuring that the resulting services reflect real research needs.

The NDI combines large-scale, distributed storage, a shared software platform, and domain-specific repositories aligned with major research clusters. These repositories are integrated with national computing, networking, and cybersecurity environments, enabling both discipline-specific and cross-disciplinary research workflows. Alongside technical components, the initiative addresses data stewardship by embedding trained data stewards and curators across research organisations and by providing coordinated national training.

This Czech chapter of the EOSC initiative is supported by national and European funding. The emerging ecosystem delivers core services such as authentication, persistent identifiers, metadata catalogues, and governance frameworks. Its central component, the National Repository Platform, provides a federated, FAIR-by-design repository backbone tightly connected to European and global data infrastructures and services.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

## CS3 Community Site Reports / 54

## Distributed data sharing infrastructure for scientific institute based on Onedata

**Authors:** Adrián Rošinec<sup>None</sup>; Tomáš Svoboda<sup>1</sup>

<sup>1</sup> *Masaryk University*

CEITEC, a research institute established at Masaryk University, hosts multiple core facilities and research groups that jointly produce large volumes of heterogeneous scientific data across diverse scientific domains. These activities require systematic recording of experimental outputs and their timely delivery to researchers for analysis and interpretation.

A key requirement is controlled and reliable data handover. Ad-hoc transfers via USB drives or personal workstations lead to data fragmentation, loss of provenance, and limited reproducibility, and therefore must be replaced by centralized storage on institutional infrastructure. At the same time, users expect an access model comparable to contemporary cloud storage services such as Google Drive or Dropbox, combined with seamless integration into institutional compute environments. Keeping data close to computational resources—such as Kubernetes clusters, Jupyter notebooks, or Galaxy workflows—eliminates repeated manual copying and significantly lowers the barrier to data analysis.

Certain data-intensive experiments, for example in cryo-electron microscopy, impose additional constraints. Measured data must be made available to users while the acquisition is still in progress, enabling real-time inspection and timely adjustment of instrument parameters. This combination of real-time access, controlled sharing, and tight coupling to compute infrastructure creates strict requirements on data management and sharing systems.

To address these challenges, we adopted the Onedata platform as the foundation of a distributed data sharing infrastructure. Onedata provides unified access, controlled sharing, and transparent integration of distributed storage resources while preserving data locality and performance. This contribution presents the concrete workflows deployed at CEITEC, describes their integration into everyday laboratory and computational practices, and reports operational experience from running the service in a multi-facility, multi-disciplinary research institute.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

## Scalable Storage Backends and Integration with Data Processing / 55

## Bringing legacy (filesystem) storage systems to the S3 cloudy world

**Authors:** David Manuel Garcia Aguilera<sup>1</sup>; Krzysztof Błoniarz<sup>2</sup>; Krzysztof Wadówka<sup>2</sup>; Maciej Brzezniak<sup>None</sup>; Piotr Modrzyk<sup>3</sup>

<sup>1</sup> *Leil Storage*

<sup>2</sup> *PSNC*

<sup>3</sup> *Leil Storage SaunaFS*

S3 protocol is the de-facto standard data protocol in distributed storage and cloud computing with its wide usage for cloud back-ends, back-up and long-term storage. It is also being adopted in HPC computing systems and AI platforms. This trend is seen in academia and is strengthening in industry with e.g. nVidia assuming S3 to be the default data protocol for accessing data from GPUs for AI data intensive workflows.

S3 services are provided using variety of products and solutions, including proprietary Amazon S3 cloud infrastructure, closed appliances (EMC's Isilon, DDN's Infinia, NetApp Storage Grid to name the few), as well as open source platforms including Ceph with its S3 RADOS gateway, OpenStack Swift's S3 interface, and standalone products and gateways such as MinIO or Versity.

Despite overall storage 'cloudification', most storage systems in industry and academia is still based on filesystems, including open source ZFS, Lustre (cloud, HTC, HPC, AI) and commercial "usual suspects" DAOS, GPFS, BeeGFS (mostly used in HPC). An open issue is how to enable S3 access to those "legacy systems" in an efficient, scalable way while ensuring non-disruptive filesystems operation, reuse of existing infrastructure and knowledge and avoiding the costly and cumbersome data migration from "legacy" to "modern" systems or duplicated storage of datasets to ensure both S3 and POSIX access.

In our work we focus on examining compatibility and performance of gateway-based system comprised of SaunaFS with HDD disk storage back-end (including CMR and SMR drives) and VersityGW providing S3 protocol integration. The experiment has been part of the technology evaluation conducted by PSNC within the nationally funded R&D project on data center and distributed storage.

Our work proves that such a combined architecture does introduce significant performance limits or overheads on the throughput (GB/s) and I/O capability (IOPS) of the combined storage system compared to its native capabilities measured and the filesystem level.

Our results also show that such a integration can be deployed and configured in reasonable time, with minimum effort, and - what is very important for the pre-existing systems and datasets - with no negative impact on filesystem operations nor intrusion or reconfiguration needed in the filesystem. In particular VersityGW we have tested seems to enable coexistence of POSIX filesystem interface next to S3 by providing an extension to existing, legacy system.

#### **Suggested Contribution Type:**

Regular Talk (15-30 min)

### **CS3 Community Site Reports / 56**

## **First experience with OpenCloud in PSNC**

**Authors:** Eugeniusz Pokora<sup>1</sup>; Jan Bródzia<sup>1</sup>; Krzysztof Wadówka<sup>1</sup>; Maciej Brzezniak<sup>1</sup>

<sup>1</sup> PSNC

OpenCloud as a new star in the sync & share open source products constellation that is in the center of interest of the CS3 community for its capability to build and operate sovereign, large scale data platforms for file sharing and data-based collaboration.

OpenCloud was created as a next-generation sync & share software stack, fully based on a microservices architecture. Since the well-known fork, it has undergone significant development aimed at strengthening functionality, scalability, and market presence.

Since 2015 PSNC operates several sync & share platforms and services based on various software stacks: Seafiler for the country-wide services for academia (box.pionier.net.pl), Nextcloud for some internal projects and user groups and ownCloud for the purposes of EOSC EU Node.

OpenCloud gained our interest as the potential next stage EFSS platform for country-wide service for academic users and as the EFSS platform that is architecturally feasible for integration with compute and processing environments including: cloud, HPC and AI platforms. Its open architecture, allows developing functional extensions as micro-services also enables integrations with data repository services, open data systems, FAIR-oriented tools, and large scale long-term storage systems, possibly including tape based cold storage.

In our work we focus on the gaining the first impressions of using and operating OpenCloud within the proof of concept initiative that started in 2025/2025 with the intention to extend the scope of work during 1-2Q of 2026. Our evaluation and testing included deployment and back-end storage integration features and will include extensive performance testing.

While our work is still a work in progress, we are willing to share the early results and impressions with the CS3 community. As OpenCloud seems to be a promising technology for dealing with our 1PB+ datasets under sync & share, we are planning to test OpenCloud at relevant scale of multiple TBs and under the high I/O stress, already in the early stage of evaluation work.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**CS3 Community Site Reports / 57**

## **DESY Sync & Share: A dCache-Based Federated Storage Service with Multi-Protocol Access, AI Assistance, and Full-Text Search**

**Author:** Peter van der Reest<sup>1</sup>

**Co-authors:** Christian Voss <sup>1</sup>; Tigran Mkrtchyan <sup>1</sup>; Tim Möller <sup>1</sup>

<sup>1</sup> *DESY*

DESY's **Sync & Share (S&S)** service provides a **scalable, federated storage solution** for both **research data management** and **organisational storage**. Built on **dCache**, it supports **high-throughput, highly available and highly resilient data access** while enabling scientific collaboration. The service has evolved to meet growing demands for **multi-protocol support, efficient data discovery**, and the **integration of state-of-the-art technologies**.

This presentation will examine the **use cases, architecture, and technical status** of the S&S service, which serves not only DESY communities but also various other communities across the Helmholtz Association of German Research Centres.

Additionally, we will describe the **core integrations within the DESY IT landscape** that are currently in production.

Finally, the talk will outline **future development efforts** aimed at enhancing the utility of DESY Sync & Share.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Digital Sovereignty / 58**

## **EFSS as Infrastructure: Foundational Layer of Europe's Sovereign Data Stack**

**Author:** Peter Szegedi<sup>1</sup>

<sup>1</sup> *European Commission*

Europe's ambition for digital sovereignty is no longer an abstract policy goal. It is being operationalized through concrete, interoperable infrastructures that span research, data, compute and AI. This keynote situates the European Open Science Cloud (EOSC) within its broader European Commission policy context and shows how Enterprise File Sync and Share (EFSS) is evolving from a productivity service into a critical, shared digital infrastructure. A particular focus will be placed on SIMPL as open-source middleware enabling secure, scalable data and AI pipelines, and on the growing integration of EOSC with EuroHPC and AI Factories. Within this landscape, EFSS - exemplified by the EOSC EU Node and its evolution towards the EOSC Federation - emerges as a common, always-available layer for data access, collaboration and automation across Europe.

A central enabler of this transformation is Open Cloud Mesh (OCM). Originating in the CS3 community and now progressing through standardization in the IETF, OCM provides a vendor-neutral, policy-aware interoperability layer for federated file sharing. Its recognition within EOSC positions OCM as a practical pathway into SIMPL, European Data Spaces and, more broadly, Europe's sovereign middleware stack. The keynote situates these developments within the European Commission policy landscape and their alignment with EOSC, SIMPL, Destination Earth and EuroHPC-enabled AI workflows. It argues that EFSS, powered by open standards such as OCM, is becoming a critical connective layer between data, compute and AI—turning federation from a policy aspiration into an operational reality.

**Suggested Contribution Type:**

Regular Talk (15-30 min)

**Keynotes / 60**

## **Digital Sovereignty or Digital Dependency: Europe's Tech Moment of Truth**

**Author:** Frank Karlitschek<sup>None</sup>

Amid intensifying geopolitical tensions, Europe is accelerating its drive toward digital sovereignty. Technology providers, enterprises, and public institutions are being reshaped by the continent's push for technological independence supported by an expanding regulatory framework that now spans the DMA, DSA, GDPR, NIS2, and upcoming European cybersecurity initiatives.

This keynote examines how these developments are redefining the landscape for collaboration platforms and File Sync & Share solutions technologies that sit at the core of modern productivity and information security.

We will explore what the next generation of collaboration software must deliver to thrive in this environment: decentralised and federated architectures, robust Open Source ecosystems, privacy respecting AI by design, transparent data flows, and open standards that empower organisations rather than lock them in.

Attendees will gain a behind-the-scenes perspective on how European tech policy is shaping the future of digital collaboration along with insights and lessons learned from more than 16 years of building Open Source collaboration technologies.

**Suggested Contribution Type:**

**Keynotes / 61**

## **The HPC-Cloud Pipeline: Data Provisioning for Destination Earth and the AI Revolution in Weather Forecasting**

**Author:** Tiago Quintino<sup>1</sup>

<sup>1</sup> *ECMWF*

The European Commission's Destination Earth (DestinE) initiative represents a paradigm shift in Earth system simulation, aiming to develop highly accurate digital twins of the Earth. At the core of this endeavour lies the Digital Twin Engine (DTE), an innovative software framework designed to connect the extreme data generation capabilities of High-Performance Computing (HPC) with the interactive, user-centric flexibility of cloud environments.

This keynote discusses the architectural challenges and solutions involved in designing the DTE to enable seamless data provisioning and sharing. A key focus will be the convergence of Numerical Weather Prediction (NWP), Climate Information, and Machine Learning (ML). We will examine how the DTE is evolving to support "AI-ready" datasets, particularly addressing the extensive data handling requirements for ML training, including the upcoming ERA6 reanalysis—the successor to ERA5 and a vital component for future AI model training. Additionally, we will outline the data throughput challenges related to operationalising ECMWF's Artificial Intelligence Forecasting System (AIFS) and explain how we are developing scalable workflows to support the next generation of data-intensive prediction systems.

**Suggested Contribution Type:**

**Keynotes / 62**

## **Clinical research and data in the AI era: Lessons and future directions from the AI-Mind project**

Concluded in February 2026, the ambitious AI-Mind project (Grant ID: 964220) was a pioneer in the new era of clinical AI research. During the five-year project period, the initiative successfully established one of the world's most comprehensive longitudinal datasets focused on Mild Cognitive Impairment (MCI) and dementia progression.

Collected through harmonised procedures across Norway, Finland, Italy, and Spain, the dataset covers over 1,000 individuals. It comprises more than 3,500 EEG recordings and digital cognitive assessments, 1,800 blood biomarker samples, and nearly 1,000 diagnostic conclusions. However, the existence of valuable data is only the first step; preparing such complex, sensitive information for clinical AI applications demands robust engineering combined with domain expertise.

In this talk, Dr. Hatlestad-Hall details how the AI-Mind project leveraged and collaborated with the Services for Sensitive Data (TSD) at the University of Oslo to build a comprehensive research platform for data curation, quality assurance, and AI model development. The talk will highlight critical lessons learned and identify essential focus areas for the future co-creation of research infrastructure that bridges the gap between clinical research and technology development.

**Suggested Contribution Type:**

**Interoperability: protocols, APIs, OpenCloudMesh (OCM) / 63**

## **Panel**

**Lightning talks / 64**

## **A Decade of Keeper: 10 Years of Secure Sync, Share & Long-Term Archiving in the Max Planck Ecosystem (2016–2026)**

**Author:** Vladislav Makarenko<sup>1</sup>

**Co-author:** Iosif Peterfi<sup>1</sup>

<sup>1</sup> *Max Planck Information and Technology (MaxIT)*

Since its 2016 launch by the Max Planck Digital Library, **Keeper** —a heavily customized Seafile platform —has delivered secure file synchronization, collaborative sharing, and automatic long-term archiving (guaranteed  $\geq 10$  years) for research datasets. It supports seamless project workflows while preserving critical materials (e.g., raw data, scripts, credentials, documentation) for institute members and invited collaborators.

Key milestones include blockchain-based provenance via bloxberg, DOI minting for archived libraries, strengthened encryption, refined group controls, the Project Catalog for enhanced discoverability of archiving-ready libraries, and the Cared Data Certificate attesting to proper data care and long-term usability.

By early 2026, Keeper serves over 8,500 active users and stores hundreds of terabytes. Keeper has become an essential component in the daily work of MPG scientists and is actively used in more than 80 Max Planck Institutes.

**Suggested Contribution Type:**

Lightning Talk (5 min)