

CS3 2022 - Cloud Storage Synchronization and Sharing

Monday, January 24, 2022 - Thursday, January 27, 2022



Book of Abstracts

Contents

| | |
|--|---|
| Introduction and Welcome | 1 |
| IPFS: Interplanetary filesystem | 1 |
| Summary of CS3 Community Site Reports | 1 |
| Moving sciebo to kubernetes: Lessons learned and practical considerations for productive workloads | 1 |
| CERN Site Report: CERNBox Horizon 2030 | 1 |
| SUNET Drive - Status and plans for Swedens storage solution | 2 |
| Our road towards self-service sync-and-share | 2 |
| Seafile 9.0 and Beyond | 3 |
| Nextcloud - State of the nation | 3 |
| Infinite Scale - A new era for the ownCloud project | 3 |
| OCM - next steps? | 4 |
| OCM test suite | 4 |
| ScienceMesh - Invitation Workflow Implementation | 4 |
| An OCM protocol extension to support data transfer in the mesh | 4 |
| Some thoughts about OCM-over-ScienceMesh | 5 |
| Group-owned shares | 5 |
| Experiencing a new Internet Architecture | 5 |
| Update from the European Commission on Future Evolution of Digital Landscape in Europe | 6 |
| ScienceMesh: An Interoperable Federation of EFFS services for European Open Science Cloud | 6 |
| HIFIS: VO Federation for EFSS | 6 |
| ScienceBox 2.0 | 7 |
| ownCloud Infinite Scale - Identity, Roles and Permissions | 8 |

| | |
|---|----|
| Introducing smart document forms for paperwork automation | 8 |
| Collabora Online: easy to deploy and manage document collaboration | 8 |
| Breaking the limits: short status update of the online spreadsheet solution SeaTable | 9 |
| Introduction & Goals | 9 |
| CERN viewpoint | 9 |
| ScienceMesh and EOSC | 10 |
| Quo Vadis CS3 Community? | 10 |
| CS3 ORG Governance Discussion | 10 |
| Digital Market: a level playing field for EU Tech sector | 11 |
| Converging Storage Layers with Virtual CephFS Drives for EOS/CERNBox | 11 |
| Sync and Share Access to HPC Resources at CERN | 11 |
| The CERN Tape Archive : Archival Storage for Scientific Computing | 12 |
| JupyterLab+ScienceMesh: Collaborative Data Science in sync-and-share environment. | 12 |
| iRODS Research Community Requirements Drive Expanded Scale Data Management Features | 13 |
| Infinite scale is a design principle | 13 |
| Welcome and Objectives: CS3 community and ScienceMesh | 14 |
| Science Mesh – where we are now with the CS3MESH4EOSC project? | 14 |
| Science Mesh for EFSS service providers | 14 |
| ScienceMesh-Nextcloud | 14 |
| ScienceMesh-Owncloud OCIS | 14 |
| ScienceMesh-Seafile | 15 |
| Technology & Development - Advancements & Innovations | 15 |
| Researchers and Use-cases | 15 |
| Reframing adoption challenges in FAIR Data Infrastructures: Science Mesh as a source of research advantage. | 15 |
| ScienceMesh 2022: What’s next and where do we take it from here | 15 |
| EOSC and Science Mesh: overcoming data challenge (EOSC Association and TFs) | 15 |
| Scientific disciplines embracing no border Research Environment thanks to Science Mesh | 16 |
| CS3MESH4EOSC Wrap-Up and Next steps | 16 |

| | |
|---|----|
| Solid storage and pod migration [CANCELLED] | 16 |
| ABEBox: end-to-end encryption for file sharing cloud services | 17 |
| Is EOS ready for enterprise companies? | 17 |
| CERNBox User Forum: Ultimate Engagement with Users | 18 |
| “Find the needle in the haystack”: how log monitoring, analyze and error handling can be done with SeaTable. | 18 |
| ownCloud Web UI: Lessons learned from implementing accessibility | 19 |
| Using Workflows for Data Preservation Using Onedata | 19 |
| ownCloud WOPI Proxy for O365 | 20 |
| Finding an optimal approach to enabling document collaboration with ONLYOFFICE Docs in integrated environments: interoperability, decentralization and limitations | 20 |
| integrate applications with the application provider | 21 |
| Summary and Conclusions | 21 |
| CS3APIs: governance and next steps (Campfire Discussion) | 21 |
| ScienceMesh workshop | 21 |

Introduction & Welcome / 59**Introduction and Welcome**

Author: Jakub Moscicki¹

¹ *CERN*

Corresponding Author: jakub.moscicki@cern.ch

Keynote / 31**IPFS: Interplanetary filesystem**

Corresponding Author: yiannis@protocol.ai

Site Reports / 43**Summary of CS3 Community Site Reports**

Author: Dr. Tilo Steiger¹

¹ *ETH Zuerich*

Corresponding Author: steiger@id.ethz.ch

Site Reports / 20**Moving sciebo to kubernetes: Lessons learned and practical considerations for productive workloads**

Authors: Holger Angenent¹; Marcel Wunderlich^{None}

¹ *University of Münster*

Corresponding Author: wunderlich@uni-muenster.de

At Sciebo we migrated the first half of our productive ownCloud instances, serving over 200k customers across the state of North Rhine Westphalia at university institutions, to our new on-premise kubernetes platform.

Last year we presented an overview of the rough architecture of the platform and promised some more insights for this year's CS3. ;-)

In this presentation we

- give a quick reminder consisting of little lies how to conceptualize all this kubernetes stuff
- discuss some choices we made in regard to our tooling
- mention some patterns and anti patterns we identified in the wild
- some practices and mantras that served us well
- address the elephant in the room and talk about some things we did roll on our own in order to move our already existing services to the cloud
- the road ahead

Site Reports / 28**CERN Site Report: CERNBox Horizon 2030****Author:** Ishank Arora¹**Co-author:** Hugo Gonzalez Labrador ¹¹ *CERN***Corresponding Authors:** ishank.arora@cern.ch, hugo.gonzalez.labrador@cern.ch

CERNBox is key enabler service for users at CERN and beyond. The service is used by more than 37K users and stores over 15PB of data, representing all the user communities at the laboratory.

In this talk we will explain the current status of the service, the challenges we faced in 2021 and we look into the future: CERNBox as the gateway for heterogeneous storage spaces at CERN and beyond.

Site Reports / 13**SUNET Drive - Status and plans for Swedens storage solution****Authors:** Micke Nordin¹; Richard Freitag^{None}¹ *SUNET***Corresponding Authors:** kano@sunet.se, freitag@sunet.se

SUNET is currently establishing SUNET Drive as their solution to store and share large amounts of scientific data. The architecture is based on a global scale setup of Nextcloud, where each university and college gets an own node, which then can be customized. The underlying storage infrastructure is based on S3 containers, and each university can manage and assign new buckets depending on their needs. The goal is to establish a service providing data-sovereignty, while being part of a larger federation of storage-services.

This community site report will focus on the current status of SUNET Drive, its level of automation to achieve a scalable solution, as well as challenges and issues to get SUNET Drive to where it currently is.

Site Reports / 32**Our road towards self-service sync-and-share****Author:** Narges Zarrabi¹¹ *SURF***Corresponding Author:** narges.zarrabi@surf.nl

Since a few years SURF has been running a sync-and-share service called Research Drive next to the personal storage based SURFdrive. Research Drive is specially tailored for the special needs of researchers. These special needs had to do with flexible quota, project-based storage rather than personal storage and multiple means of authentication. The latter was an absolute necessity in order to allow people accessing the service outside of the Dutch SURFconext identity federation. On one hardware infrastructure now almost 30 instances are running of a sync-and-share service for equally many institutes.

The institutes use this service to manage their research data. Apart from the regular users like students, teachers and researchers there are also the departmental administrators, central IT and primary investigators. Each having their role in the research data management process. Since Research Drive aims to be self-service we have developed a dashboard where these different roles have been implemented, each with the different capabilities suiting their role. The dashboard allows users to invite other users, primary investigators to manage their project folders and monitor data accesses, central IT to hand out chunks of storage to departments and departmental administrators to provide project folders to primary investigators. In addition, central IT is now also able to configure the settings for their sync-and-share instance themselves.

In this presentation we will give an overview of the progress we made on our dashboard.

EFSS Products / 1

Seafile 9.0 and Beyond

Author: Jonathan Xu^{None}

Corresponding Author: xjqkilling@gmail.com

Seafile is a popular open source cloud storage solution widely used by European educational institutes, such as Homboldt University of Berlin, Marx Planck Digital Library and INRIA.

In 2021 we released Seafile 9.0. This new version contains a few important improvements to performance and interoperability. In this talk we'll present these new features and also future development plan for Seafile.

EFSS Products / 4

Nextcloud - State of the nation

Author: Frank Karlitschek^{None}

Corresponding Author: frank@nextcloud.com

This talk will give an overview of the big improvements that happened in Nextcloud in the last year. In the last 12 month Nextcloud Hub 21, 22 and 23 were made available. During this time a lot of significant improvements in functionality, performance, scalability and security were released. This talk will give an overview together with some real world example how the new capabilities can be used.

EFSS Products / 17

Infinite Scale - A new era for the ownCloud project

Authors: Jörg Eberwein¹; Patrick Maier^{None}

¹ *ownCloud GmbH*

Corresponding Authors: jeberwein@owncloud.com, pmaier@owncloud.com

With the announcement of ownCloud Infinite Scale, a new era was born for ownCloud and its community. In this talk we will explain the big picture behind the new product generation and shed light on how it will accompany and support organizations on their data strategy. Going forward

we'll talk about differences to the classic ownCloud product, celebrate the achievements since the initial Tech Preview release and discuss the roadmap to general availability and beyond.

OCM Interoperability Workshop / 45

OCM - next steps?

Author: Jakub Moscicki¹

¹ *CERN*

Corresponding Author: jakub.moscicki@cern.ch

OCM Interoperability Workshop / 10

OCM test suite

Author: Michiel de Jong^{None}

Corresponding Author: michiel@unhosted.org

Last year, we had a first version of the OCM test suite, testing three flows of OCM v1.0 between Nextcloud, ownCloud, and a stub server. These tests were running between a number of live test instances, deployed to virtual private servers for this purpose.

This year, we present:

- * the Dockerized version of these same tests
- * the addition of Reva/IOP as an OCM v1.0 implementation
- * the addition of the "invite-first" flow

OCM Interoperability Workshop / 37

ScienceMesh - Invitation Workflow Implementation

Authors: David Antoš^{None}; Jan Hornicek^{None}; Milan Danecek¹; Miroslav Bauer^{None}

¹ *Data Storage Specialist*

Corresponding Author: milan.danecek@cesnet.cz

The Invitation workflow is one of the elementary scenario on how to enable file sharing among users from different EFSS systems. Invitation workflow eliminates the necessity of knowing the exact identity of the user (share receiver) in the target EFSS system. You can generate a share invitation and distribute it via email or distribute the link via any other channel, chat app, etc. The target user then clicks on the link, selects his "home" EFSS, and log in. Once the incoming sharing is accepted the user can see all shared files in his "home" EFSS.

OCM Interoperability Workshop / 38

An OCM protocol extension to support data transfer in the mesh

Author: Antoon Prins^{None}

Corresponding Author: antoon.prins@surf.nl

In the OCM specification the share message specifies the protocol to be used for establishing the synchronization. For regular shares the webdav protocol is supported.

We present a (custom) protocol 'datatx' to support data transfer in the mesh. Using this protocol signifies that the share message is in fact a data transfer.

This presentation will be about this new OCM protocol extension for data transfer and the way we have implemented it in Reva, the reference implementation of CS3APIs.

OCM Interoperability Workshop / 35

Some thoughts about OCM-over-ScienceMesh

Author: Michiel de Jong^{None}

Corresponding Author: michiel@unhosted.org

Existing implementations of OCM generally implement the public-link workflow and the share-with workflow.

Reva implements the share-with workflow and the invitation workflow.

Should other OCM implementations add this third workflow too?

Should Reva add the public-link workflow?

Can we keep OCM-over-ScienceMesh and generic OCM-over-WWW in lock-step?

OCM Interoperability Workshop / 39

Group-owned shares

Author: Michiel de Jong^{None}

Corresponding Author: michiel@unhosted.org

OCM assumes the sender of a share is a specific user.

But in some situations it would be useful to think of shares as owned by a group. Would this be a feature we could add to OCM? What would be needed? What issues can we foresee?

See also <https://github.com/cs3org/OCM-API/issues/53>

Keynote / 42

Experiencing a new Internet Architecture

Corresponding Author: adrian.perrig@inf.ethz.ch

Imagining a new Internet architecture enables us to explore new networking concepts without the constraints imposed by the current infrastructure. What are the benefits of a multi-path inter-domain routing protocol that finds dozens of paths? What about a data plane without inter-domain forwarding tables on border routers? What secure systems can we build if a router can derive a symmetric key for any host on the Internet within nanoseconds?

In this presentation, we invite you to join us on our 12-year long expedition of creating the SCION next-generation secure Internet architecture. SCION has already been deployed at several ISPs and domains, and has been in production use since 2017. On our journey, we have found that path-aware networking and multipath communication not only provide security benefits, but also enable higher efficiency for communication, increase network capacity, and even reduce power utilization.

Born 1972, Perrig is a Swiss computer science researcher specialising in the areas of security, networking, and applied cryptography. He received his BSc degree in Computer Engineering from EPFL in 1997, MS and PhD degrees from Carnegie Mellon University in 1998 and 2001, respectively. He spent three years during his PhD working with his advisor Doug Tygar at the University of California, Berkeley. From 2002 to 2012, he was a Professor of Electrical and Computer Engineering, Engineering and Public Policy, and Computer Science (courtesy) at Carnegie Mellon University, becoming Full Professor in 2009. From 2007 to 2012, he served as the technical director for Carnegie Mellon's Cybersecurity Laboratory (CyLab). During this time he built a research project called SCI-FI (Secure Communications Infrastructure for a Future Internet). A research project aimed at building a next-generation secure internet architecture. The project later got renamed into SCION (Scalability, Control, and Isolation On Next-generation networks). Since 2013, he is Professor at ETH Zurich, leading the Network Security Group, whose research "revolves around building secure and robust network systems—with a particular focus on the design, development, and deployment of the SCION Internet architecture.

Federated Infrastructures & Clouds / 34

Update from the European Commission on Future Evolution of Digital Landscape in Europe

Corresponding Author: peter.szegedi@ec.europa.eu

Federated Infrastructures & Clouds / 41

ScienceMesh: An Interoperable Federation of EFSS services for European Open Science Cloud

Authors: Jakub Moscicki¹; Pedro Ferreira¹

¹ CERN

Corresponding Author: pedro.ferreira@cern.ch

ScienceMesh (sciencemesh.io) is an interoperable research platform developed for the European Open Science Cloud by the cs3mesh4eos.eu project.

ScienceMesh enables seamless sharing and collaboration on data between sites running different EFSS platforms (Owncloud, Seafile and Nextcloud).

In this presentation we will give a summary of the status of the integration with EFSS platforms and outlook for 2022.

Federated Infrastructures & Clouds / 5

HIFIS: VO Federation for EFSS

Authors: Andreas Klotz^{None}; Matthias Leander-Knoll¹; Sander Apweiler²

¹ *KIT*

² *FZJ*

Corresponding Author: andreas.klotz@helmholtz-berlin.de

Following the first rough ideas on Virtual Organisations (VO; Community AAI [1] based group of any size) based Enterprise File Sync&Share (EFSS) Federation [2], which were presented by HIFIS [3] on CS3 Conference 2021, we have since moved further along working on a first implementation. During Summer 2021, we have clarified the use case and identified the basic technical architecture for this future VO Federation App in Nextcloud.

When users who are distributed across multiple institutes want to collaborate within a Virtual Organisation they currently have two options: Make use of the OCM protocol [4] to share files and folders with the individual VO members who are based on remote EFSS instances. This would cause considerable effort on the sharer's side, as they need to keep track of to whom they have shared which content with. Or, as second option, all VO members have to convene on one institution's local EFSS instance, which would cause many redundant accounts and confusion on the user's side. Especially, as they need to know where to log in for working on a specific project and as they have no central entry point for all of their projects on their local EFSS instance.

We want to tackle this issue by enabling users to use federated shares with entire VOs instead of individual users. This way, every user within a VO will receive the share, no matter which EFSS instance they are based on. Updates to VO membership will also be communicated between federation members, resulting in new VO members automatically receiving existing VO shares and former VO members losing access to VO shares. Based on a new interface between EFSS and Community AAI. This whole process is planned to be GDPR compliant, too. To ensure that this interface will also work with other Community- or Infrastructure AAIs, we are collaborating with AARC to create an AARC guideline with the aim of standardizing the interface specifications.

While the initial implementation is set to be done within a Nextcloud environment, the new features will be based on existing CS3 APIs [5] and consequently be ready to also be implemented by further EFSS vendors.

[1] AARC Blueprint for Community AAIs: <https://aarc-project.eu/architecture/>

[2] CS3 Contribution: <https://indico.cern.ch/event/970232/contributions/4157924/>

[3] HIFIS Website: <https://hifis.net/>

[4] OCM Project documentation: <https://wiki.geant.org/display/OCM/Open+Cloud+Mesh>

[5] CS3 APIs GitHub page: <https://github.com/cs3org/cs3apis>; CS3 APIs are implemented in the REVA middleware: <https://reva.link/>

Federated Infrastructures & Clouds / 24

ScienceBox 2.0

Author: Enrico Bocchi¹

¹ *CERN*

Corresponding Author: enrico.bocchi@cern.ch

This contribution reports on the recent revamping of ScienceBox: The container-based stack for science with EOS, CERNBox, and SWAN services.

ScienceBox has been rebuilt from its foundations using modern cloud-native technologies for better service configuration and improved reliability, without compromising on deployment flexibility. Rethinking the whole package also allowed for better alignment of the production services at CERN with their container-based version.

Sciencebox has been tested and deployed on a variety of infrastructures, ranging from tiny deployments on developers' laptops to orchestrated Kubernetes clusters on commercial cloud providers with GPU accelerators and 100s of TBs of storage.

Federated Infrastructures & Clouds / 16**ownCloud Infinite Scale - Identity, Roles and Permissions**

Author: David Christofas^{None}

Corresponding Author: dchristofas@owncloud.com

ownCloud Infinite Scale (oCIS) will be used in many different environments. Many of those environments already have existing role definitions. To best support the individual existing definitions we designed and implemented a system which is open enough to be fitted to the environment. oCIS extensions will also benefit from that, because they can use this system for their permissions instead of implementing their own.

This talk will give an overview over the concepts, considerations, decisions we have made.

Collaboration Products / 3**Introducing smart document forms for paperwork automation**

Author: Mikhail Korotaev^{None}

Corresponding Author: galina.goduhina@onlyoffice.com

A big share of daily documents are model, universally structured files: agreements, briefs, contracts, budget plans, etc. Every process that involves such repetition has a room for automation – with this understanding in mind, ONLYOFFICE has been working on smart forms aimed at optimization of file creation and sharing in organizational document flow.

ONLYOFFICE presents new formats, DOCXF and OFORM, built on the basis of DOCX with the purpose of creating standardized document templates and working with them through specifically designed UI segment of ONLYOFFICE Docs.

This presentation will cover:

- First prototype: creating forms using Content Controls;
- Differences between smart forms and Content Control-based forms;
- OFORM and DOCXF;
- How smart forms work in ONLYOFFICE Document Editor;
- Mechanics of form sharing;
- Data protection;
- Creating and filling PDF files in ONLYOFFICE Docs;
- Roadmap for smart form development.

Collaboration Products / 6**Collabora Online: easy to deploy and manage document collaboration**

Author: Michael Meeks¹

¹ *Collabora*

Corresponding Author: mejmeeks@gmail.com

Come and hear how Collabora Online can deliver scalable, secure, on-premise editing of your documents with a simple, easy to deploy and manage architecture.

Hear about the work we've done to improve both server and client performance and scalability over the last year, with many startling improvements for users.

Hear about our User Experience improvements, from bring faster native, client-side javascript rendering to the sidebar and various dialogs, to improving document rendering crispness.

Hear some thoughts on we can allow easy deployment, simple scaling, high availability, live-upgrade-ability, and more for your EFSS, with some examples of how that is in-use around the the world.

Hear updates on improvements to features and integrations with other EFSS that have been implemented in the last year.

Collaboration Products / 8

Breaking the limits: short status update of the online spreadsheet solution SeaTable

Author: Christoph Dyllick-Brenzinger^{None}

Corresponding Author: cdb@datamate.org

In this presentation I will give an overview of the improvements that happened in SeaTable in the last year.

In the last 12 month SeaTable put focus on the development of a new archiving backend, that allow millions of records per base.

At the same time the second priority was to add more option for data visualization and automation.

SeaTable is like a lego kit that enables you to develop and build efficient business processes in the shortest possible time. SeaTable is a low code / no code platform for you and your team.

There will be a second talk during CS3 with concrete examples how to use SeaTable to visualize logs and make error handling an easy task.

CS3 Org: Governance Campfire Discussion (CS3APIs, REVA, OCM, WOPI,...) / 60

Introduction & Goals

Author: Jakub Moscicki¹

¹ CERN

Corresponding Author: jakub.moscicki@cern.ch

CS3 Org: Governance Campfire Discussion (CS3APIs, REVA, OCM, WOPI,...) / 26

CERN viewpoint

Author: Hugo Gonzalez Labrador¹

¹ CERN

Corresponding Author: hugo.gonzalez.labrador@cern.ch

REVA is an implementation of CS3 APIs which has become a key component of several systems:

- 1) CERNBox service (where it was originally developed)
- 2) Interoperability Platform (IOP) for ScienceMesh with 3rd party connectors to Owncloud, Seafiler and Nextcloud
- 3) Core module and dependency of ownCloud Infinite Scale

A roadmap and agreement on governance is needed to define the direction in which REVA will continue to evolve and to ensure that needs and perspectives of all REVA users are harmoniously reconciled:

- 1) open platform welcoming contributions from the FOSS community at large;
- 2) vendor-neutral interoperability component for European Open Science Cloud;
- 3) efficient implementation layer for commercial products supported by interested vendors;
- 4) efficient implementation layer for specific service deployments in the CS3 community.

CS3 Org: Governance Campfire Discussion (CS3APIs, REVA, OCM, WOPI,...) / 61

ScienceMesh and EOSC

Author: Pedro Ferreira¹

¹ CERN

Corresponding Author: pedro.ferreira@cern.ch

CS3 Org: Governance Campfire Discussion (CS3APIs, REVA, OCM, WOPI,...) / 19

Quo Vadis CS3 Community?

Authors: Klaas Freitag^{None}, Michael Barz^{None}

Corresponding Author: kfreitag@owncloud.com

The CS3 community and the reference implementation Reva was started a few years ago, and meanwhile the community has grown significantly. Code- and other contributions are coming in frequently. ownCloud has based its completely new product ownCloud Infinite Scale on Reva and took a significant share on Reva and other projects of the CS3 community as well.

This talk discusses how the recent developments have influenced the work on the CS3 project and how the evolving community would benefit from changes of the project.

Concretely, it raises questions and tries to propose answers in the areas of

- A (re-) definition of what CS3 and Reva want to be
- The layout of the project's code and its modules
- The release cycle and maintenance promises
- QA improvements and best practises for quality assurance
- The governance of the technical direction
- Community management and communication

The hope is that this talk will accompany a fruitful discussion of the bright future of Reva as the base of collaboration.

CS3 Org: Governance Campfire Discussion (CS3APIs, REVA, OCM, WOPI,...) / 63

CS3 ORG Governance Discussion

Keynote / 30

Digital Market: a level playing field for EU Tech sector

Corresponding Author: frank@nextcloud.com

Scalable Storage Backends / 23

Converging Storage Layers with Virtual CephFS Drives for EOS/CERNBox

Authors: Andreas Joachim Peters¹; Dan van der Ster¹; Roberto Valverde Cameselle¹

¹ CERN

Corresponding Author: roberto.valverde.cameselle@cern.ch

The CERNBox service is currently backed by 13PB of EOS storage distributed across more than 3,000 drives. EOS has proven to be a reliable and highly performing backend throughout. On the other hand, the CERN Storage Group also operates CephFS, which has been previously evaluated in combination with EOS as a potential solution for large scale physics data taking [1]. This work seeks to further explore the operational benefits of a combined EOS/CephFS solution as a CERNbox backend. First, we present the functional validation work done using a canary instance and existing micro benchmarks. Next, we show how the solution was gradually introduced to production, observing the relative impacts of metadata and backend storage on user perceived small op performance. Finally, the qualitative impact of the solution is discussed: potential for enhanced QoS (e.g. policy driven low latency vs low-cost areas), simplification of hardware operations across the entire lifecycle, and how the work may enable future cloud-based deployments.

[1] <https://doi.org/10.1007/s41781-021-00071-1>

Scalable Storage Backends / 25

Sync and Share Access to HPC Resources at CERN

Authors: Dan van der Ster¹; Theofilos Mouratidis¹

¹ CERN

Corresponding Authors: daniel.vanderster@cern.ch, theofilos.mouratidis@cern.ch

CERN Storage team has been experimenting with unified storage environments for HTC, HPC and interactive computing.

Practical examples at the prototyping and experimenting stage will be presented:

1. Easier access to user data in HPC storage (CEPHFS) via Sync/Share
2. Integration of HPC storage with the web-based analysis service environment
3. Open Source Storage backend synergy: physics (EOS) and HPC (CEPHFS)

This contribution builds upon the talk presented at HPC IODC 21:
https://hps.v4io.org/_media/events/2021/iodc21-11-kuba.pdf

Scalable Storage Backends / 33

The CERN Tape Archive : Archival Storage for Scientific Computing

Author: Michael Davis¹

¹ *CERN*

Corresponding Author: michael.davis@cern.ch

The CERN Tape Archive (CTA) is the tape back-end to EOS disk. CTA went into production in June 2020 and currently stores around 400 Petabytes of physics data. During 2022, CTA will ramp up to full production data-taking volumes with the start of Run-3 of the Large Hadron Collider (LHC). CTA is an open-source system which is being evaluated and adopted by a number of scientific institutes besides CERN.

This presentation will cover the outlook for archival storage and give an overview of how tape storage fits into CERN's integrated storage strategy and the suite of storage and data transfer products/services provided by CERN's IT Department.

User Stories / 22

JupyterLab+ScienceMesh: Collaborative Data Science in sync-and-share environment.

Author: Marcin Sieprawski¹

¹ *Software Mind*

Corresponding Author: marcin.sieprawski@softwaremind.com

Collaborative Data Science becomes increasingly important, as organizations continue to become more data-driven, and Data Science projects/models become more complex. In the report **Critical Capabilities for Data Science and Machine Learning Platforms** (March 2021) Gartner predicts, that in near future collective intelligence in Data Science and cloud-based AI infrastructure will be among key factors for competitive advantage.

This talk presents Distributed Data Science environments (part of ScienceMesh), which allow collaboration on Jupyter Notebooks in sync-and-share environment.

Jupyter Notebook has become No1 platform used by data scientists to build interactive applications and to work with big data and AI. It is widely used in CS3 institutions, many successful applications have been presented in CS3 conferences.

ScienceMesh, developed in CS3MESH4EOSC project, creates the Federated Scientific Mesh providing federated sharing of data across different sync-and-share services, federated use of applications (such as collaborative document editing, data archiving, and data publishing), fast transfer of large datasets and remote data analysis (Data Science environments).

For Data Science environments ScienceMesh delivers a JupyterLab extension, integrating JupyterLab environment with ScienceMesh. File browsing and additional share and collaboration functionalities for notebooks and resources across federated cloud are now possible in JupyterLab environment. JupyterLab is considered a complete, full-fledged IDE for Data Science tasks and interactive computing, where data scientists can do all their work in one tool, so the point is that functionalities for sharing (full cs3apis client) and concurrent editing are available inside this environment. On the other hand, Data Science environments are integrated with a comprehensive suite of Data Services in ScienceMesh, to support complete research and Data Science workflows with the use of existing

collaboration tools.

The relevance and benefits of ScienceMesh Data Science Environments will be discussed in the context of two scientific use cases (High Energy Physics and Earth Observation), along with various business-related scenarios.

User Stories / 40

iRODS Research Community Requirements Drive Expanded Scale Data Management Features

Author: Terrell Russell^{None}

Corresponding Author: tgr@renci.org

Several years ago, the entire process of data management and collaboration could only be performed with the use of proprietary software products that were expensive to license. To maintain a collection, data sites required a file system, hierarchical storage management system, and some means of sharing the data over several geographically diverse sites using purchased software, often from a single vendor to ensure compatibility. Data site managers were placed in a difficult position facing quickly growing data capacity and transmission demands with limited budgets. Constraints from funding agencies and governments became very difficult, if not impossible, to manage and audit.

The iRODS (Integrated Rule-Oriented Data System) Consortium was started as an open-source software development organization in 2013 by members of the research and storage communities. The technology has roots from an earlier project started in 1995. The Consortium was launched as a response to a major scale increase in management and storage needs driven by the advent of “big data”. The member community is now comprised of over 30 members and spans the globe from the Australia to Japan and much of the EU. Recent innovations as a result of community requirements will be discussed including graphical interfaces and methods to ensure data persistence and replication management. In addition, partnerships will be discussed with Globus and others to enable large scale collaboration. Today, worldwide, FAIR discovery and directed dissemination of HPC results are being accomplished in sites controlling tens of petabytes of data with this open-source technology.

User Stories / 14

Infinite scale is a design principle

Authors: Jörn Dreyer¹; Michael Barz^{None}

¹ *ownCloud GmbH*

Corresponding Authors: jfd@owncloud.com, mbarz@owncloud.com

When working on the spaces feature we reorganized reva’s internal path semantics. While the current global path based namespace looks efficient, it ties namespace organization to a single instance. This prevents true federation. By replacing absolute paths with relative paths and a corresponding root we can delegate building a user individual namespace to the clients. This allows them to present a more meaningful layout to the end user, even aggregating spaces from multiple instances. Furthermore, operations like quota, trash and change propagation now also operate on individual spaces.

We are moving this approach forward on the “edge” branch and will propose changes to the cs3api to optimize the implementation. We consider spaces the logical next step in enterprise file sync&share.

ScienceMesh workshop / 47

Welcome and Objectives: CS3 community and ScienceMesh

Author: Jakub Moscicki¹

¹ *CERN*

Corresponding Author: jakub.moscicki@cern.ch

ScienceMesh workshop / 48

Science Mesh – where we are now with the CS3MESH4EOSC project?

Author: Pedro Ferreira¹

¹ *CERN*

Corresponding Author: pedro.ferreira@cern.ch

ScienceMesh is an interoperable research platform developed for the EOSC, that enables seamless sharing and collaboration on data between sites running different EFSS platforms (Nextcloud, Own-Cloud and Seafile). In this presentation we will give a summary of the status of the integration with of these EFSS platforms into the Science Mesh and outlook for 2022

ScienceMesh workshop / 49

Science Mesh for EFSS service providers

Author: Ron Trompert^{None}

Corresponding Author: ron.trompert@surf.nl

ScienceMesh workshop / 11

ScienceMesh-Nextcloud

Author: Michiel de Jong^{None}

Corresponding Author: michiel@unhosted.org

As a subcontractor to the CS3MESH4EOSC project, Ponder Source developed a bridge that allows existing Nextcloud sites to join ScienceMesh. This talk will show how it works, and why you as a Nextcloud site will want to join ScienceMesh.

ScienceMesh workshop / 50

ScienceMesh-Owncloud OCIS

Corresponding Author: hugo.gonzalez.labrador@cern.ch

The status of the integration of Science Mesh with OwnCloud EFSS will be explained.

ScienceMesh workshop / 56

ScienceMesh-Seafile

Corresponding Author: maciekb@man.poznan.pl

ScienceMesh workshop / 52

Technology & Development - Advancements & Innovations

Corresponding Author: hugo.gonzalez.labrador@cern.ch

ScienceMesh workshop / 51

Researchers and Use-cases

Corresponding Author: holger.angenent@uni-muenster.de

ScienceMesh workshop / 62

Reframing adoption challenges in FAIR Data Infrastructures: Science Mesh as a source of research advantage.

Corresponding Author: gozal.ahmadova@esade.edu

This presentation explores what is the role of digital infrastructures in the FAIR movement? How can we improve the adoption of digital infrastructures by researchers? Presentation done by ESADE Business School, which leads the Science Mesh “Assessment of Business Impact” task.

ScienceMesh workshop / 58

ScienceMesh 2022: What’s next and where do we take it from here

Corresponding Author: jakub.moscicki@cern.ch

ScienceMesh workshop / 53

EOSC and Science Mesh: overcoming data challenge (EOSC Association and TFs)

Corresponding Author: ron.trompert@surf.nl

The session will be opened by a distinguished member of EOSC Association Board who will provide an overview about the EOSC Association structure, goals, as well as its next year's work plan to advance open science in Europe. Afterwards, members from EOSC Task Forces present the main priorities of their task forces and brainstorm how the work from the Task Forces and existing infrastructures and solutions developed by the CS3 Community can be brought together

ScienceMesh workshop / 54

Scientific disciplines embracing no border Research Environment thanks to Science Mesh

Corresponding Author: silvana.muscella@gmail.com

Introduction made by session chair on the importance of collaboration and joining forces between European initiatives to unlock Open Science. This session will have an esteemed panel from the different RI science clusters to discuss how the Science Mesh, by teaming up with different research infrastructures, can support them in addressing their challenges related to data sync and sharing, while increasing the long-term sustainability of their services.

ScienceMesh workshop / 55

CS3MESH4EOSC Wrap-Up and Next steps

Author: Jakub Moscicki¹

¹ CERN

Corresponding Author: jakub.moscicki@cern.ch

Decentralized Web and Storage Architectures / 12

Solid storage and pod migration [CANCELLED]

Author: Yvo Brevoort^{None}

Corresponding Author: yvo@muze.nl

Solid is quickly shaping into a solution for bringing data ownership back into the hands of the user. And yet, despite all the available storage options that are already provided, moving that data from one solution to another is not a solved problem.

The Solid project provides specifications for a different kind of web. By allowing people to store their own data in decentralized data stores (pods), it puts users back in control of which applications or people can access their data. Having a means to migrate data from one pod to another amplifies that control.

In this talk, we will touch on these subjects:

- what is Solid and why it is important
- how Solid is different compared to projects like Mastodon and Diaspora
- which storage solutions the current Solid server implementations provide and related challenges.
- what does a Solid Pod migrator solve

Decentralized Web and Storage Architectures / 7

ABEBox: end-to-end encryption for file sharing cloud services

Authors: Lorenzo Bracciale¹; Giuseppe Bianchi¹; Pierpaolo Loreti¹

¹ *University of Rome "Tor Vergata"*

Corresponding Author: lorenzo.bracciale@uniroma2.it

Besides providing data sharing, commercial cloud-based file-sharing services (e.g., Dropbox) also enforce access control, i.e. permit users to decide who can access which data.

In this work, we advocate the separation between the sharing of data and the access control function. We specifically promote an overlay approach that provides end-to-end encryption and empowers the end users with the possibility to enforce access control policies without involving the cloud provider itself. To this end, our proposal, named ABEBox, relies on Ciphertext-Policy Attribute-Based Encryption (CP-ABE) for custom policy definition and key management.

Using CP-ABE, users can encrypt and share files and folders with others without the need of handling also the sharing of the related cryptographic keys for all the resources to be shared, thus implementing a flexible many-to-many end-to-end encryption which perfectly fits the need of adding privacy to a file sharing service.

We developed a multi-platform client which seamlessly performs data encryption/decryption on top of any arbitrary cloud storage provider and takes care of the key management.

The project has been funded by the GÉANT Innovation Programme and with support from the European Commission under European Project BPR4GDPR under grant agreement No.787149.

Decentralized Web and Storage Architectures / 57

Is EOS ready for enterprise companies?

Author: Gregor Molan¹

¹ *COMTRADE D.O.O (SI)*

Corresponding Author: gregor.molan@cern.ch

On the highest level, we are presenting different views to problems and different views to solutions related to

(c) Cloud, (s_1) Storage, (s_2) Synchronization, (s_3) Sharing.

To achieve this goal, we are presenting a merge of two actors with different background and different philosophy:

Academic approach to development of a storage software to collect data from CERN experiments
Industry approach to development of a storage for enterprise companies

Actor for (*Academic*) is CERN community that developed EOS, de facto standard for collection of data for all CERN experiments, while actor for (*Industry*) is Comtrade 360 with almost 30 years of track record on storage development for enterprise customers – until 1996 Comtrade delivered 4000 engineer years of storage software.

The goal of collaboration of (*Academic-Industry*) actors is to shape the excellent EOS software to the file system for enterprise customers. To reach this goal, this collaboration must merge different worlds of (*Academic-Industry*), where there are not only different political approaches, like Linux vs Windows, and open source vs proprietary code.

Awareness that EOS is the only distributed file system that is fast enough, reliable enough, and has latency low enough to capture data from CERN experiments, urge us to (*Academic-Industry*) merge and allow EOS to be used by enterprise companies. Moreover, data storage is just as important for enterprise companies as it is at CERN in terms of importance and quantity. For this reason, Comtrade 360 marks out the road to adopt EOS to enterprise companies in terms of the complexity of setting up EOS and the complexity of using EOS.

This presentation of Comtrade 360 will highlight this road of development of

(c) Cloud, (s_1) Storage, (s_2) Synchronization, (s_3) Sharing.

from academia to industry with milestones as EOS-wnc and documentation for EOS.

User Stories / 29

CERNBox User Forum: Ultimate Engagement with Users

Author: Hugo Gonzalez Labrador¹

¹ CERN

Corresponding Author: hugo.gonzalez.labrador@cern.ch

This year the CERNBox team organised the 1st CERNBox User Forum.

This forum allowed the community to meet the CERNBox team and share their experiences and use-cases, engaging with them in this period of remote working.

In this talk we'll talk about our motivations for organising this gathering and how we deal with the vast amount of user feedback.

This talk will be of special interest for other institutions deploying EFSS platforms to their communities.

User Stories / 9

“Find the needle in the haystack”: how log monitoring, analyze and error handling can be done with SeaTable.

Author: Christoph Dyllick-Brenzinger^{None}

Corresponding Author: cdb@datamate.org

The larger log files become, the more difficult it is to keep track of them. SeaTable can be an ideal solution here, because as a database solution it has no problems to hold hundreds of thousands of rows and at the same time it offers multiple visualization options to find what you are looking for. In this presentation I will demonstrate the possibilities of log analysis with SeaTable.

User Stories / 18

ownCloud Web UI: Lessons learned from implementing accessibility

Author: Tobias Baader^{None}

Corresponding Author: tbaader@owncloud.com

Web accessibility is often understood as making websites accessible to e.g. blind people. Due to extra efforts and because it's prescribed by law, implementing accessibility measures is often not the most popular UX task.

On the other hand, accessibility advocates fiercely argue to put more effort into this - despite the presumably small target audience.

In my talk I will speak about how to overcome both views with a pragmatic "how to start with accessibility" approach and show you how this improves the everyday usage of web applications for nearly anyone.

User Stories / 36

Using Workflows for Data Preservation Using Onedata

Authors: Michal Orzechowski¹; Bartosz Kryza²; Lukasz Dutka²

¹ ACC Cyfronet-AGH

² ACC Cyfronet AGH

Corresponding Author: lukasz.dutka@cyfronet.pl

Onedata [1] is a distributed, global, high-performance data management system, which provides transparent and unified access to globally distributed storage resources and supports a wide range of use cases from personal data management to data-intensive scientific computations. Due to its fully distributed architecture, Onedata allows for creation of complex hybrid-cloud infrastructure deployments, with private and commercial cloud resources. It allows users to share, collaborate and publish data as well as perform high performance computations on distributed data. Onedata allows users to collaborate, share, and perform computations on data using applications relying on POSIX compliant data access.

Onedata comprises the following services: Onezone - authorisation and distributed metadata management component that provides access to Onedata ecosystem; Oneprovider - provides actual data to the users and exposes storage systems to Onedata and Oneclient - which allows transparent POSIX-compatible data access on user nodes. Oneprovider instances can be deployed, as a single node or an HPC cluster, on top of highperformance parallel storage solutions with the ability to serve petabytes of data with GB/s throughput.

Recently, Onedata was enhanced with a powerful workflow execution engine, powered by OpenFaas [2]. It allows for creation of complex data processing pipelines that can leverage the transparent access to distributed data provisioned by Onedata. In particular the workflow functionality can be used to create a comprehensive, OAIS [3] compliant, data archiving and preservation system, covering all archival requirements including ingestion, validation, curation, storage and publication. The workflow function library contains ready to use functionalities (implemented as Docker images), covering typical archiving actions such as metadata extraction, format conversion, checksum validation, virus checks and others. New custom functions can be easily added and shared among user groups. The solution was thoroughly tested running on auto-scalable Kubernetes clusters.

Currently Onedata is used in European EGI-ACE [4], PRACE-6IP [5], and FINDR [6] project, where it provides data transparency layer for computation, data processing automation deployed on dynamically hybrid clouds containerised environments.

REFERENCES:

- [1] Onedata project website. <https://onedata.org>.
- [2] OpenFaaS - Serverless Functions Made Simple. <https://www.openfaas.com/>.
- [3] David Giarretta, CCSDS Group, and CCSDS Panel. Reference model for an Open Archival Information System (OAIS). 06 2012.
- [4] EGI-ACE: Advanced Computing for EOSC. <https://www.egi.eu/projects/egi-ace/>.
- [5] Partnership for Advanced Computing in Europe - Sixth Implementation Phase. <http://www.prace-ri.eu>.
- [6] FINDR: Fast and Intuitive Data Retrieval for Earth Observation

Technology for Application Integration / 21**ownCloud WOPI Proxy for O365****Author:** Mark Carioscio^{None}**Corresponding Author:** mcarioscio@owncloud.com

Content collaboration is an essential component of modern companies. Using WOPI protocols, companies can allow employees to directly edit their cloud files in the web browser.

Microsoft has allowed the editing of Word, Excel, and PowerPoint files within a web browser for several years, but the company using this feature was initially required to host their own Microsoft Office Online Server (OOS).

For companies that already have an Office subscription and do not wish to host their own OOS, ownCloud (oC) has proudly joined Microsoft's Office Cloud Storage Partner Program (CSPP). Now, via a proxy server set up by oC, customers will soon be able to view/edit documents online via the oC Web UI.

Technology for Application Integration / 2**Finding an optimal approach to enabling document collaboration with ONLYOFFICE Docs in integrated environments: interoperability, decentralization and limitations****Authors:** Alex Miheev^{None}; Mikhail Korotaev^{None}**Corresponding Author:** mikhail.korotaev@onlyoffice.com

Intentions to increase interoperability and expand functionality often clash in a conflict where a more universal standardized approach to service integration limits the unique capabilities of the developer's technologies.

Despite WOPI having its limits in delivering the whole amount of functionality into the receiver system's interface, it is an open standard that makes integration considerably easier due to abundant and standardized documentation, ready ways to perform connectivity checks, and ability to integrate the services into protected systems where API integration is simply not possible.

With ongoing research in WOPI-based ONLYOFFICE integration, we don't consider API and WOPI interchangeable alternatives, as our default API integration provides opportunities to accommodate the growing functionality that in many cases goes off-limits.

In this presentation, we will discuss:

- Two approaches of integrating ONLYOFFICE Docs into sync&share environments: API and WOPI;
- Limitations of WOPI and ways to overcome or adapt to them;
- ONLYOFFICE Docs integration using WOPI: ownCloud Infinite Scale, SharePoint, OpenKM and Filecloud;

- WOPI integration structure and what it means to third-party ONLYOFFICE integrators;
- Recent updates in functionality of ONLYOFFICE Docs available for integrated solutions;
- Roadmap for future development and integrations.

Technology for Application Integration / 15

integrate applications with the application provider

Author: Willy Kloucek^{None}

Corresponding Author: wkloucek@owncloud.com

The CS3 apis are all about files. Large parts of it focus on how to make files accessible to users. But once users have access to a file, they want to do something with them. Therefore external applications bring a real value to an EFSS solution.

The solution on how applications can integrate themselves in the CS3 ecosystem is called App Provider ("cs3.app.provider").

Since the last CS3 conference, the possibilities for external applications have been improved or added in the CS3 apis, REVA and ownCloud Web.

This progress on the App Provider and how you can use it today will be demoed in this session.

Summary and Conclusions / 64

Summary and Conclusions

Corresponding Author: jakub.moscicki@cern.ch

CS3APIs Interoperability Campfire Discussion / 27

CS3APIs: governance and next steps (Campfire Discussion)

Authors: Hugo Gonzalez Labrador¹; Jakub Moscicki¹

¹ CERN

Corresponding Author: jakub.moscicki@cern.ch

CS3APIs have been initially developed within the CERNBox service to flexibly connect Storage and Applications Providers. CS3APIs adopt the industry standards for backend microservice architectures and consequently have been used to interlink a growing number of independently developed components, for research and commercial services.

A governance model needs to be discussed and agreed by the community to ensure stable and predictable evolution of the APIs.

ScienceMesh workshop / 46

ScienceMesh workshop