

CS3 2024 - Cloud Storage Synchronization and Sharing



Monday, 11 March 2024 - Wednesday, 13 March 2024

CERN

Programme

The 2024 CS3 Conference will take place on 11-13 March 2024 at CERN. This is a way for all of us to mark the 10th anniversary of this event and of our community (the very first conference in this series took place in 2014 at CERN).

We look forward to meeting old and new colleagues within the CS3 community. Reconnect, inspire and get inspired, learn from each other and have some fun together, too.

The final agenda will be compiled once we collect all the contributions.

This is the 10th anniversary edition and it part of an larger series of events: **CERN's TechWeek on Storage and Data Technologies**.

Keynotes

From Data To Knowledge: Computing For High-Energy Physics Experiments

Mario Lassnig (CERN)

Modern large-scale sciences have become increasingly complex and face unprecedented data & compute challenges. The number of data-intensive instruments generating substantial volumes of data is growing and their accompanying internal and external workflows are becoming more complex every day. Their storage and computing resources are usually heterogeneous and are distributed at numerous geographical locations belonging to different administrative domains and organisations.

In this presentation, we will try to give an insight how we solved these problems in High Energy Physics: from data taking, distributed processing, infrastructure, and user analyses, with specific examples from the ATLAS Experiment.

Dr. Mario Lassnig has been working as a Software Engineer at the European Organisation for Nuclear Research (CERN) since 2006. Within the ATLAS Experiment he is in charge of all aspects of its distributed computing systems. In his previous life, he developed mobile navigation systems for transportation at the Austrian Institute of Technology and on-board satellite data systems at the Austrian Space Forum. He holds a Master's degree in Computer Science from the University of Klagenfurt, and a doctoral degree in Computer Science from the University of Innsbruck

Neutrality, Impartiality, and Independence of Humanitarian Action in the Digital Age at ICRC

Presented by: ICRC Delegation for Cyberspace, Luxembourg

Challenges and opportunities for Neutral Impartial and Independent Humanitarian Action in a world increasingly characterized by pervasive connectivity and connectivity denials, digital services, adverse cyber operations, debates around digital sovereignty, and global competition for digital infrastructure.

European Strategy for Data and the Resulting Landscape

Presented by: DG-CNECT, European Commission

Topics of interest to the CS3 community:

Bird's eye view on structure, objectives and scope of Europe's Strategy for Data. The resulting the landscape of EC-supported incentives and projects is based on several pillars:

Scientific and research community: research and innovation actions (RIA), EOSC
Industry and public services: data spaces (DEP — Digital Europe Programme) for long-term standard operations
Flagship initiatives with high-societal impact: Destination Earth, AI4Europe,...

How this translates into services? New platforms are being launched in 2024, such as:
AI-ON-DEMAND
Copernicus Data Space
EOSC EU Node

Are these services supposed to be interoperable? What would be the role of SIMPL protocols?

What is the long term economic model and market incentives to ensure successful implementation of single market for data; B2G, G2B, B2B data sharing. How shall this happen?

What is the long-term strategy to ensure European competitiveness vis-a-vis established or emerging technology superpowers in other parts of the world? How can we be leveraging on the European assets and expertise to achieve this goal. Is European ICT infrastructure / technology / expertise also an asset (in addition to domain specific assets such as Weather Models)?

Panel discussion: EOSC — what's in there for the CS3 community?

The first results of EOSC funding programme becoming available and the question is: how is this relevant and will this be relevant for the CS3 community — as a whole and as individual actors. What's in there also for the actual users (research groups, academic, ...), with (or without) access to their home institute services.

Some of the existing results are:

EOSC EU Node — Exchange Infrastructure and Application Services (Lot 2 and 3), including managed services for end-users: File Synchronisation and Sharing, Interactive Notebooks, Large File Transfer

ScienceMesh community federation: ScienceMesh is part of EOSC Exchange, how do ScienceMesh and EOSC node services relate and mutually leverage for the common community benefit and to support the European Strategy for Data?

Core Federation Services for EOSC EU Node (Lot1), including Helpdesk, SSO, Accounting: can this be leveraged or provide services for our community and ScienceMesh?

Interoperability protocols and platforms within RIA (EOSC services) and interoperability with DEP (Industrial Data Spaces)?

Main sessions

User Voice: Innovative Applications, Data Science Environments & Open Data

Enterprise File Sync and Share (EFSS) services play a role in the research data lifecycle workflow, from data acquisition, collection, processing, analysis to publication, preservation and Open Data.

This track is intended to discuss integration between EFSS and other research services to cover the full data lifecycle: interactive notebooks, metadata aggregators, data repositories, preservation services, specialised data viewers, collaboration tools, documentation and more.

How can our services help in implementing and encouraging FAIR practices in real life? What should be the role of human interfaces and machine APIs?

This track is also a place to discuss novel applications and user scenarios which are enabled by the CS3 services with innovative data access and sharing functionality.

Keywords: JupyterLab & Notebooks, FAIR, ORCID, OpenAIRE, GPUs, Spark, Analytics, DTN, FTS, Grid.

File Sync & Share Solutions and Requirement from the Community

This is the track for software companies developing Enterprise File Sync and Share products (EFSS): evolution and latest releases, planned new features and development roadmap.

Past speakers included: Dropbox, Nextcloud, Owncloud, Powerfolder, Pydio, Seafile, Syncany

We expect to discuss community requirements in connection with the different solutions, such as scalability, technology evolution, interoperability, operational costs and the like.

We invite contributions from software companies and users.

Scalable Storage Backends and Integration with Data Processing

Effective storage solutions are a key factor for successful integration of CS3 services with processing farms for HPC, Machine Learning and Global Science.

In this track we provide a forum for providers and integrators of innovative storage solutions across all storage tiers: low-latency storage for data analysis and machine learning, online and near-line storage for data products and archival storage.

Solutions from vendors and experience from the sites will be discussed.

Collaborative Applications, Data Privacy and Data Classification

This track focuses on collaborative applications and techniques to enhance sharing and user experience for Office, Groupware and Productivity.

For this class of applications data privacy is a challenge and data classification is becoming a topic of interest.

While applications are typically provisioned fully on-premise some CS3 sites are exploiting hybrid models in which certain applications (e.g. office) are provisioned by external clouds while the storage stays on-premise.

We encourage input from application developers as well as site managers and policy owners.

CS3 federations and synergies with eResearch infrastructures.

There is a large overlap between CS3 sites and other eResearch Infrastructures (such as WLCG, EGI, SKAO, EU-DAT...) and services resulting from science cluster projects (such as ESCAPE, SSHOC, EOSC-Life,...).

This is a session to discuss co-evolution of CS3, Enterprise File Sync and Share (EFSS) services, EFSS federations and eResearch Infrastructures. Contributions are encouraged from all stakeholders.

CS3 Future in the European Open Science Cloud (EOSC)

European Open Science Cloud (EOSC) is the European Commission's Digital Single Market strategic initiative to remove virtual borders, boost digital connectivity, and make it easier for researchers, innovators, companies and citizens to access cross-border online content across the European Union.

In this session we will discuss the EOSC roadmap, available services and opportunities, as well as the role of EFSS platforms and federations such as ScienceMesh in EOSC.

Interoperability: protocols, APIs, OpenCloudMesh (OCM)

The community is involved in ongoing activities in the area of federation and interoperability, for example technologies such as OCM, CS3APIs and Reva.

In this track we welcome contributions regarding the status of developments and experience with prototypes and services.

Technology & Research

The traditional CS3 track hosts technical presentations of the building blocks of CS3 services, as:

- * Algorithms and protocols for file sync and sharing;
- * Sharing and metadata semantics;
- * Service reliability and data integrity;
- * Innovative desktop and mobile integration;
- * Monitoring and performance analysis;
- * New user interfaces;
- * APIs and command-line tools.

CS3 Community Site Reports

There is a growing number of services deployed and operated in the CS3 community. This session is an opportunity to share the current status and plans, user feedback as well as operational experience. This session will provide a sort of family-photograph and a competence map of all CS3 services.

In particular we encourage newcomers to the community to introduce themselves and to present the status of their system and plans.

ScienceMesh Workshop: status and plans

ScienceMesh (sciencemesh.io) is a global collaboration service for researchers, educators, data curators and analysts based on OCM (Open Cloud Mesh) and CS3APIs (Connect Storage and Application Providers).

We are planning to have a workshop involving project participants and users.

SIG-CISS meeting of GEANT Association

Co-located event: 13th SIG-CISS meeting

The SIG brings together those who are building/operating R&E clouds and are willing to share strategy, design, deployment, performance optimization, application integration, interoperability, security and other related information, knowledge and best practices as well as participate in joint efforts aimed at addressing needs of academic environment related to building, operating, brokering cloud services and infrastructure and ensuring their interoperability.

More information:

<https://wiki.geant.org/display/CISS/SIG-CISS+Home>

Interoperability: protocols, APIs, OpenCloudMesh (OCM)