

CS3 2020 - Cloud Storage Synchronization and Sharing Services



Monday, January 27, 2020 - Wednesday, January 29, 2020

Programme

The 6th edition of the CS3 conference (Cloud Storage Services for File Synchronization and Sharing) will take place in Copenhagen (January 27-29 2020).

A lot happened since the last meeting in Rome. Most importantly, CS3 participants prepared the CS3MESH project proposal which was approved as a 3-year EU project starting January 2020.

This edition of the annual CS3 conference is immediately preceding the first CS3MESH project meeting. This was a deliberate choice: we believe that CS3MESH is a tool and an opportunity for all the actors of the CS3 community. The project aims to deliver the initial and coherent implementation of the long-term CS3 collaborative infrastructure. This is an important opportunity to give input and help co-design the common future.

We believe that original formula focusing on innovative storage systems and their integration with user environments remains the key to enable progress in data sciences at all levels: local laboratory, regional collaborations and global science. Therefore CS3 will continue to follow applications ranging from innovative big-data analysis to science outreach and education promoting exchange of experience, innovative ideas and collaboration across all actors in this industry. In short: CS3 will continue to drive the evolution in the field via an inclusive community effort and CS3MESH project will help to structure this effort.

The sessions we propose for this conference are listed below and we invite you to submit contributions (oral presentations or posters) via the conference web site.

Keynote: May we please store your personal data?

Michiel de Jong will talk about his work on Terms of Service; Didn't Read, Unhosted web apps, and more recently, Solid.

Invited talk: Machine learning in particle physics, Troels Petersen, Niels Bohr Institute

Experimental particle physics is notable for producing large amounts of data. The ATLAS-detector at the CERN Large Hadron Collider is truly exceptional in this respect: The amount of data produced is still many orders of magnitude larger than what can meaningfully be consumed with today's data processing mechanisms. For this reason ATLAS stores only 1 out of every 100000 collision events recorded, and afterwards it is the job of a complex data reduction and analysis chain to further reduce data without losing events of scientific interest. With the advent of large-scale machine-learning technologies this chain has been considerably enriched and has allowed an expansion of the size and amount of large datasets with detailed information that can be explored. Putting this into practice requires the testing of many training configurations. This, in turn, puts strains on storage and computing power. I will show, from the point of view of particle physics, how critical infrastructure is to obtaining scientific results.

Meet CS3MESH

This track will be animated by a general overview of the CS3MESH programme putting in evidence collaboration opportunities for the CS3 community at large. Flexible CS3MESH federation across

installations will promote global scientific collaboration and integration, avoiding disconnection of infrastructures. The main strength of CS3MESH is the CS3 community and the project will be listening to all proposals!

The spirit of the CS3 conference has been the main driver in preparing the CS3MESH proposal and one of the main factors in getting it approved. This track is the initial step to open and keep the effective communication channel between the project and the whole community.

We expect the project to deliver the initial implementation of the CS3 infrastructure, hosting all sites interested to use it and to contribute to its evolution. In this session we will encourage input on specific use cases: requirements ("my users desperately want this"), success stories ("this is the way we operate our system and we believe it might be of interest of others") or other experience ("it looked a good idea but eventually we abandoned it"). So, if you want something delivered from CS3MESH, or you are prepared to co-design it or you are offering experience or resources, this is your session. We expect that some contributions might be called to give more technical details in one of the following tracks while concentrating here on the high-level messages.

The background will also be introduced in this session. In essence: the European Science Cloud is in the full swing, digital sovereignty became one of the critical topics and literally tens of organization are looking for opportunities to offer services based on open-source software, on-premise data storage and collaborative work environments cross-institutionally. This opens up new opportunities for the CS3 community.

The CS3MESH project will jump-start the entire CS3 community to become a permanent and integral part of the newly shaping landscape of European cloud computing for research and education.

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

This track is for novel applications and user scenarios which are enabled by the CS3 services with innovative data access and sharing functionality.

Many CS3 institutes are experimenting with new ways to support data science on their collaborative storage fabric. Activities such as quick-prototyping, educational and outreach tools have been quite successful.

One such example is the usage of interactive notebooks which enable collaborative data processing. Notebooks naturally become environments for data curation, data preservation, educational and outreach. The ease of access and the self-documenting feature of notebook-based environments complement and cooperate with sync and share environment.

Likewise, examples of successful production-grade data analytics environments are also available. Analysis platforms have the potential to become the aggregation point for other services, notably specialised data viewers, collaboration tools, documentation and more.

More recently direction has been emerging where CS3 services may become the fabric to implement new classes of services focusing on open-data access and data preservation.

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

Collaborative Platforms

This track focuses on collaborative platforms and techniques to enhance sharing at the application level (Office, Groupware and Productivity). As a matter of fact more and more web-based tools are becoming available and become accessible as web-based applications within Sync&Share platforms. CS3 sites are proposing ways to host such services in a coherent way augmenting their final value, e.g. via combining Office functionality and sharing capabilities.

File Sync&Share Products for Home, Lab and Enterprise

This is the presentation session for software companies developing File Sync&Share products: evolution and latest releases, planned new features and development roadmap.

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

Scalable Storage Backends for Cloud, HPC and Global Science

This storage track is the place for providers, advanced users and integrators of innovative storage solutions. The need of selecting and supporting effective storage solutions (notably in the multi-PB area) should not overshadow the difficulty and costs to maintain these solutions without creating long-term support nightmares. Nowadays cloud storage is required to deliver multiple functionalities within a single data repository, e.g. serving sync&share mobile access along with high-performance HPC access. Solutions from vendors and experience from the sites will be discussed in this track.

CS3 Community Site Reports

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

Technology & Research

Classic CS3 track presenting and discussing technical building blocks of CS3 services: technology, design, experimentation and engineering results. It includes topic like:

Interoperability: CS3 APIs, OCM
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.