

# CS3 2019 - Cloud Storage Synchronization and Sharing Services



**Monday, January 28, 2019 - Wednesday, January 30, 2019**

**CNR**

## **This conference**

The Cloud Storage Services for File Synchronization and Sharing (CS3) in about innovative storage systems and their integration with user environments to enable progress in data sciences at all levels: local laboratory, regional collaborations and global science. CS3 applications range from innovative big-data analysis to science outreach and education.

In addition to the well-established format we also added new tracks this year to include new areas of interest which emerged lately:

Open Data Ecosystems and CS3

Cloud infrastructure and software stacks for data science (CISS)

A number of displays will be available for **digital posters**. They have to be in JPG format, 1080x1830 (width\*height) pixels in size. Posters will be shown at the conference premises and uploaded on the conference site, and can be posted here: <https://pandora.infn.it/public/09ec1e> before January 18, 2019.

## Keynote

**Open Deep Learning and Data Management of large datasets in hybrid Clouds: a practical view** by Davide Salomoni, Istituto Nazionale di Fisica Nucleare.

Davide is now Director of Technology at the INFN in Bologna (CNAF). It is an excellent news that Davide accepted our invitation given his experience in large-scale distributed computing (Grid, Cloud) within large international multiscale project.

Davide will let us explore how an open "Deep Learning as a Service" paradigm can be applied to public or private cloud infrastructures. The keynote will cover topics such as dynamic, on-demand orchestration, data placement and instantiation of DL environments for the efficient analysis of very large datasets over distributed clouds, real-time handling of ingested data, and publication of reusable, trained datasets into open catalogues - with emphases on concrete cases and experience.

## Invited Talk

**Building an EOSC in practice** by Isabel Campos Plasencia, Spanish National Research Council - CSIC

Isabel will give us a complete overview of the European Open Science Cloud (EOSC) and discuss the main challenges and opportunities.

## User Voice: Novel Applications

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

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.

Analysis platforms have the potential to become the aggregation point for other services, notably specialised data viewers, collaboration tools, documentation and more.

## 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, motivated by several scenarios described below.

High-performance and cost-effective storage solutions are important to scale up and evolve data and synchronization services in the context of Cloud, HPC and global scientific environments.

The separation between the storage backend used by sync&share services and analytics environments brings no user benefit: it prevents the users to easily share algorithms and results; it also complicates data correlation and full-statistics access; ultimately hardware resources are not optimally used and managed.

Seamless integration of storage into sync&share environment may facilitate egress and ingress of data to specialized systems such as HPC.

Modern scalable storage backends should easily support many thousands of concurrent clients and have multi-PB storage capacity. To allow federating distinct storage resources, multi-site capabilities are quite important; cache capabilities to improve user experience and system resilience are also interesting.

## Synchronization/Sharing Technology & Research

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

Main topics:

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

...

## 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: Nextcloud, Owncloud, Powerfolder, Pydio, Seafiler, Syncany

## Open Data Ecosystems and CS3

This is a new discussion track on the future open data ecosystems and CS3 services, given the evolutionary path described below.

It has taken five years for CS3 type services to mature from the GB-range niche sync&share platforms for early adopters they were initially, into what they currently are for many R&E einfra providers: their default collaboration and live-data holding platform often in the PB range.

In those same 5 years, other aspects of the digital Open Data / Open Science landscape were abuzz as well; open publishing, data packaging and portability, identity minting, data citation; and around these topics and principles, too, organisations have sprung up, such as RDA, ORCID, DataCite, GO-FAIR, the Open Science Foundation etc.

Up until recently these developments happened largely in isolation from each other. At the current state of maturity and community uptake however, it seems the next logical step is to combine these various systems and services and investigate what progress can be made in open science if a joint-up, coordinated system is presented to researchers.

This session invites submissions on implementations of and experiences with such joint-up open science / open data systems, where a CS3 type service acts as the live data fulcrum.

## Cloud infrastructure and software stacks for data science

This new track explores CS3 services in a broader context of modern cloud infrastructure and software stacks (CISS).

Integration between CS3 and CISS aims at providing uniform environments shared across different researchers; access to computing facilities or workflow engines like batch facilities, Openstack&container services, Spark clusters, Cloud-based resources, GPU hardware and more.

Keywords:

New Research Environments

Data Management and Workflows

File Transfer & Distribution

Virtualization: Open Stack, Open Nebula

Containers and Orchestration: Kubernetes, Mesos

Analytics: Hadoop, Spark Compute and Grid services

## Sharing and Collaborative Platforms

This track focuses on collaborative platforms and techniques to enhance sharing at the application level (Office, Groupware & Scientific Apps) as well as between cloud infrastructures (Open Cloud Mesh federating standard).

The case for deeper integration of data-intensive cloud services with desktop-like services has clearly emerged.

The need of flexible federations across installation is also an absolute need to promote global scientific collaboration and integration and to avoid fragmentation of infrastructures.

## 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.