# CS3 2018 - Workshop on Cloud Storage Synchronization and Sharing Services



**Monday, January 29, 2018 - Wednesday, January 31, 2018**

**AGH Computer Science Building D-17**

# Scientific Program

The event consists of 7 sessions in a single-track: 5 regular oral sessions + 2 special-format sessions. In between sessions there will be plenty of time for discussions and networking.

In this edition we added an exciting new session on **Cloud Infrastructure&Software Stacks for Data Science.**

For all regular sessions oral presentations are accepted. The number of slots is limited and not guaranteed.

# User Voice: Novel Applications

*Format: prime time presentations, 25 minutes + 10 minutes QA.*

There is an increasing number of scientific-, engineering-, collaborative- and office applications closely integrated with CS3 services (cloud storage and file sync/share services). This session is designed for (power-)users of such novel applications to share their experience with synchronized-, online- and offline-storage: benefits and areas for improvement of CS3 services in their respective application domains. This session is an opportunity for service and technology providers to understand opportunities and new requirements but also shortcomings of existing CS3 implementations from the most important perspective: the one of the user.

Application domain examples:
Image collection, processing and analysis Educational applications
Scientific notebooks for data processing
Collaboration and communication in team
Decision support systems
Problem solving environments
Data curation and publication workflows
...

# Scalable Storage Backends for Cloud and HPC: Foundations

*Format: oral presentations, 20 minutes + 5 minutes QA*

High-performance and cost-effective storage solutions are important to scale up and evolve synchronization services.

The separation between the storage backend used for offering sync&share services and the ones for analytics is usually not desirable. This separation 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.

This track focuses on the lower layer of the stack: storage foundations.

In the storage track we call for contributions from innovative storage providers. Interesting storage systems should promote seamless integration with synchronization infrastructures. They should scale above many thousands of 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

# Scalable Storage Backends for Cloud and HPC: Integration

*Format: oral presentations, 20 minutes + 5 minutes QA*

High-performance and cost-effective storage solutions are important to scale up and evolve synchronization services.

This track focuses on the upper layer of the stack: efficient integration of storage into sync&share environment.

# Synchronization/Sharing Technology & Research

*Format: oral presentations, 20 minutes + 5 minutes QA*

This is a classic CS3 session to present technology, design, experimentation and research results relevant for development and operation of synchronization and sharing services. The topics include:
algorithms and protocols for file sync and sharing
sharing and metadata semantics
data integrity
reliability
conflict resolution
network-aware design
software QA hardening and resilience
desktop and mobile integration
backup strategies
monitoring tools
testing frameworks
security frameworks and procedures
new user interfaces
...

# File Sync&Share Products for Home, Lab and Enterprise

*Format: oral presentations, 20 minutes + 5 minutes QA*

In this session software companies present their File Sync&Share products: latest releases,

planned new feature and development roadmap.

Past speakers included:

Nextcloud, Owncloud, Powerfolder, Pydio, Seafile, Syncany

# Cloud Infrastructure&Software Stacks for Data Science

*Format: oral presentations, 20 minutes + 5 minutes QA*

Classical networked storage systems typically accepted science data in bulk uploads, often after processing; as a consequence, stored data usually wasn't live in the sense of fresh from the instrument. Similarly, efforts at building Virtual Research Environments (VREs; essentially cloud-based science toolchains) haven't seen great uptake, again because the tools are only useful if they have fresh data to operate on, and users typically do not good discipline at regular uploading of data taking runs.

In contrast, synched data stores hold what can be considered live data – thereby offering the possibility of performing first-line scientific munging / workflow / analytics on the cloud platform, rather than on researchers' desktops. This opens up interesting possibilities of transparent compute scaling, GPU compute, science package management etc. not normally available on researcher-managed (desktop) platforms. This stream is intended to showcase such novel opportunities.

Keywords:
Virtual 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 & Scientific Apps) as well as between cloud infrastructures (Open Cloud Mesh).

# 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 family photograph or a map of all CS3 services in operation to give a clear and concise picture of the whole service community.

*Format: A summary of all site reports will be presented by a session convener in one single presentation. There will be few selected lightning talks on service highlights. This will be followed by the plenary discussion involving all participants. Immediately after the discussions will continue around the posters in the lobby. The posters will be on display for the whole time of the event.*

How you can contribute:

You are asked to provide basic information about your service according to this template By entering providing the information you will enter into the official CS3 contributor list.

You are entitled to prepare a poster which will be displayed at the venue.

Optionally, if you believe there is a particular highlight of your service, you may consider a 5 minutes lightning talk + 1 minute QA slot. The number of such slots is limited and not guaranteed.

# Future of Sync&Share (Panel Discussion)

Format: oral presentation, 10' minutes + panel

# Welcome