Cloud Services for Synchronisation and Sharing (CS3) Conference Report

Oliver Keeble

Inc material from Javier Ferrer, Elizaveta Ragozina, Gianmaria Del Monte
CS3 2023

- Cloud Services for Synchronisation and Sharing conference @ESADE, Barcelona
- From 06 March 2023 to 08 March 2023
- Attendance: ~150 people
- https://indico.cern.ch/event/1210538/program
Plan

• Overview of the community and conference
• Where CERNBox fits in
• The Mesh
## Timetable

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collaborative Data Science and Visualisation</strong></td>
<td><strong>EFSS Products</strong></td>
<td><strong>Site Reports</strong></td>
</tr>
<tr>
<td><strong>EFSS Products</strong></td>
<td><strong>Interoperable Cloud Infrastructure Stacks</strong></td>
<td><strong>EFSS Products</strong></td>
</tr>
<tr>
<td><strong>Site Reports</strong></td>
<td><strong>Education and Research Use-cases</strong></td>
<td><strong>Interoperable Cloud Infrastructure Stacks</strong></td>
</tr>
<tr>
<td><strong>Interoperable Cloud Infrastructure Stacks</strong></td>
<td><strong>FAIR Data Management</strong></td>
<td><strong>Federated Infrastructures &amp; Clouds</strong></td>
</tr>
<tr>
<td><strong>Education and Research Use-cases</strong></td>
<td><strong>Federated Infrastructures &amp; Clouds</strong></td>
<td><strong>Sciencemesh workshop</strong></td>
</tr>
<tr>
<td><strong>FAIR Data Management</strong></td>
<td><strong>Collaboration Products</strong></td>
<td><strong>Security and Authentication</strong></td>
</tr>
<tr>
<td><strong>Federated Infrastructures &amp; Clouds</strong></td>
<td><strong>Security and Authentication</strong></td>
<td><strong>Scalable Storage Backends</strong></td>
</tr>
<tr>
<td><strong>Sciencemesh workshop</strong></td>
<td><strong>Scalable Storage Backends</strong></td>
<td></td>
</tr>
</tbody>
</table>
Keynote: Stop data sharing

- Interesting observations on FAIR data from one of the founders of the concept
  - From “Findable, Accessible, Interoperable & Reusable”
  - To “Findable and AI Ready”
  - Stop copying and start visiting data
Enterprise File Sync ‘n’ Share (EFSS) providers
Nextcloud. State of nation
Frank Karlitschek

An overview of Nextcloud Monitoring
Pietro Marini

Seafile, what’s new
Jonathan Xu

Status Update of the no-code platform SeaTable

Visualize big data and create individual customer frontends
Christoph Dyllick

OwnCloud Infinite Scale
Holger Dyroff

A View on the CS3 APIs
Klaas Freitag

New scenarios for Research and Higher education
Reinhard Schüller
Performance improvements

Integrations

Open Collaboration Services
Develop apps and integrations in any language and connect to a content collaboration platform
Encryption
- Server-side: S3 primary storage, group folders
- End-to-end: performance, control access, shared folders

Collabora deep integration

Advanced versioning: name, preview, mark to keep/delete

Automatic file locking for Office apps (Text, Collabora online), showing lock status for users, manual locking/unlocking
File server rewritten in Go, **performance improvements**

- Notification server (websockets) on file upload, lock, change of permissions
- MacOS sync client implements VFS (fetch on demand)
- OCM (Mesh) integration with Nextcloud
- Release of oCIS (ownCloud Infinite Scale)
- New architecture: Microservices based, Infinite scalability, parallel and async processing
- Full text search
- Deny access for folders
- Full practice for Kubernetes deployment
- Integrations MS Office, OnlyOffice, Collabora
Spaces: a "revolutionary new" way of collaboration, Shared maintainership, data ownership decoupled from uids --> use case in higher education (Bavarian schools)

"Databaseless"

LibreGraph API (inspired by MS Graph API for accessing their cloud resources)

Currently mixed unstructured use of LibreGraph, CS3, other APIs

Performance improvements
Nextcloud clearly user-oriented, feature rich with tons of integrations

Seafile: focus on performance + SeaTable

OwnCloud: focus on scaleability and "spaces"
Analysis/Apps
Office and workflow

Collabora Online
A real alternative office suite
- Smoothly transition to something better
- Collaborate on documents in the browser
- Excellent interoperability with Microsoft formats
- Easy, on-premise deployment & scaling
- Document + Spreadsheet + Slides + Drawing
- On-premise integration

Regain control of your data
Over 80m pulls from docker image

Collabora Online

About ONLYOFFICE project

Products
ONLYOFFICE Docs
ONLYOFFICE Workspace

Solutions for
SaaS, On-premises, Desktop, Mobile

Supported platforms
Windows, Linux, macOS, Android, iOS

License
Open source / commercial

Status Update of the no-code platform SeaTable
CS3 2023
March 7th, 2023

from the developers of SeaTable
Integrating (CERN) services

Why an analysis facility?
And why SWAN needs to evolve
Voilà/VOIS – Turn notebooks into web apps

Consistency: themes

VOIS library manages the light and dark theme for all the widgets.
With a single setting a dashboard theme can be changed.

Layered widgets: dialog-boxes

Modal dialog-boxes can be opened to request user input.
Any widget can be added to the dialog content.
On the right an example of some overlapping dialog-boxes that implement a legend editor for geospatial vector dataset thematicization.
Infrastructures
INFN Cloud

- INFN Cloud
- Sync’n’share as a service, self managed

What is INFN Cloud

A production-quality set of resources and solutions providing:
- A core backbone, with ancillary and special-purpose services.
- A multi-site, federated Cloud infrastructure:
  - INFN Cloud can transparently federate INFN sites as well as public or private Clouds (e.g., AWS, Google Compute Cloud, Microsoft Azure, and others)
- A customizable portfolio of services accessible via web interfaces, terminal or API.
- A fully distributed organization for the support and management of both infrastructure and services.
- A set of rules that define access resources and policies, according to INFN, national and European laws.

The INFN Cloud Dashboard is backed by a cloud service orchestrator.

The Orchestrator is in charge of distributing Paas level services among the available IaaS infrastructures.

The INFN Cloud IaaS infrastructures are configured to accept service requests from the Orchestrator.

The INFN Facilities

- 4 National Laboratories
- 20 Divisions
- 6 Associated groups
- 3 National Centres and Schools
- 1 International consortia
- 35 instances of Research Drive (oC10 based)
- 2600+ project folders
- 300TB+ institutional instances
- Dashboard to control different settings of projects
  - Delegate admin privileges
  - Permission to share a project
  - Strict auditing
- Connection to institutional repositories through Sciebo RDS
  - Archiving and publishing data to any repository (SURF sharekit, SURF repository, iRODS, Figshare, Dataverse)
- Moved on Kubernetes, keycloak and OIDC, migration to Ceph S3
- SUNET drive based on Nextcloud
- Openstack and CEPH (s3 buckets)
- Each university has its own SUNET drive node
  - Shares are allowed between different universities
- Separation of access and ownership
  - Ownership of data remains with the original uploader (even if moved to other university)
- Data lifecycles
- 54(x2) nodes on 300 servers in 2 data centres
- 100% powered by green energy
- 120TB data stored (+0.5TB daily)
- Sciebo, 40 ownCloud instances for academic institutions
- 235k users
- 4PB storage (IBM spectrum scale)
- Many insights in running with kubernetes + ArgoCD
  - Split dbs per instance
- Integrations
  - Onlyoffice (in production)
  - Collabora (in production)
  - Overleaf (testing)
  - sciencemesh
Certifying research data combining Seafile and blockchain technology

René Ranger - Max Planck Digital Library

- 87 institutes, 5k users and 179TB storage
- Protect intellectual property!!
- Integration with blockchain + founded BLOXBERG association (50+ universities/institutions)
  - Each contributes a node to verify transactions
- User can certify a given file
- The consortium issues a certificate
CERNBox
CERNBox today

Sync and Share
- Dropbox-like use case
- Offline access to data
- Universal access to data

Web Apps
- Real-time collaboration
- Office documents

Online filesystem access
- CERNBox (Windows)
- Samba (Linux)

Physics integration
- ROOT
- RDA
- RELIC

The tech stack

C) Support scientific workflows (Batch computing)

D) Federated “dropbox” service for HEP community and beyond
CERNBox notifications

Alice sets uploader permissions and enables notifications
CERNBox backup restore

- Data stored in an **other geographical location** in the CERN campus
- +850M of files processed every day
- +5.3PB of files backed up
- ~40k backup job
- Support for different sources
  - EOS
  - CephFS
  - ...

---

GDB | CS3 Conference 2023 Report
ScienceBox rebooted

The demonstrator package for some of CERN’s storage and analysis services:

We present this update as a self-contained and easy-to-use package with minimal dependencies and with the same goals as the original ScienceBox: Provide a sandbox to evaluate the storage, sharing, and analysis services we run at CERN on external premises to non-CERN users.

We believe there is not only a great value in releasing and contributing back to the open source projects that sustain these services, but also in describing the configuration and artifacts that make operating such complex software systems at scale possible.
The Science Mesh
OpenCloudMesh – common file access layer across organizations
Implemented by ownCloud, Nextcloud and Seafile
Perfect for 1:1 communication between nodes

We “just” need to know the other person’s user name on the target system
The Mesh: invitations and apps

The **invitation workflow** was implemented in Reva, and exploited in the NC-Reva and OC-Reva integrations by PonderSource.

- Based on a new endpoint, **yet not part of any tagged OCM protocol version**

- However, no provision was made to support multiple protocols/access methods when creating OCM shares

- Only WebDAV was possible, even data transfers were an “implementation detail”

- Extensions have been defined to cover all that

- More details in [OCM-API#54](#) and [OCM-API#57](#)

**Applications over OCM shares**

- **Model**: a user at site EFSS-A shares
  - A resource, accessible via WebDAV
  - An application to manipulate that resource, accessible via a Web App URL

- **Consequence**: remote users are enabled to
  - Browse the remote storage from their **local EFSS**
  - Access the application(s) available at the **remote EFSS**, via “public” link
  - **Local** applications might be enabled in read-only mode, to prevent conflicts with remote ones

---

GDB | CS3 Conference 2023 Report

32
<table>
<thead>
<tr>
<th>Topic</th>
<th>Presenter</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ScienceMesh Demos Introduction</td>
<td>Pedro Ferreira</td>
<td>14:45 - 14:50</td>
</tr>
<tr>
<td>Connecting Nextcloud &amp; OC-10 to the ScienceMesh</td>
<td>Michiel de Jong</td>
<td>14:50 - 15:05</td>
</tr>
<tr>
<td>Invitation Workflow and Federated Sharing: NC, OC10</td>
<td>Milan Dinecek</td>
<td>15:05 - 15:15</td>
</tr>
<tr>
<td>User-friendly OCM Invitation and Sharing: OCIS, CERNBox</td>
<td>Elizaveta Ragozina</td>
<td>15:15 - 15:30</td>
</tr>
<tr>
<td>Data Science Environments: JupyterLab sharing and collaborative editing</td>
<td>Marcin Sieprawski</td>
<td>15:30 - 15:50</td>
</tr>
<tr>
<td>Data Transfers: Connecting Science Mesh and ESCAPE Data Lakes</td>
<td>Ron Trompet</td>
<td>15:50 - 16:10</td>
</tr>
<tr>
<td>OCM Test Suite current status</td>
<td>Michiel de Jong</td>
<td>14:00 - 14:15</td>
</tr>
<tr>
<td>A View on the CS3 APIs</td>
<td>Klaas Prellag</td>
<td>14:15 - 14:30</td>
</tr>
</tbody>
</table>
Sessions this week

• Wed
  - Sharing applications and data between institutions with Open Cloud Mesh
  - Hugo Gonzalez Labrador

• Thu (Science Mesh public meeting)
  - https://www.egi.eu/article/egi2023-welcomes-CS3Mesh4EOSC/
  - Science Mesh - Unlocking Open Science and Collaborative Research Landscape
    • 11:00 – 17:45
    • Room: Paris
THANK YOU!