



CS3 Copenhagen 2020

Short summary

Jakub T.Mościcki IT/ST

with the notes from:

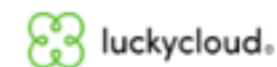
G.Lo Presti, D.Castro, H.Labrador, E.Bocchi

GDB Meeting, CERN, 12 February 2020

Workshop on **Cloud Services** for **Synchronisation** and **Sharing**

27-29 January 2020
Copenhagen, Denmark

Photo // Ulrich Jakobson





www.cs3community.org

Cloud Storage Services for Synchronization and Sharing (CS3)

This is a community of providers, developers and users of innovative storage and sync&share systems. The CS3 services are integrated with user environments and higher-level application services. CS3 reports on the progress in data science at all levels: local laboratories, regional collaborations and global science. CS3 applications range from innovative big-data analysis to science outreach and education.

Conferences organized



2019

Rome, IT
Conference Organized
by
INFN

DOI [10.5281/zenodo.2545482](https://doi.org/10.5281/zenodo.2545482)

[Website](#)

[Programme](#)



2018

Krakow, PL
Conference Organized
by
Cyfronet

DOI [10.5281/zenodo.1157141](https://doi.org/10.5281/zenodo.1157141)

[Website](#)

[Programme](#)



2017

Amsterdam, NL
Conference Organized
by
SURFSara

DOI [10.5281/zenodo.254064](https://doi.org/10.5281/zenodo.254064)

[Website](#)

[Programme](#)



2016

Zurich, CH
Conference Organized
by
ETH Zurich

DOI [10.5281/zenodo.44783](https://doi.org/10.5281/zenodo.44783)

[Website](#)

[Programme](#)



2014

Geneva, CH
Conference Organized
by
CERN

DOI [10.5281/zenodo.2546420](https://doi.org/10.5281/zenodo.2546420)

[Website](#)

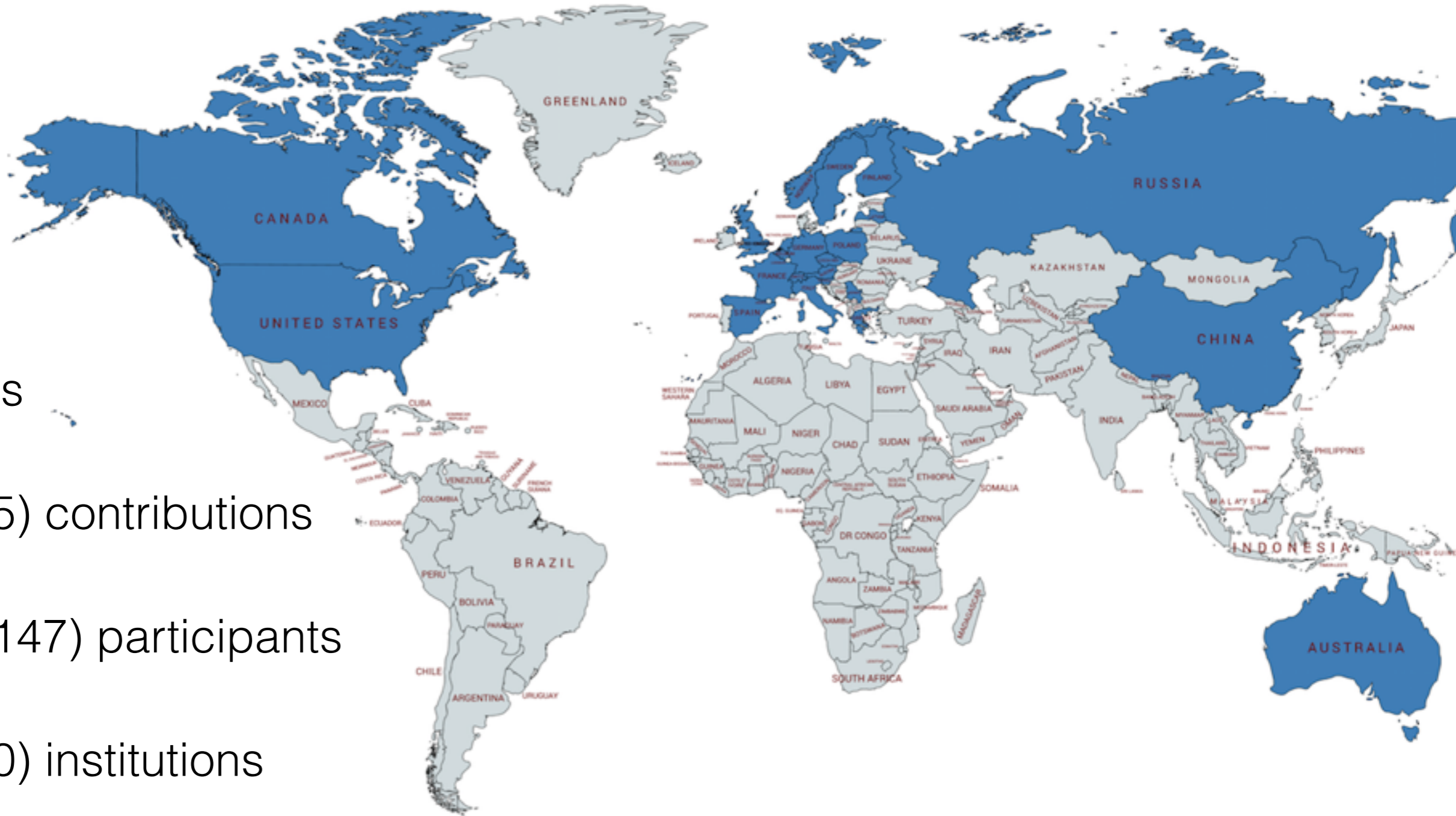
[Programme](#)

Hosted
by





2020— Worldwide Community



- 3 days
- **65** (55) contributions
- **134** (147) participants
- **80** (70) institutions
- **22** (25) countries

2019 logos

científica



NRENS



LABORATORI NAZIONALI DEL GRAN SASSO

HEP&Physics institutes

SURF

SARA

DeiC

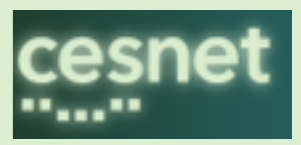


CYFRONET



CNRS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



Istituto Officina dei Materiali



HPC Centers



Organizations



ONE DATA



Universities

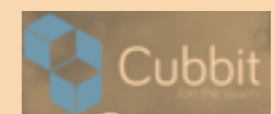
HUMBOLDT-UNIVERSITÄT ZU BERLIN



Companies&SMEs



cynny space



ETH zürich

Collabora Online

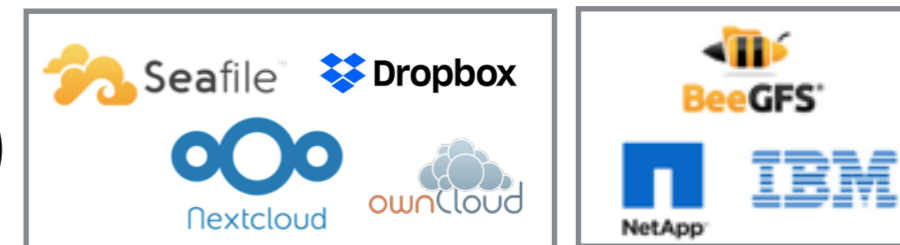


ORACLE



Main themes

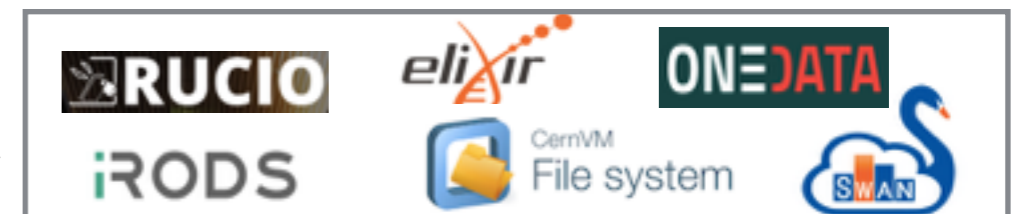
- Site reports
- Digital sovereignty (keynote, panel)
- Cloud storage & sync/share (EFSS)



- Collaborative applications



- Science Environments & Fabric



- CS3 Mesh Project & Interoperability Campfire



Site Reports

Information provided on voluntary basis by EFSS site admins

- Statistics for **23** sites
- **1/2** OC, **1/4** NC, **1/4** Seafire
- **450k** user accounts
- **16PB** stored data
- **10274** project spaces
- **3.5B** stored objects (files and directories)
- **1.5M** shares
- **617k** link shares



HUMBOLDT-UNIVERSITÄT ZU BERLIN



universität wien



UNIVERSITÉ DE NANTES

Software transition

Operational Aspects

Service control

(strongly based on **AAI**

(bwIDM = SAMLbased IdPs

=> scaling), e.g.:

- registration („Ack“ of „terms of use“),
- de-/provisioning of accounts, ...

- „Ack“ of sharing, ownership transfer, ...

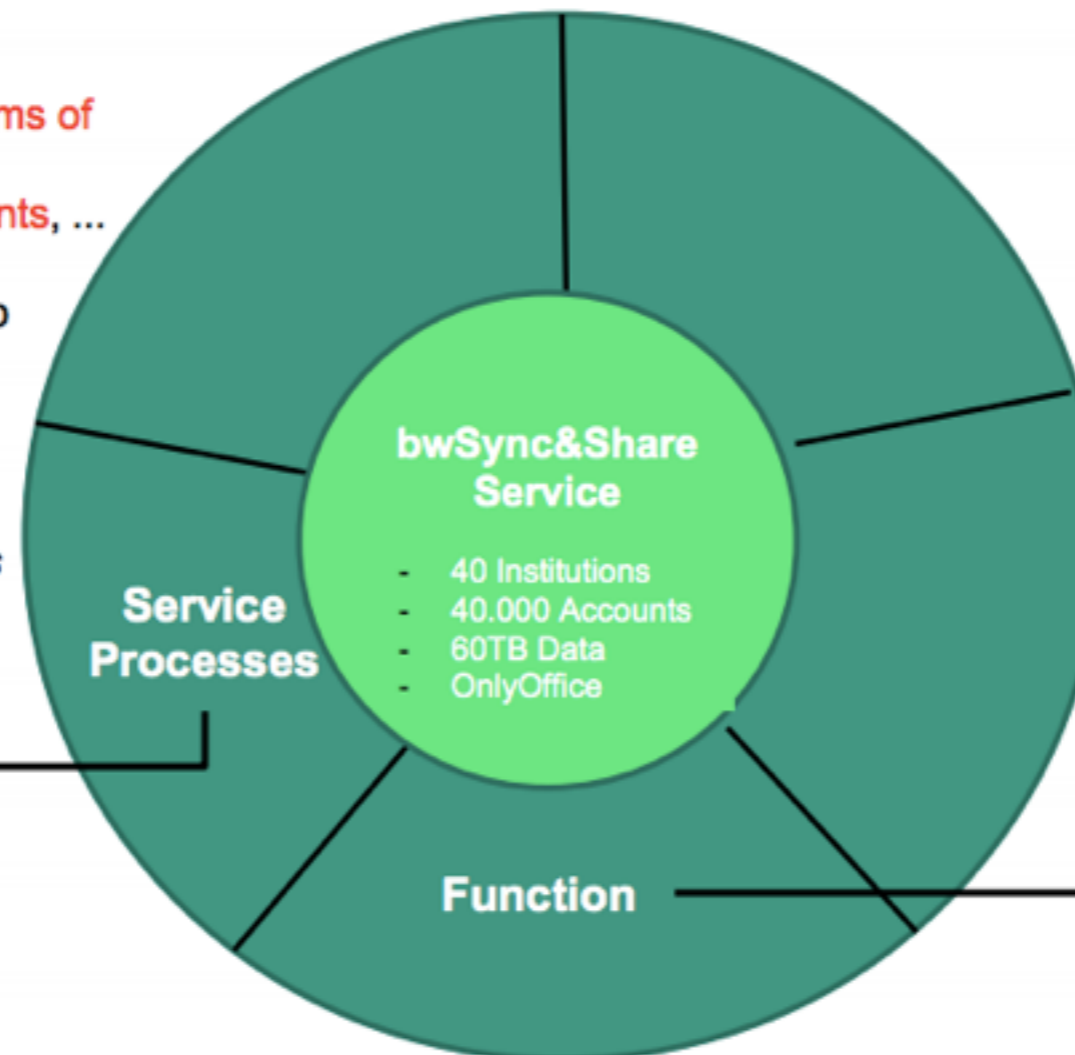
- Documentation: service description, FAQ, ...

(Coincide with other aspects => next slide)

bw = Baden-Württemberg, Germany

bwSync&Share = country-wide

Sync&Share Service for Institutions (edu)



Software

- Actual Powerfolder
- Imminent migration to Nextcloud

Some Standardtopics

- Webediting:
OnlyOffice
- Merging of accounts (several mailing addresses)
- Ownership transfer
- **Server based / E2E encryption**
- **Monitoring**
- **S3/NFS, ...**

BNL and Cloud Storage Requirements

- **BNL: Brookhaven National Laboratory**
 - Multi-purpose US Department of Energy National Laboratory located in NY
 - Conducts a wide range of scientific activity: medical and biological studies, chemistry and nano-material studies, powerful light source, nuclear and particle physics.
 - Hosts Relativistic Heavy-ion Collider, **RHIC**.
 - Will host Electron-Ion Collider, **EIC**.
 - Hosts National Synchrotron Light Source II, **NSLS-II**
 - Hosts National Center for Functional Nano-materials, **CFN**
- **Cloud Storage Requirements**
 - Allow not only nuclear and particle physics BNL users to have cloud storage, but all BNL scientific and non-scientific users
 - ▶ KeyCloak with OpenID
 - Allow BNL users to share with non-BNL users
 - ▶ Guest features
 - Allow users to create and manage their own groups
 - ▶ Circle features
 - Provide archival storage
 - ▶ Lustre with external storage

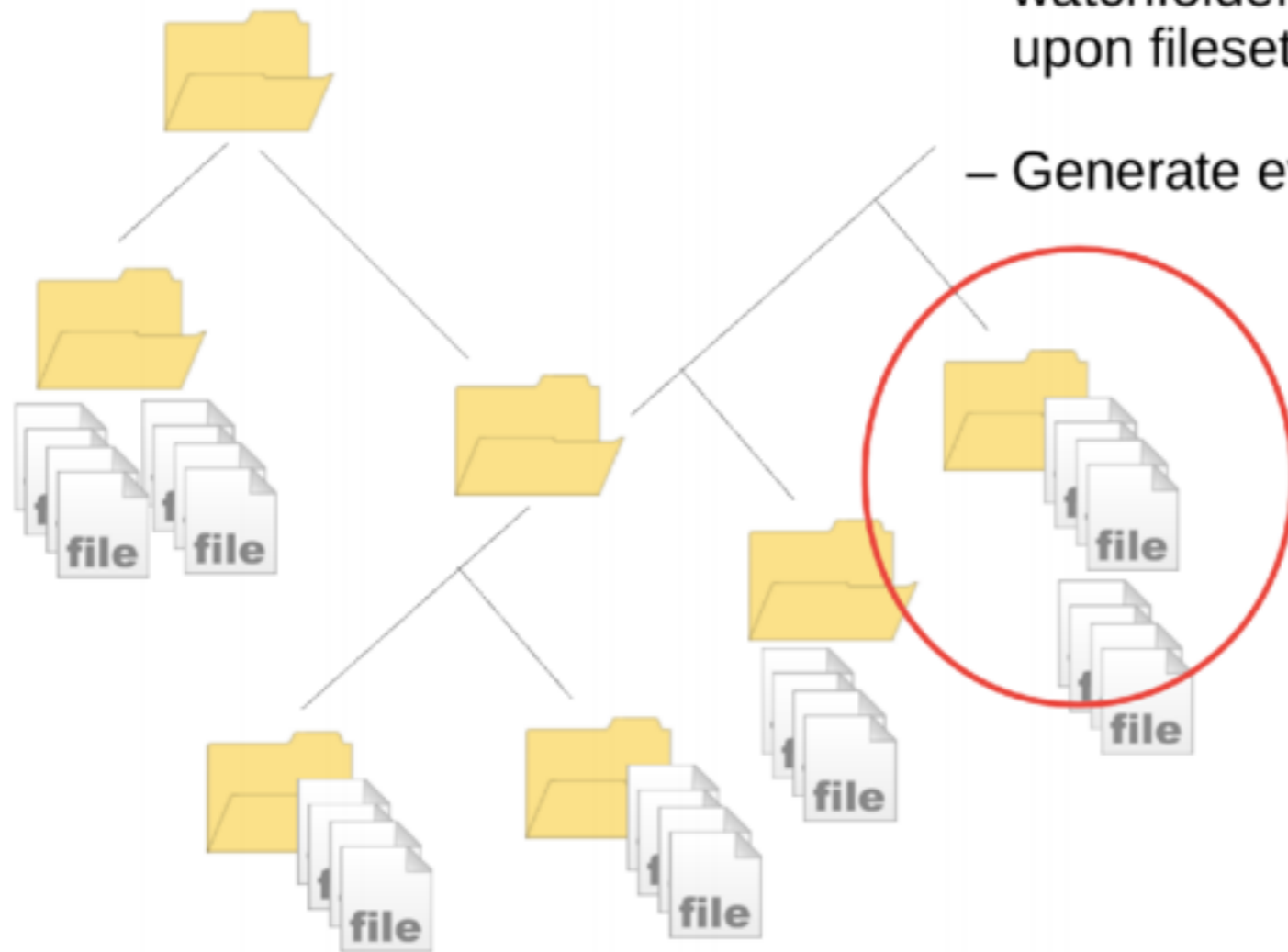
Seafire Educational Users

- Europe
- Canada
- China



Storage Integration

Spectrum Scale Watch Folder



– watchfolder can be enabled upon filesets or directories

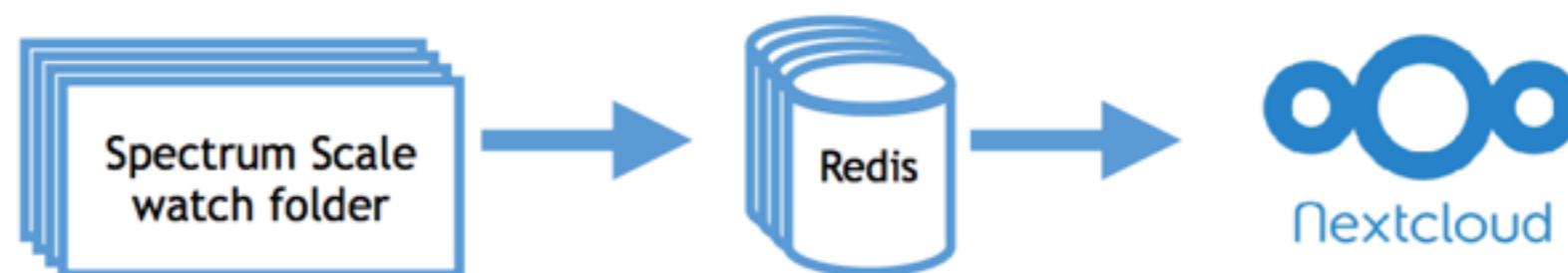
– Generate event on access for

... open, read, write ...

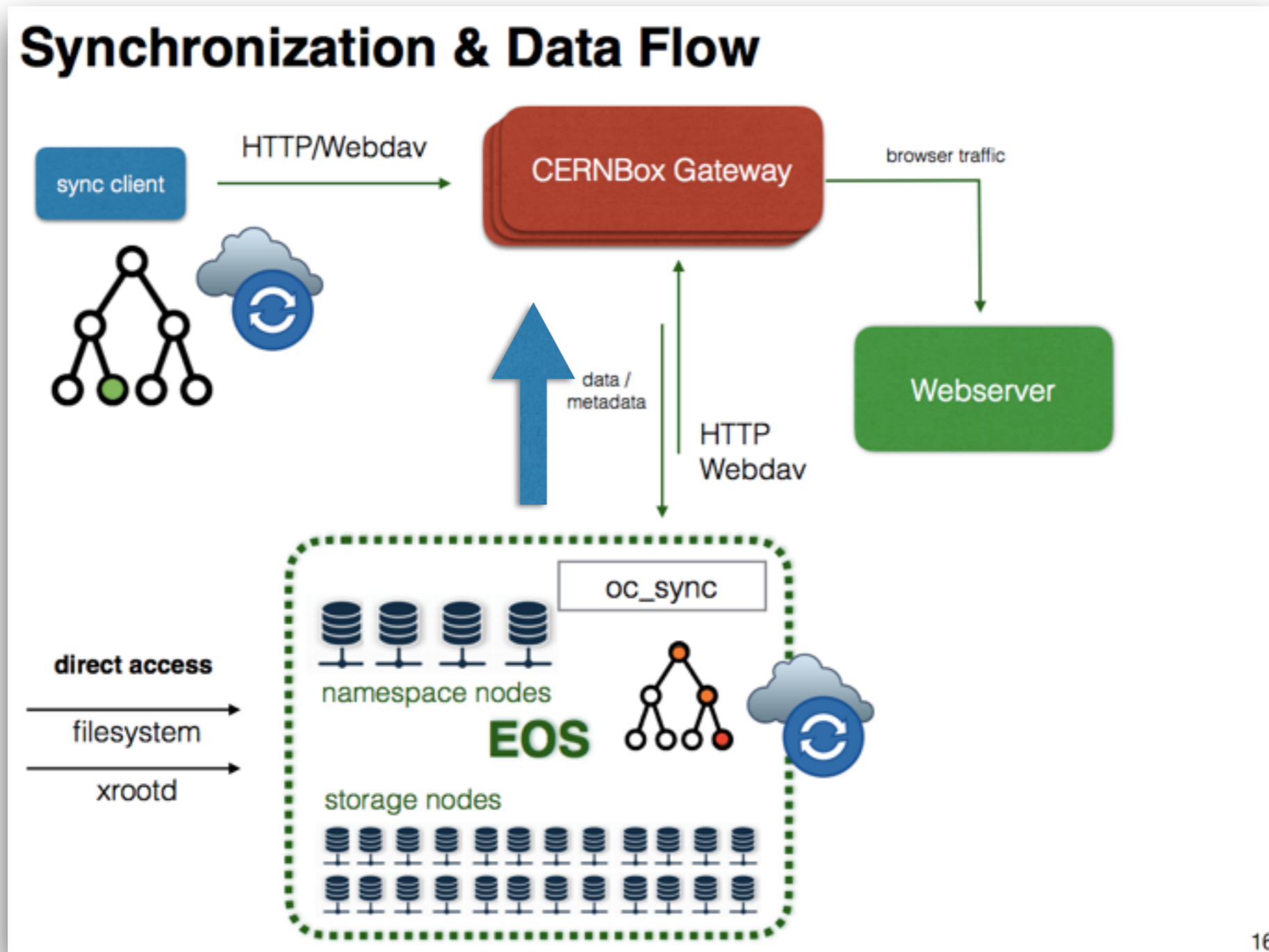
– flexible according to Workload

– uses policy engine / LWE

IBM Spectrum Scale Integration

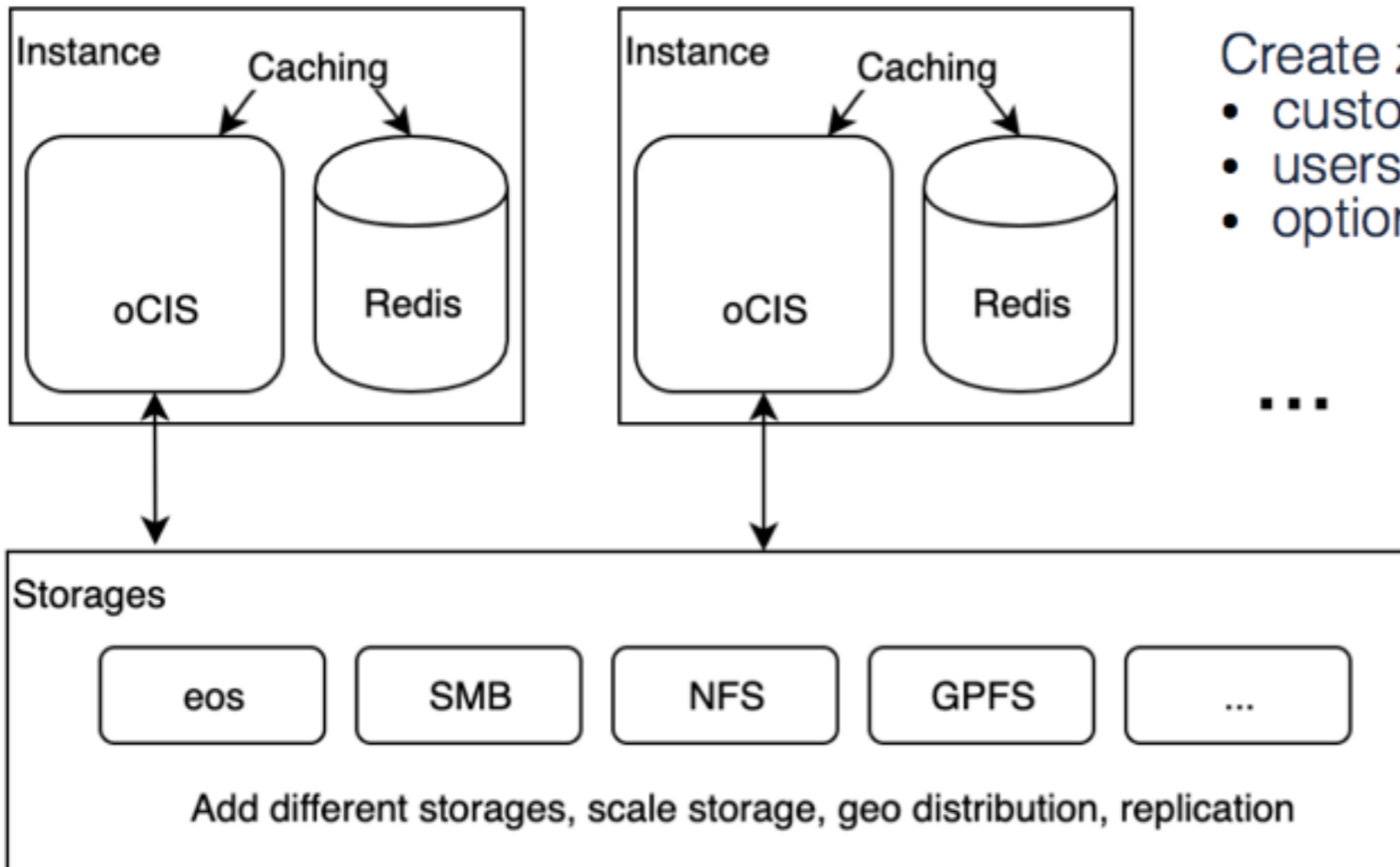


2017 flashback



OCIS scale and deploy

Setup independent instances - Easy scale out



- Create zones for
- custom services
 - users and access rights
 - options and configurations

...

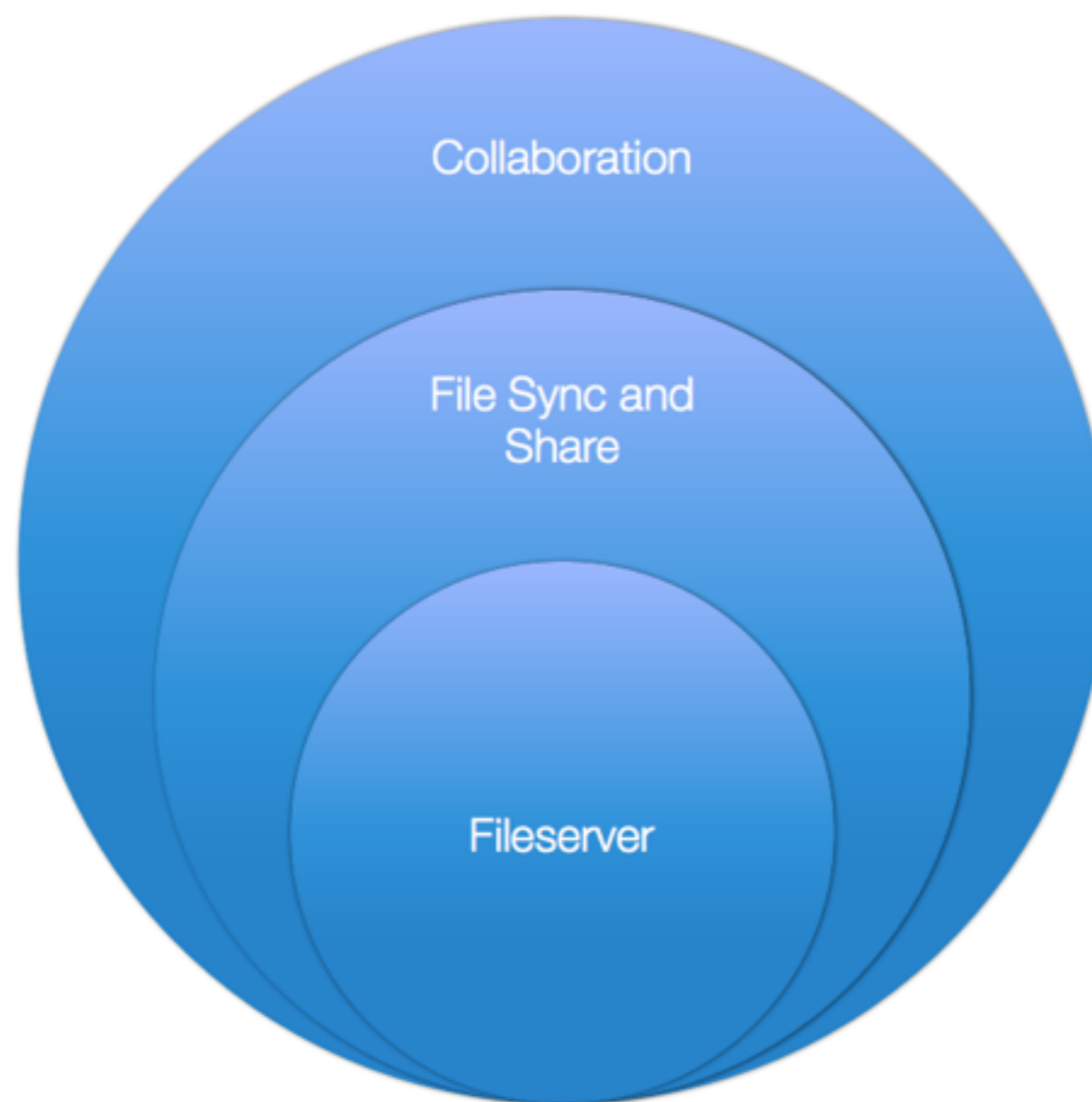


AARNet: S3 shards using Minio and EOS

MINIO DASHBOARD

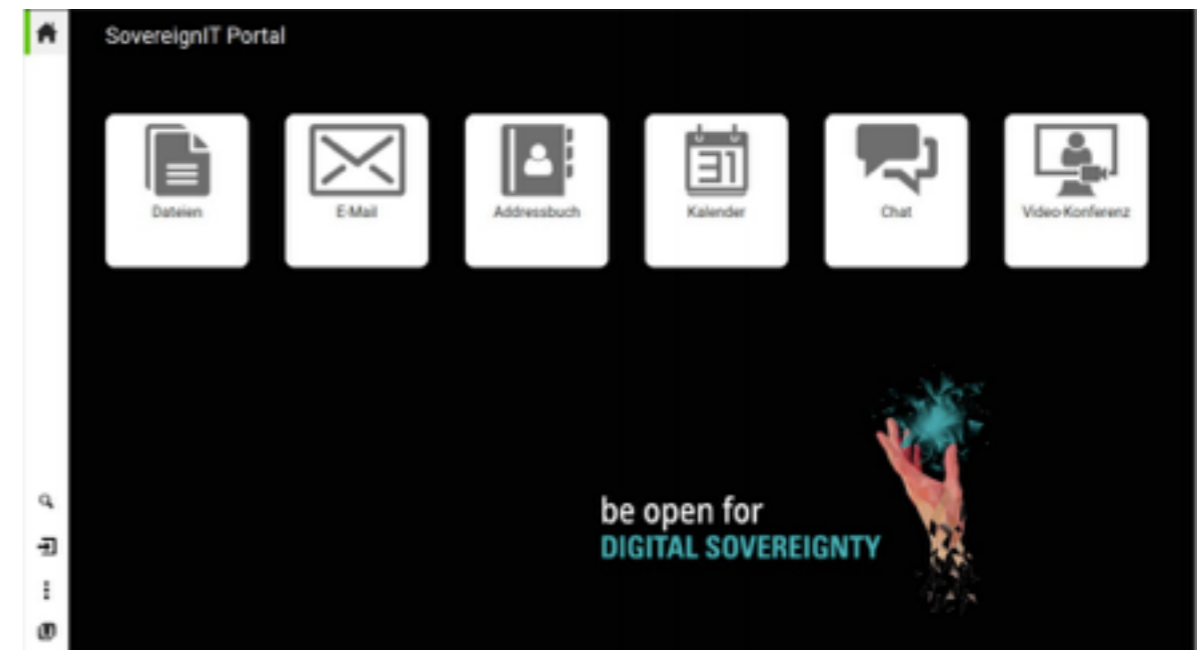
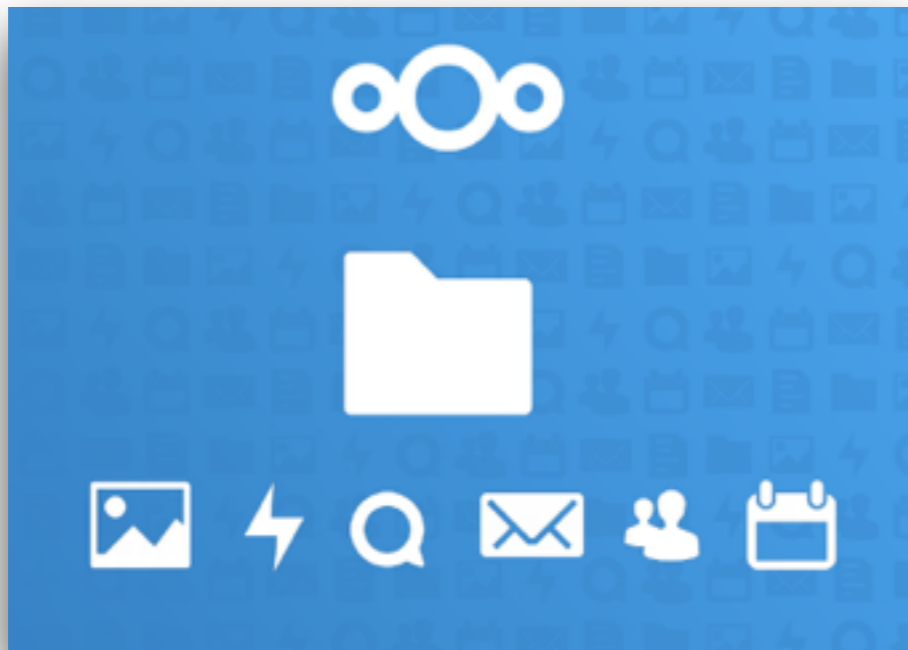


Collaborative Environments





Different approaches

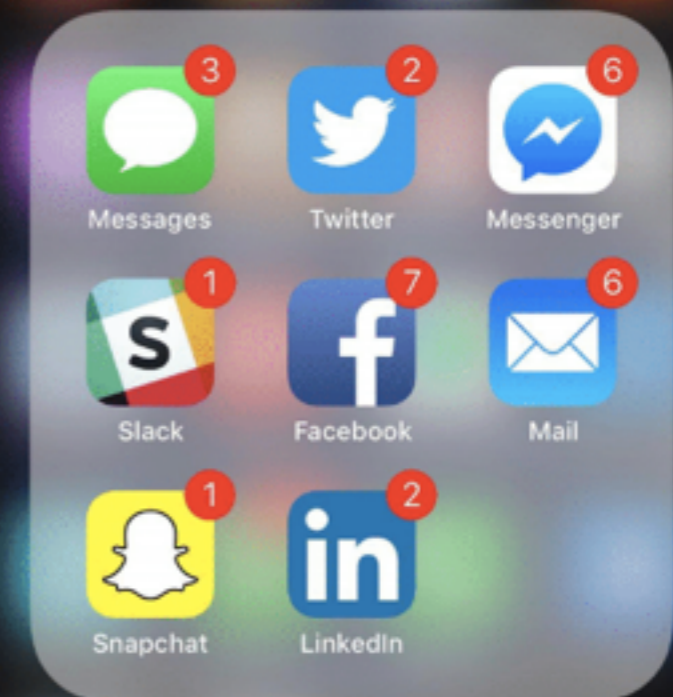
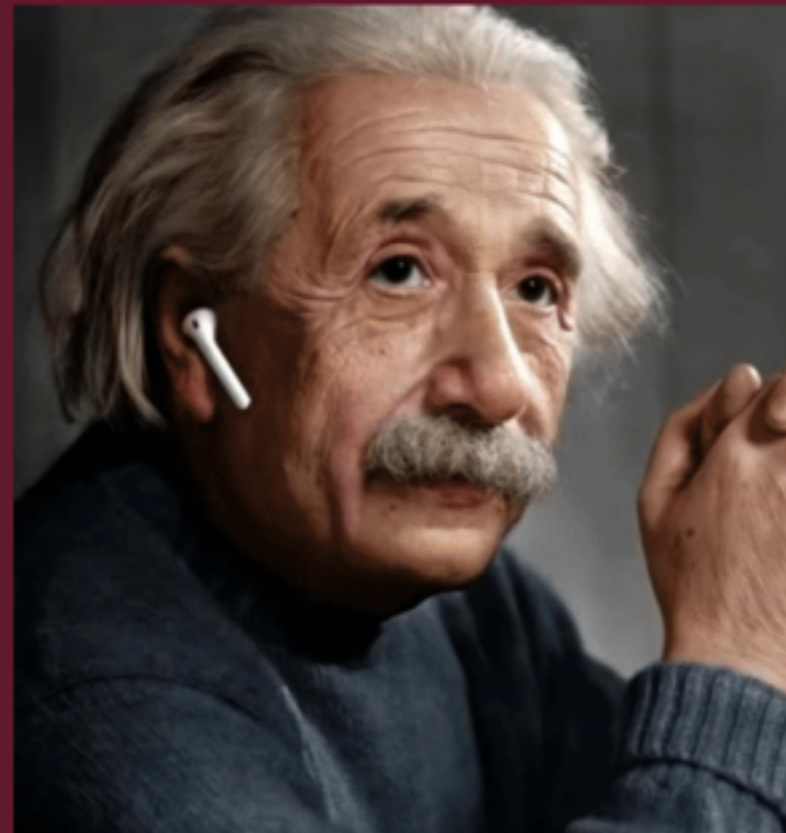


Our ability to focus is broken





#2 Notification-first design gets users





Over 60% of our time is wasted
on work about work



© McKinsey & Company

<https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy>



Dropbox is building the smart workspace

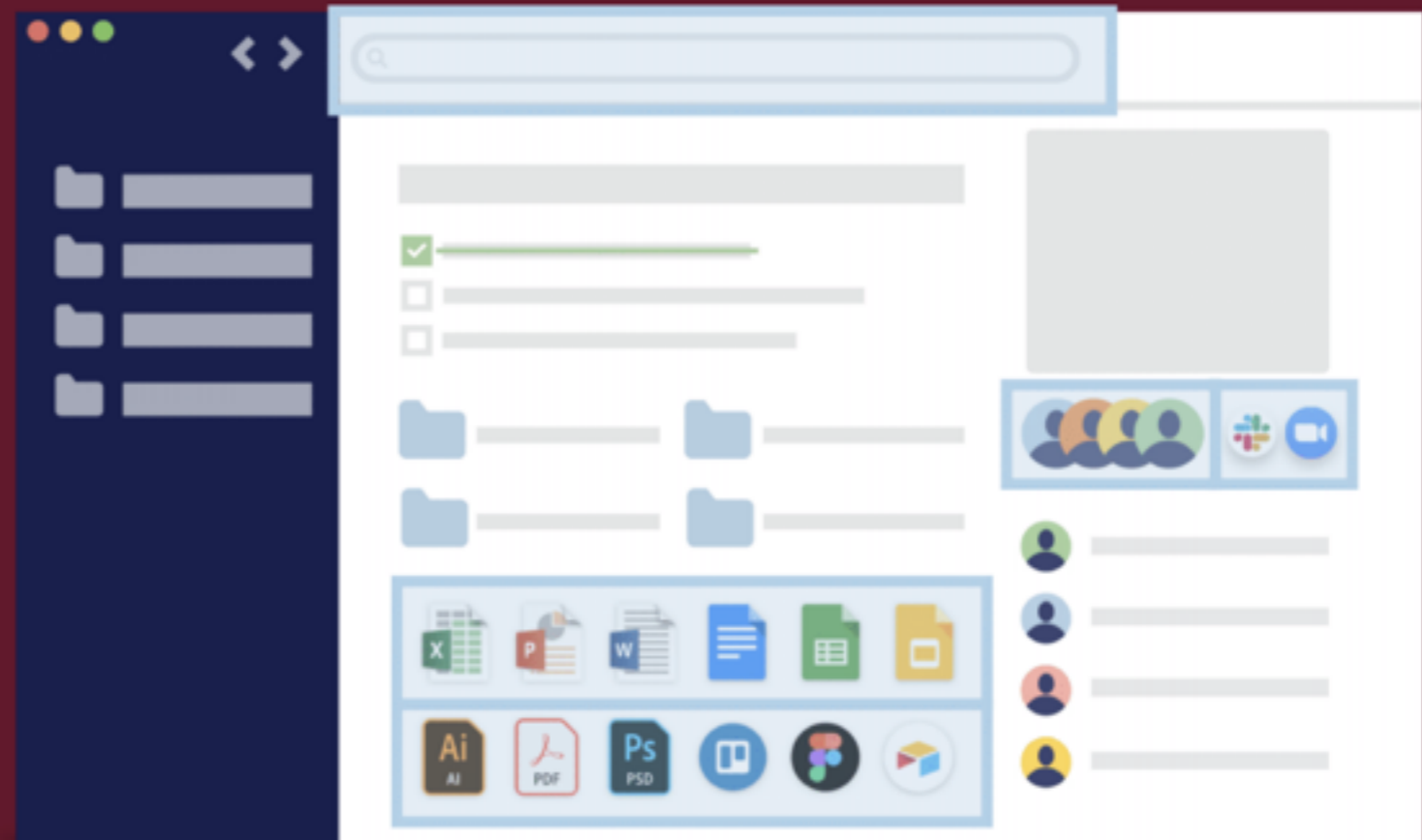
Traditional files

Cloud files

Search across all files

Teams

Collaboration tools



Science Environments & Research Services

A SOLID Distributed Architecture for Sciebo Research Data Services



Cloud

RDS core

Service



100K users



Research Data Lifecycle



WWU MÜNSTER

Upload
Folder
File from Zenodo

	Size	Modified
...	35 KB	a month ago
...	...	a month ago
...	< 1 KB	8 months ago
...	4.8 MB	a month ago
...	5.5 MB	

features.txt
ownCloud Manual.pdf

2 folders and 2 files

Details
Rename
Download
Export to Zenodo



iRODS and Owncloud are integrated

- Files and metadata inserted into iRODS through Owncloud is visible and accessible through iRODS CLI and API
- Files and metadata inserted into iRODS directly is visible and accessible in Owncloud
- Metadata is configurable per customer

Researcher copies folder into iRODS drop zone

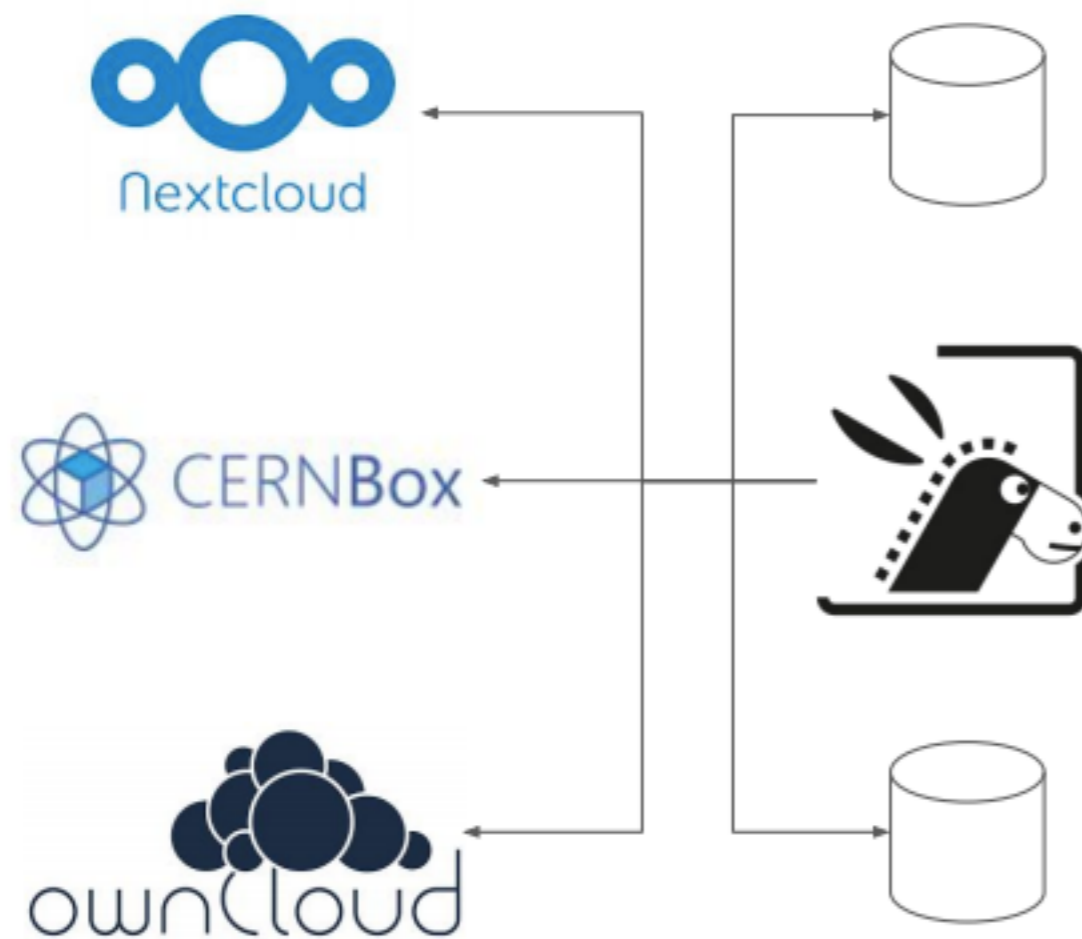
The screenshot displays the OwnCloud web interface. The top navigation bar is green and contains the 'SURF RESEARCH DRIVE' logo, a search icon, and the user name 'Mara'. Below the navigation bar, a list of files and folders is shown:

Item	Size	Time
Documents	35 KB	8 minutes ago
iRODS	Pending	5 minutes ago
Photos	663 KB	8 minutes ago
ThermalSensation	268 KB	seconds ago
ownCloud Manual.pdf		

A context menu is open over the 'ThermalSensation' folder, with the 'Copy' option selected. The menu options are: Details, Rename, Download, Cut, Copy, and a trash icon. A hand cursor is pointing at the 'Copy' option.

Below the main interface, a smaller window shows the 'iRODS > DropZone' view. A 'Paste 1 item' tooltip is visible, and a hand cursor is pointing at the 'Paste 1 item' button. The DropZone currently contains no files, with the text 'No files in here' at the bottom.

Rucio as a large-scale sharing mechanism



- Things to do in Rucio
 - Register Box/Next/Own instances
 - Do things as usual
 - Caches/Scheduling/Recovery/...
- Things to do in Box/Next/Own
 - Sync namespaces with Rucio
 - Notify Rucio about share event
 - Listen to scoped Rucio events
 - User data registration
 - User data movement
 - User data deletion

Rucio namespace integration



playground > sandbox
Last Checkpoint: 20 minutes ago (unsaved changes)

FILE EDIT VIEW INSERT CELL KERNEL WIDGETS HELP

Code mc15_13TeV:HITS.068280...

```
In [1]: TFile *tmp1 = TFile::Open("mc15_13TeV/HITS.06828093._000096.pool.root.1");
        tmp2->ls();
```

TFile**	mc15_13TeV	/HITS.06828093._000096.pool.root.1
TFile*	mc15_13TeV	/HITS.06828093._000096.pool.root.1
KEY: TTree	##Shapes;1	##Shapes
KEY: TTree	##Links;1	##Links
KEY: TTree	##Params;1	##Params

2020-01-28 Mario Lassnig :: Rucio :: CS3 21

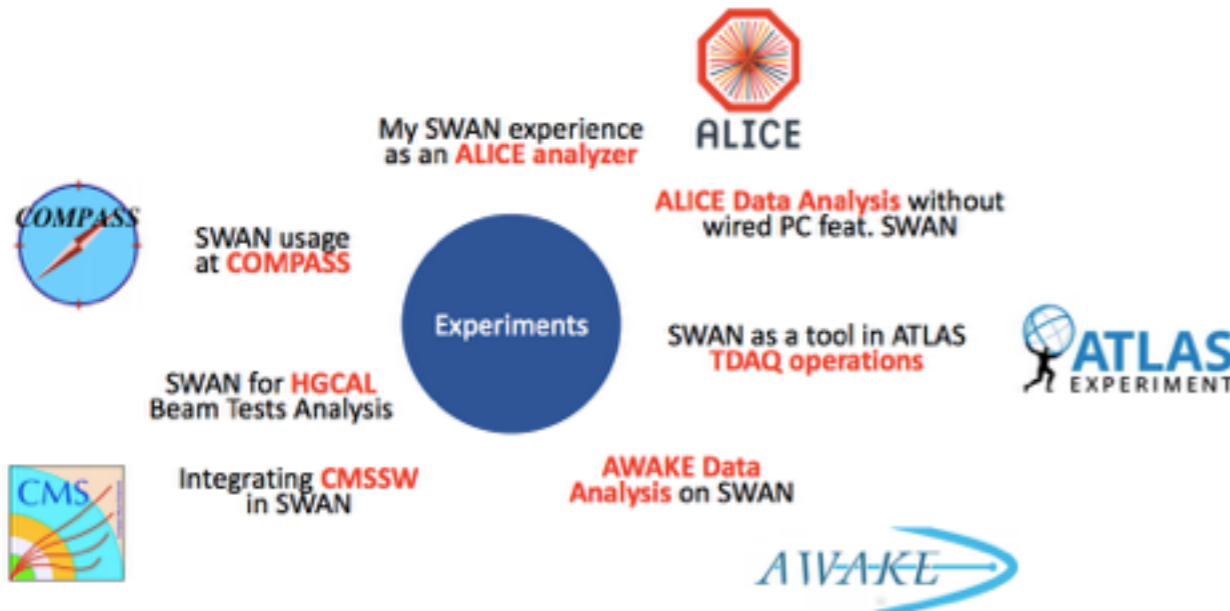


Integrating (CERN) services



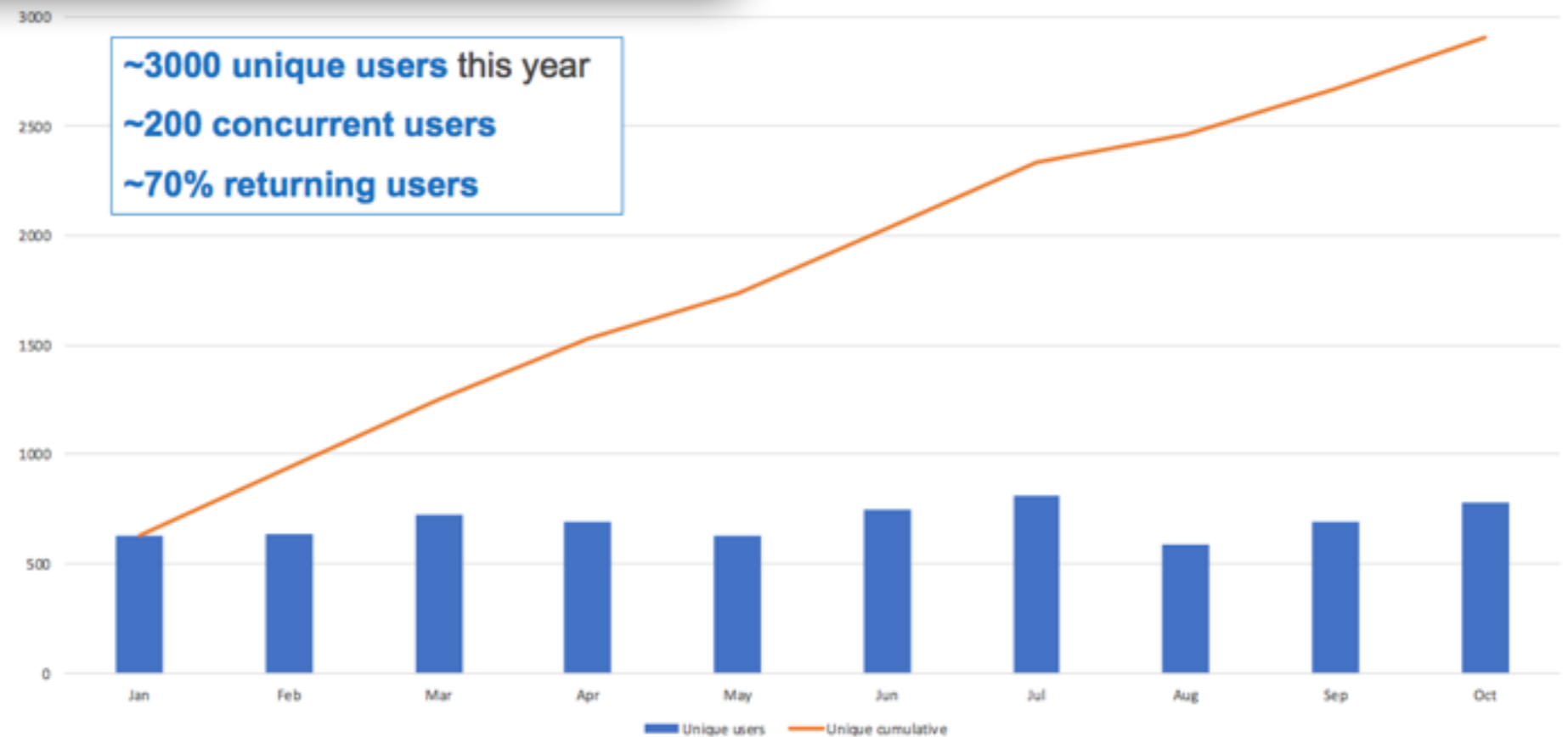


SWAN Users' Workshop



<https://indico.cern.ch/event/834069>

9



The service has been successfully and actively used.

4



CS3 Mesh Project



**EUROPEAN OPEN
SCIENCE CLOUD**

- New EU-funded project to **interconnect CS3 services**
 - Starting January 2020, 3 years
- Deliver a **Global Collaboration Service** for researchers, educators, data curators, analysts, ...
- **Provide an interoperable platform** to easily share and deploy applications and software components within the full CS3 community to extend functionality of the service.

Project Consortium


Coordinator



EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

Address

Route De Meyrin Cern
1211 Geneva 23

 Switzerland


Activity type

Research Organisations




RESEARCH

DANMARKS TEKNISKE UNIVERSITET

 Denmark

UNI / NREN

SURFSARA BV

 Netherlands




HPC / NREN

INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII
NAUK

 Poland

HPC

CESNET ZAJMOVE SDRUZENI PRAVNICKYCH OSOB

 Czechia


NREN

AARNET PTY LTD

 Australia


NREN

SWITCH

 Switzerland

NREN

WESTFAELISCHE WILHELMS-UNIVERSITAET MUENSTER

 Germany

UNI

AILLERON SA

 Poland


TECH

CUBBIT SRL

 Italy

TECH

JRC -JOINT RESEARCH CENTRE- EUROPEAN COMMISSION

 Belgium

EC RESEARCH

FUNDACION ESADE

 Spain

BUSINESS SCHOOL

Starting point

- 12 institutions to **create initial infrastructure**
 - connect existing, sustainable services
 - all major EFSS platforms included (multivendor)
 - 200K+ existing users, 10PB of sync&share data, >1billion files and objects
- The infrastructure will be **gradually expanded** and **integrating the entire community**, education and research in Europe and beyond.

Collaborative Workflows

Integrate existing experience and technology



Share, access, synchronize



Metadata&tagging, Open Data (OpenAIRE, Zenodo,...)



Data Science: Jupyter Notebooks (SWAN,...)




Collaborative editing, Latex, Markdown, Indico, ...












On-demand data transfers (Rucio, FTS, FileSender,...)

Pilot users

Students, educators and researchers at large

-  • connecting large university campuses

Target specific research and application areas

-  • Earth Observation (Copernicus Sentinel)
-  • High Energy Physics (LHC)
-  • Astroseismology (NASA Kepler telescope)
-  • Cultural Heritage and Archival Collections
-  • Material Science
-  • Astrophysics (LOFAR)
-  • Plasma Physics
-  • Video processing technology development
-  • Diabetes Research



LOFAR

Kepler



WLCG
Worldwide LHC Computing Grid



PARADISEC

Pacific and Regional Archive for
Digital Sources in Endangered
Cultures



Edu & Outreach

“Early adopters”

Backbone Services

Physical Layer

- Helmholtz VPN

Services

- Trust and Security
- Compatibility, nationally and internationally
- Common Authentication and Authorization Infrastructure (AAI)
- Backup and data transfer services
- Operational Services



Cloud Services

Why Mesh

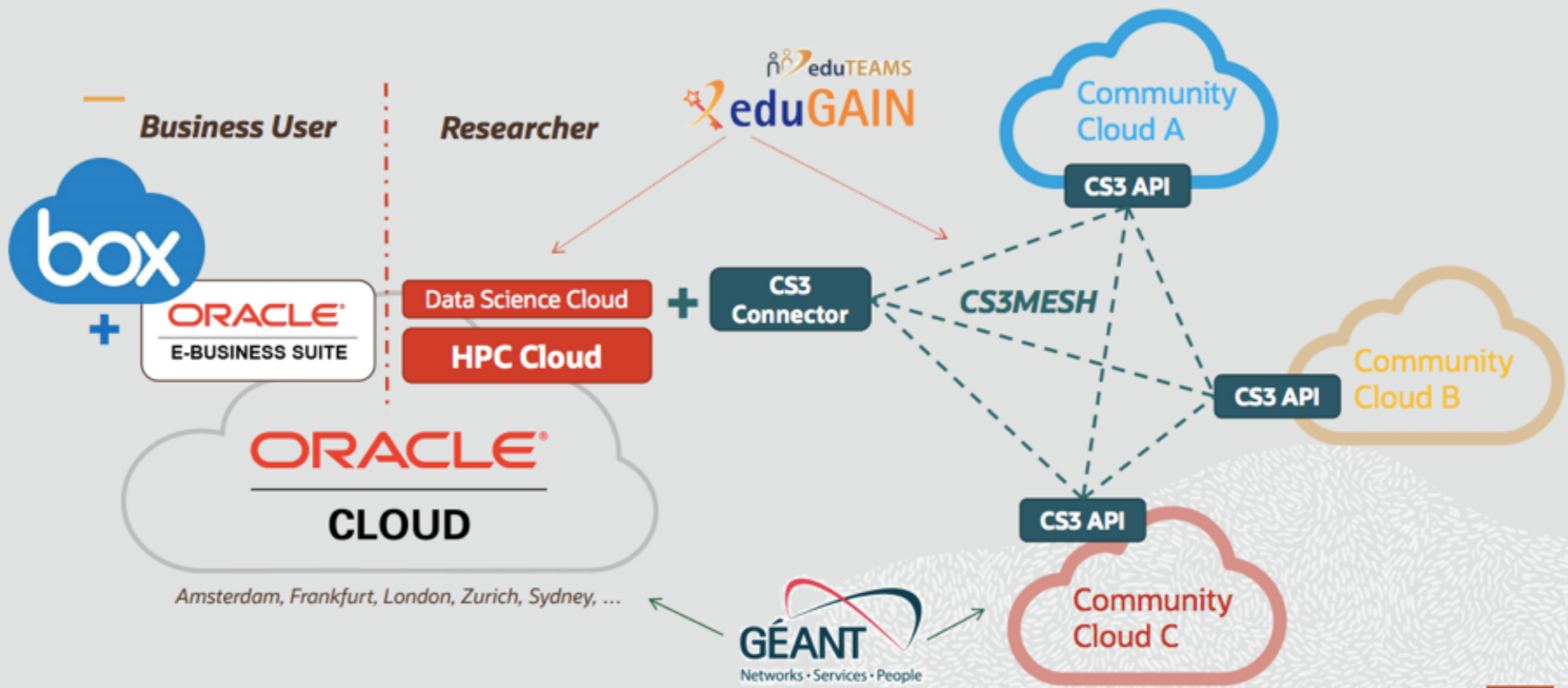
Mesh of Sync&Share and Collaboration



Mesh for Scientific Workflows



Vision for CS3MESH connector to Oracle Cloud Infrastructure



To a federation of Drives



Final thoughts

HEP community

- Better overview of EFSS services across HEP community
- <https://indico.cern.ch/event/854707/surveys/1173>
- Your view on federated collaboration services across all HEP sites...

contact@cs3community.org