



MeetCS3MESH

Jakub T. Mościcki

CERN IT, Storage

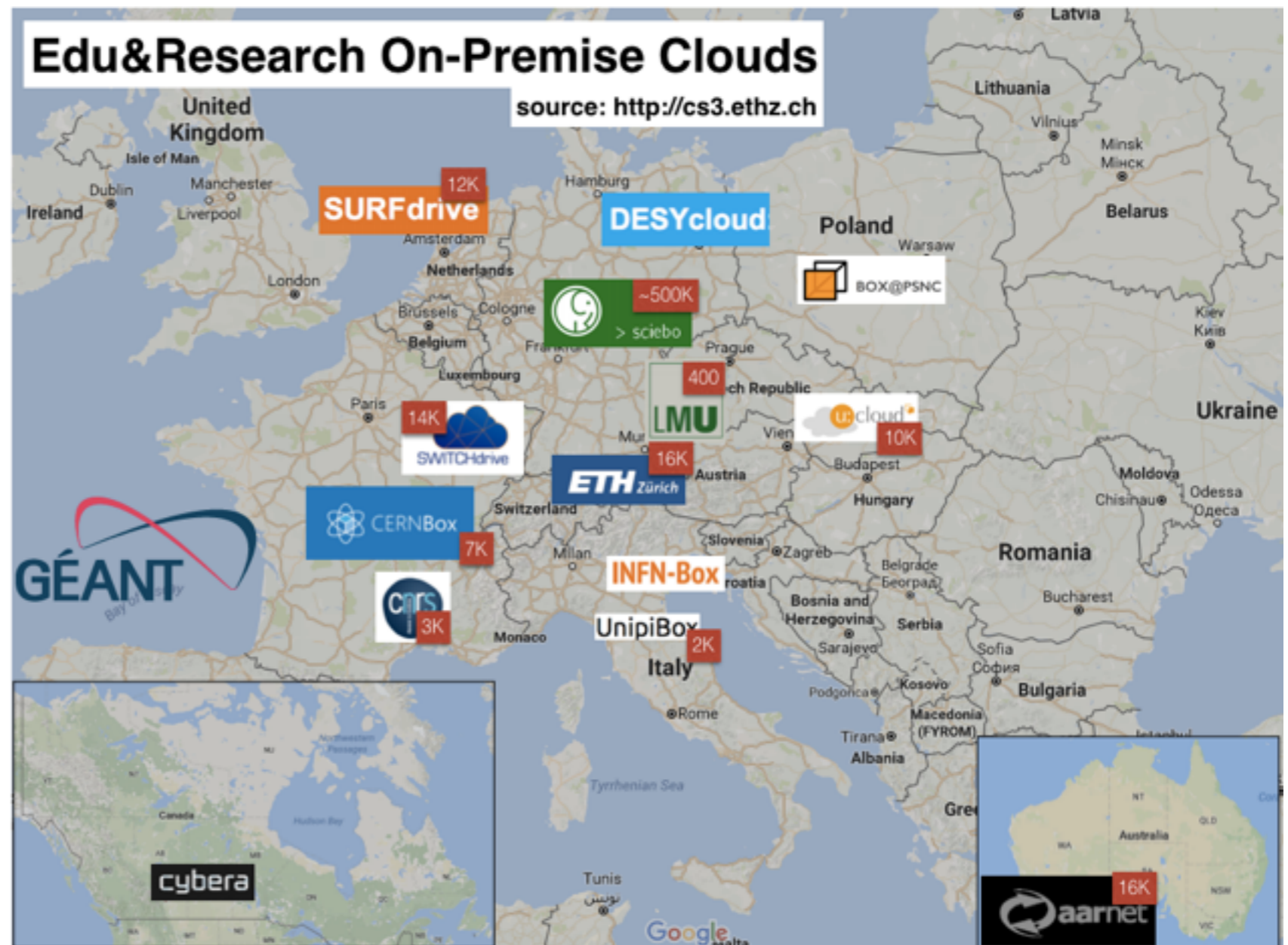
CS3, 27 January 2020, Copenhagen

Sync&Share in European Research Landscape in 2010s

Working groups

- TF-Storage
- CS3
- SIG-CISS
- ...

Innovative companies



Sync&Share today:
large-scale, indispensable, daily use

CS³ Site Reports **2019**

- **25** sites
- **5** EFSS technology stacks
- **400k** users
- **2.7 billion** files and directories
- **11.5 PB** storage
- Yearly storage increase: **5-250%**

Information provided on voluntary basis by CS3 site admins (January 2019)

Many CS3 sites have gone the extra mile...

- + applications, collaborative tools: application hubs ...
- + research interfaces: interactive notebooks, data management, computing clusters, file transfer, HPC bridges,....
- + integration with big data repositories
- + Open Data publishing: implementing FAIR in practice

Clear evolution: from simple file stores to rich, data science environments and Research Data Services (RDS)

Innovative EFSS industry

- Moving higher up the functionality stack, e.g.
 - Rich collaborative environments
 - Windows Drive integration
 - Federated sharing and the OCM protocol
 -



Great success, but...

- Efforts are currently largely disconnected across CS3 community
 - **researchers** face **isolated service islands**, may not use other site services in spite of existing research collaborations across institutions
 - CS3 **sites** do not benefit from research services developed elsewhere = **risk of high cost or inadequate/obsolete service**
 - CS3 **application developers** **duplicate effort** over multiple EFSS technology stacks
 - EFSS **technology companies** **do not benefit from the community know-how**: poor transfer back to commercial and business applications

Ongoing discussions in last 2 years...

2018

Thoughts on CS³

- **We have a unique community via a series of successful events**
 - **How do we preserve this capital for the future?**
- **CS3 areas of excellence**
 - **Data-intensive science and education:** from prototyping to publication,...
 - **Non-traditional working storage:** synchronisation, sharing, federation,...
 - **Storage technology:** protocols, files and objects, disks, tapes, SDS
 - **Collaborative platforms**

New opportunities 2020



- **European Open Science Cloud**

- Create a trusted environment for hosting and processing research data to support EU science in its **global leading role**
- Open Science = the EU aims to **make research more efficient, reliable, collaborative and transparent.**
- **Data centric** & FAIR (Find, Access, Interoperate, Reuse)
- **Market & innovation potential** for European Tech industry



- **Digital sovereignty**

- Who controls the data controls the future
- Many organizations looking into solving this problem, often with on-premise services: governments, big institutions, ...



**EUROPEAN OPEN
SCIENCE CLOUD**

- New EU-funded project to **interconnect CS3 with EOSC**
 - 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.

..with a wide community support

- Over 20 support letters from CS3 community & beyond — **big thank you!**
 - including all major on-premise technology companies present at this conference!
 - many user & research groups

EC perspective

*The project delivers the core of a **scientific and educational** infrastructure for cloud storage services in Europe [...]*




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.

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

Future Federated Analysis Platform

Advancing state of the art

One-click to create user groups, share projects and data



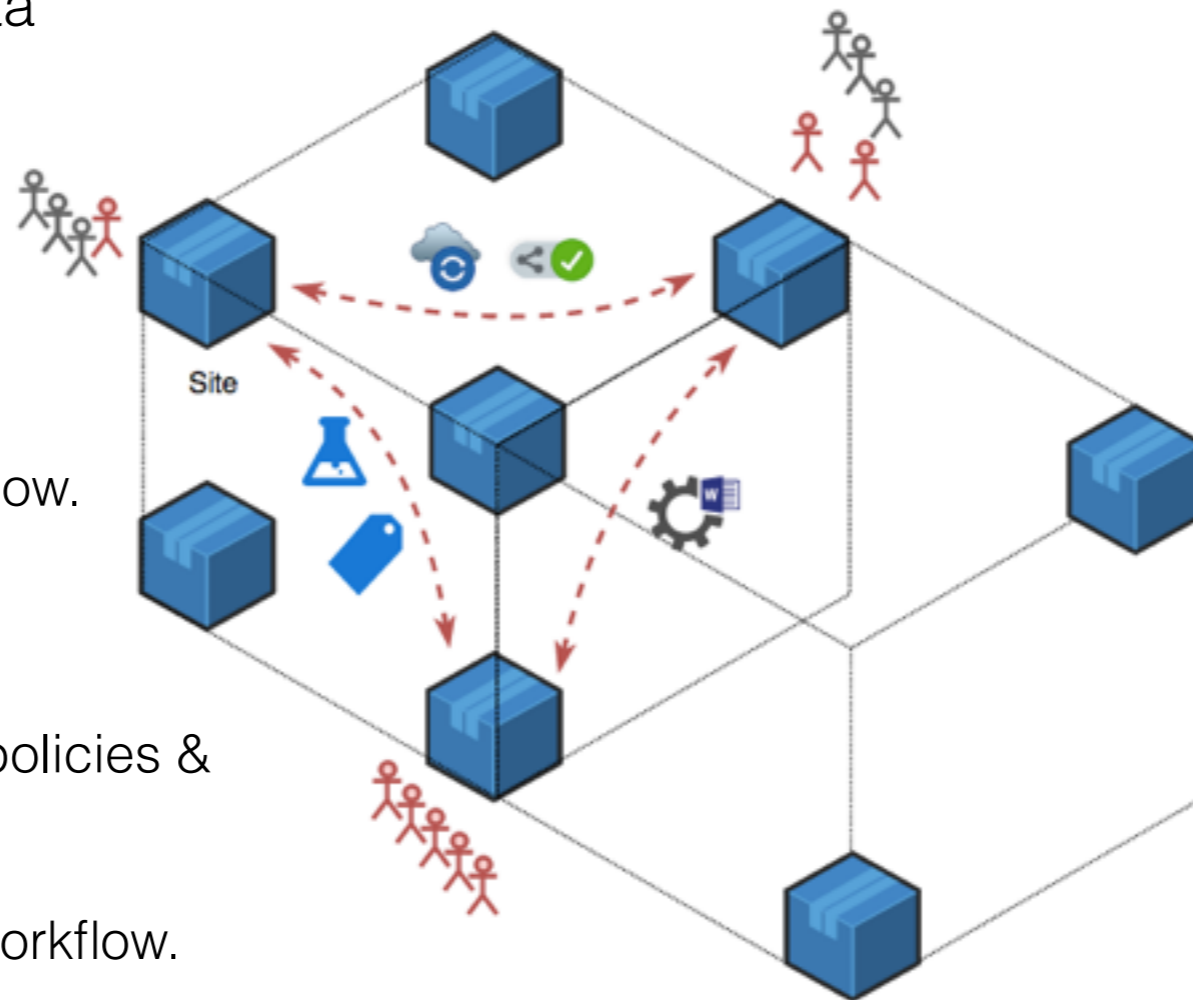
Domestic and remote users in the same collaborative workflow.

Collaborate on data across institutions, respecting local policies & GDPR

Application&data workflow.

Data available on **all devices**: mobile, laptops, desktops

Full metadata awareness in the research workflow.



Global Computing Data Fabric & Digital Repositories

Integrated with data sharing capabilities

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,...)

Technology Integration

- **Connect EFSS platforms with research services and digital repositories**
 - Promote vendor-neutral APIs and protocols
 - Open-source software development and service delivery
- **Sustainable technology**
 - Collaboration on technology with **all** EFSS, storage and application vendors

Interoperability

- Add thin layer on top of existing services



- Use existing fabric



- authentication, monitoring,...



- close collaboration with GEANT, EGI, EOSC-hub,...



OPENCLOUDMESH

EFSS Native

A vendor neutral standard under the GÉANT umbrella

- Use existing standards



- Introduce new APIs only if needed



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

Expectations

- **Users** will gain a global collaboration platform which will be seamlessly integrated in their current working environments.
- **Institutions and service providers**
 - freedom to implement their service strategy without compromising on functionality and creating disconnected service silos for their users
- **Service managers and site administrators**
 - streamlined environment for stable service operation at scale.
 - cumulated knowledge, operational experience and support of a large community
- **Tech Industry in Europe and beyond**
 - opportunity to extend the user base
 - new markets & opportunities


Early adopters

This is an invitation to co-design the service

- Contributions and discussions are welcome via open forum and within CS3
- New sites are encouraged to join and provide feedback
- Involving new user groups, research workflows and applications
- EOSC projects: GN4.3, OCRE, ARCHIVER, ESCAPE
- Public and commercial cloud providers

Input from community

15:00

The HIFIS Cloud Competence Cluster	<i>Dr Andreas Klotz et al.</i> 	14:30 - 14:50
FAIR Implementation Profiles: Driving Convergence onto an Internet of FAIR Digital Objects	<i>Dr Erik Erik Schultes</i>	14:50 - 15:10
ESCAPE ESFRI Science Analysis Platform	<i>Zheng Meyer-Zhao</i>	15:10 - 15:30
Oracle for Data Science – A Vision for CS3MESH Connector	<i>Mr Peter Szegedi</i>	15:30 - 15:50
ScienceBox as Deployment Model for CS3MESH	<i>Enrico Bocchi</i>	11:15 - 11:30
Elettra Drive(s) data far and FAIR	<i>Dr Ivan Andrian</i>	11:30 - 11:45
Drive RENATER	<i>Alexandre Salvat</i>	11:45 - 12:00
OSF and CS3MESH	<i>Nicole Pfeiffer</i>	12:00 - 12:15

