MEET CS3 – MESH
A GLOBAL CYBER INFRASTRUCTURE

Carina Kemp - AARNet

Jakub T. Mościcki – CERN

Guido Aben - AARNet

eResearch NZ

14th February, 2020

SYNC & SHARE IN THE RESEARCH LANDSCAPE IN 2010S

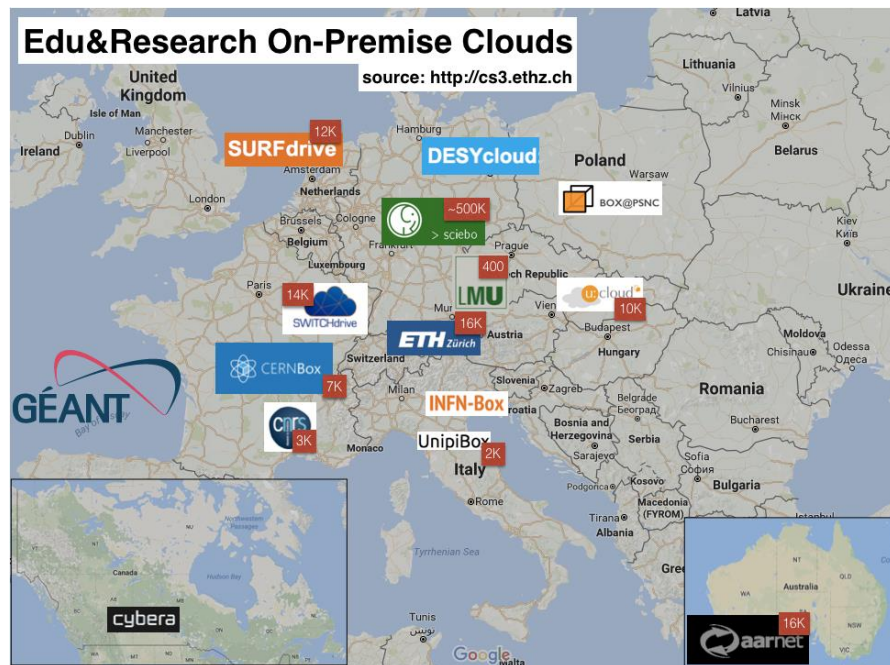## *Cloud Storage Services for Synchronization and Sharing (CS3)*

Working groups

- TF-Storage
- CS3
- SIG-CISS
- …

Innovative companies

CS³ 2016

Edu&Research On-Premise Clouds
source: http://cs3.ethz.ch

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)

Moving higher up the functionality stack, e.g.

Rich collaborative environments

Windows Drive integration

Federated sharing and the OCM protocol

…

GREET 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, …

- 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.

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
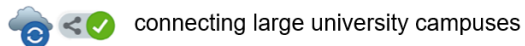
connecting large university campuses

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
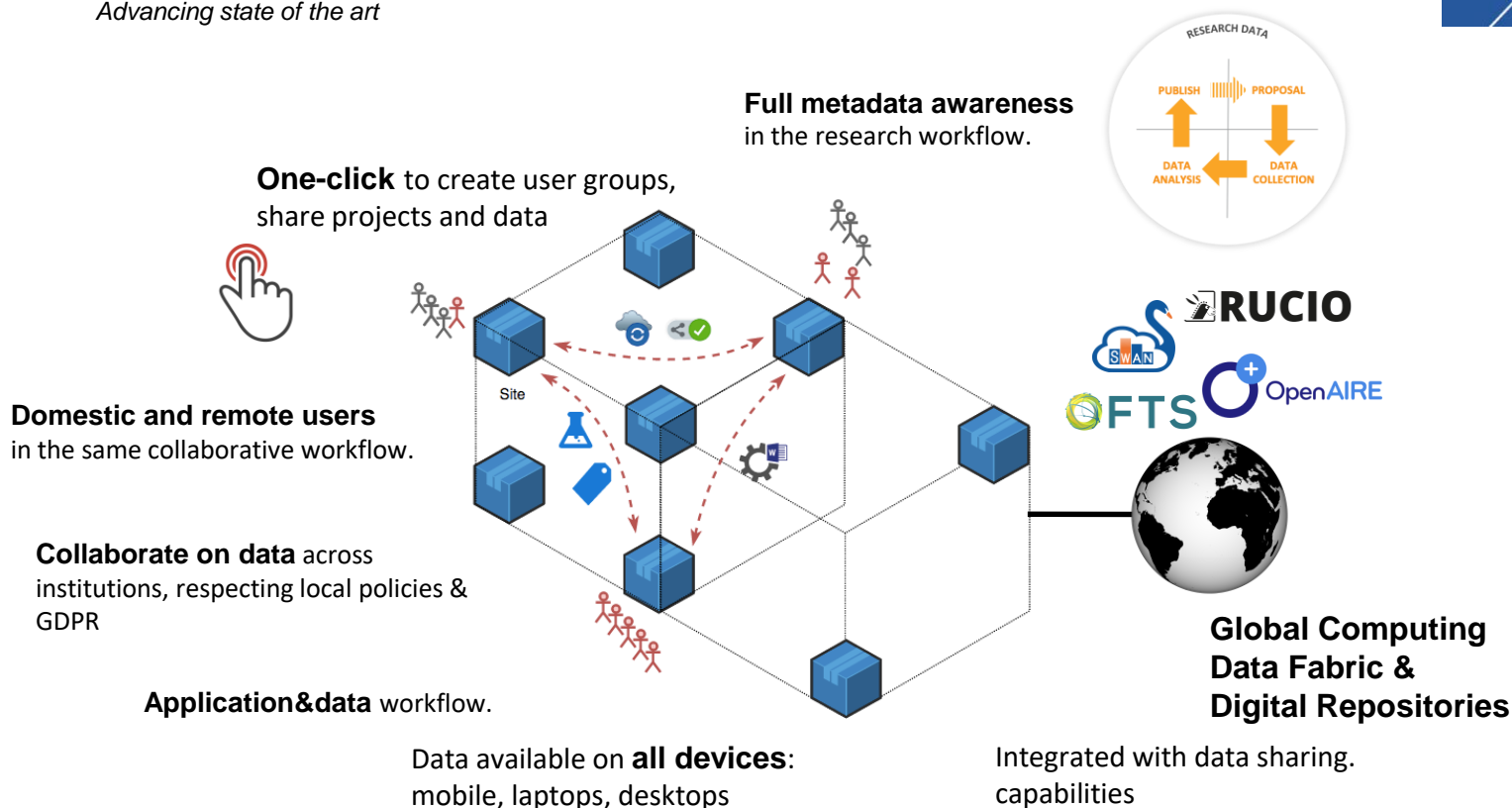
· Video processing technology development

Diabetes Research

Edu & Outreach

Pacific and Regional Archive for Digital Sources in Endangered Cultures

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 sync&share 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** sync&share, 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,…

- **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

Activity type
Research Organisations

Switzerland

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 – JOIN US…

This is an invitation to co-design the service

- Contributions and discussions are welcome via open forum and within CS3 community

- 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

- International beyond Europe….

THANK YOU