CS3 2019

Programme Development, SIG-CISS & Strategic Directions

Jakub Moscicki, CERN/IT-ST

3rd SIG-CISS meeting
Berlin, November 2018

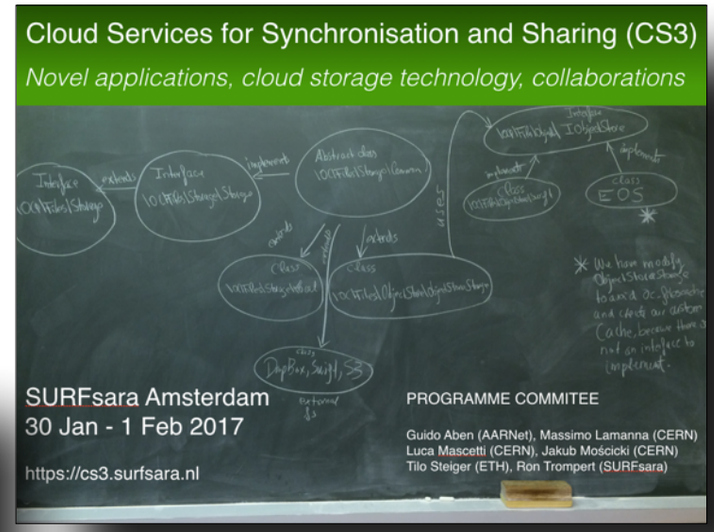
What is CS3 about?

- **“Dropbox for science”** -> sync/share on premise
  - innovative storage systems and their integration with user environments
  - to enable progress in data sciences at all levels: local laboratory, regional collaborations and global science
  - CS3 applications range from innovative big-data analysis to science outreach and education

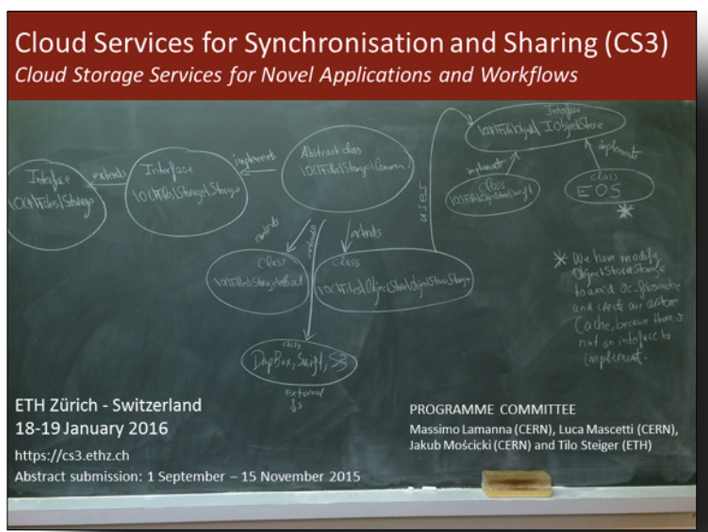
5 editions since 2014

Cyfronet
Kraków 2018

INFN
Rome 2019



SURFsara
Amsterdam 2017



ETH
Zurich 2016

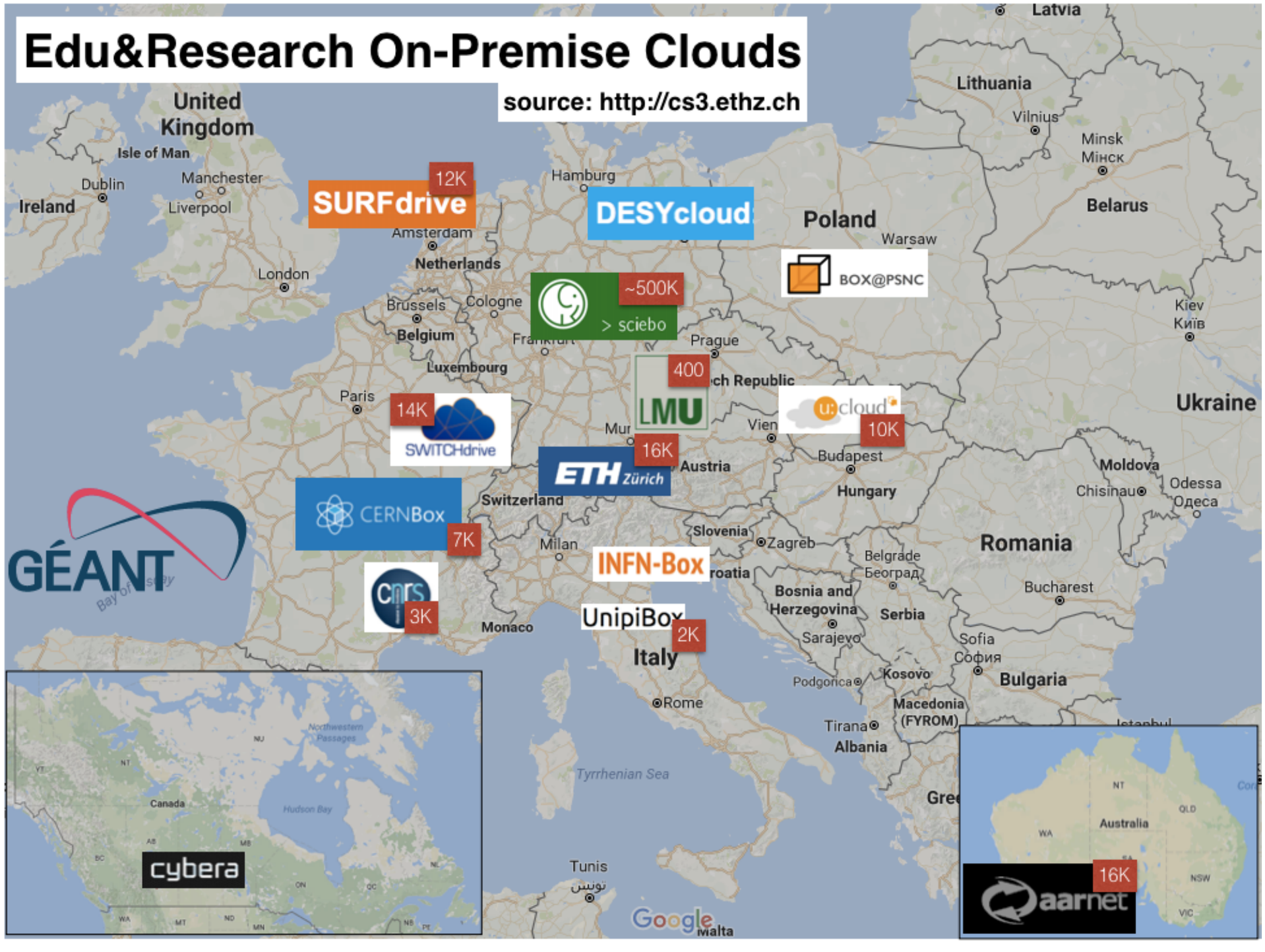
CERN
Geneva 2014

Evolution of CS3

- Kick-started by CERN in collaboration with TF-Storage and GEANT partners, designed to be **inclusive**
- Technical focus, share ideas, speak up about problems,…
- Neutral ground for competitors: multiple software providers from early days
- Evolved
  - from small sync/share core technology focus
  - to a place for creating opportunities for everyone

Initial nucleus

2016



a place for creating opportunities

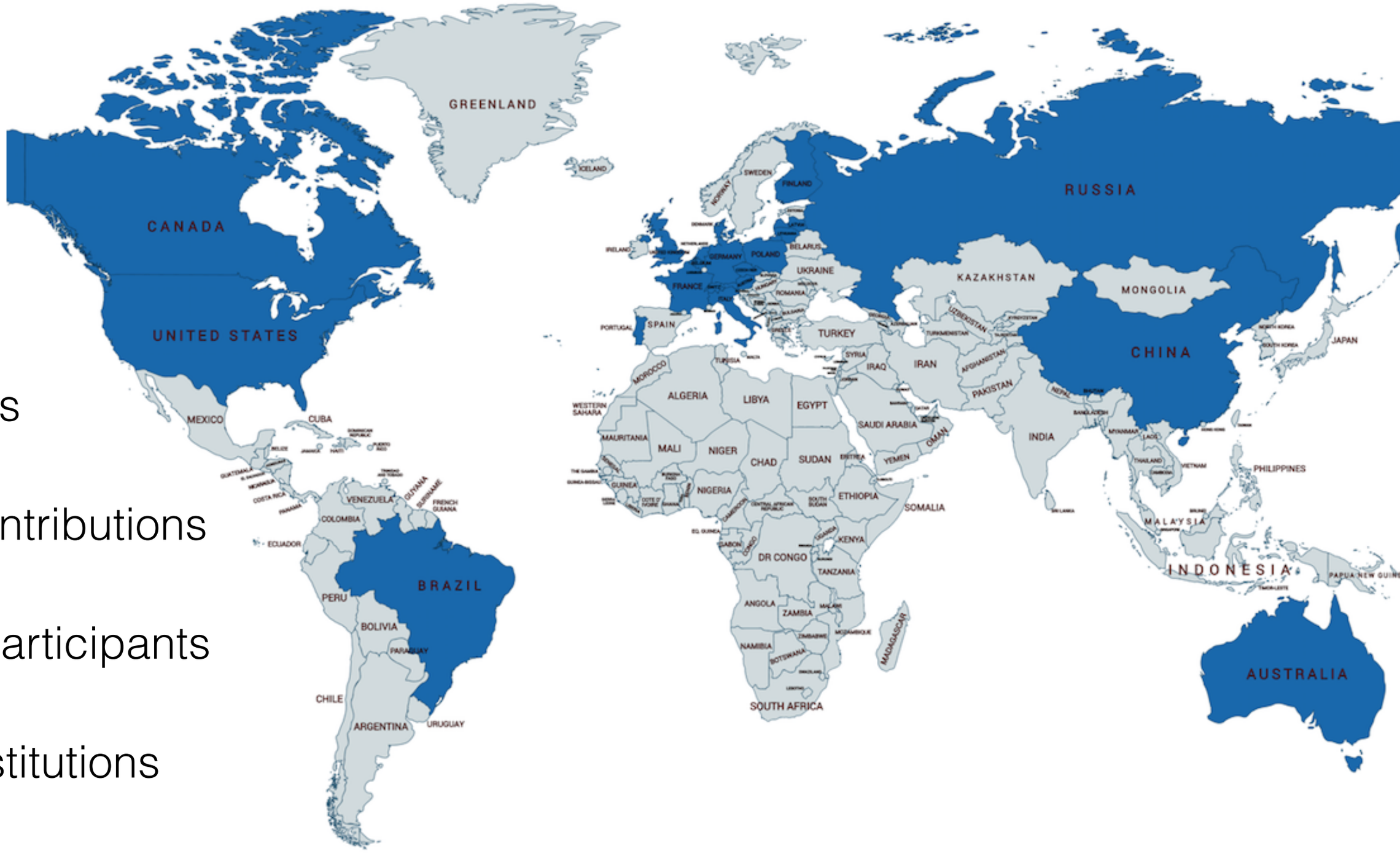
2017

CS3 2018

ONLYOFFICE
CSC – IT CENTER FOR SCIENCE
HUMBOLDT-UNIVERSITÄT ZU BERLIN
cynny space
Politechnika Krakowska im. Tadeusza Kościuszki
CERN
CSC
pydio
PUT YOUR DATA IN ORBIT
PowerFolder
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
DESY
universität wien
GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
SWITCH
aarnet
CYFRONET
UNIVERSITEIT GENT
DDN STORAGE
INFORMATYKA
Akademia Górniczo-Hutnicza w Krakowie
PSNC
JRC
EUROPEAN COMMISSION
cybera
GÉANT
Polcom
cesnet
SURF SARA
U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO
Collabora Online
ProtonMail
Nextcloud
egi
Max Planck digital library
Politechnika Wrocławska
DeiC
renci
LABORATORI NAZIONALI DEL GRAN SASSO
Seafile
ONEDATA
ETH zürich
INFN
RNP REDE NACIONAL DE ENSINO E PESQUISA
DONDERS INSTITUTE
Cubbit Join the swarm
APTIV
cscs Centro Svizzero di Calcolo Scientifico Swiss National Supercomputing Centre
iRODS
PAWSEY supercomputing centre
BROOKHAVEN NATIONAL LABORATORY
ktu kauno technologijos universitetas 1922
360 ICT, gegarandeerd betrouwbaar
ownCloud

CS3 2018 — Worldwide Community

- 3 days
- 55 contributions
- 120 participants
- 47 institutions
- 21 countries



1 track: 1 main room

networking area

splinter meetings

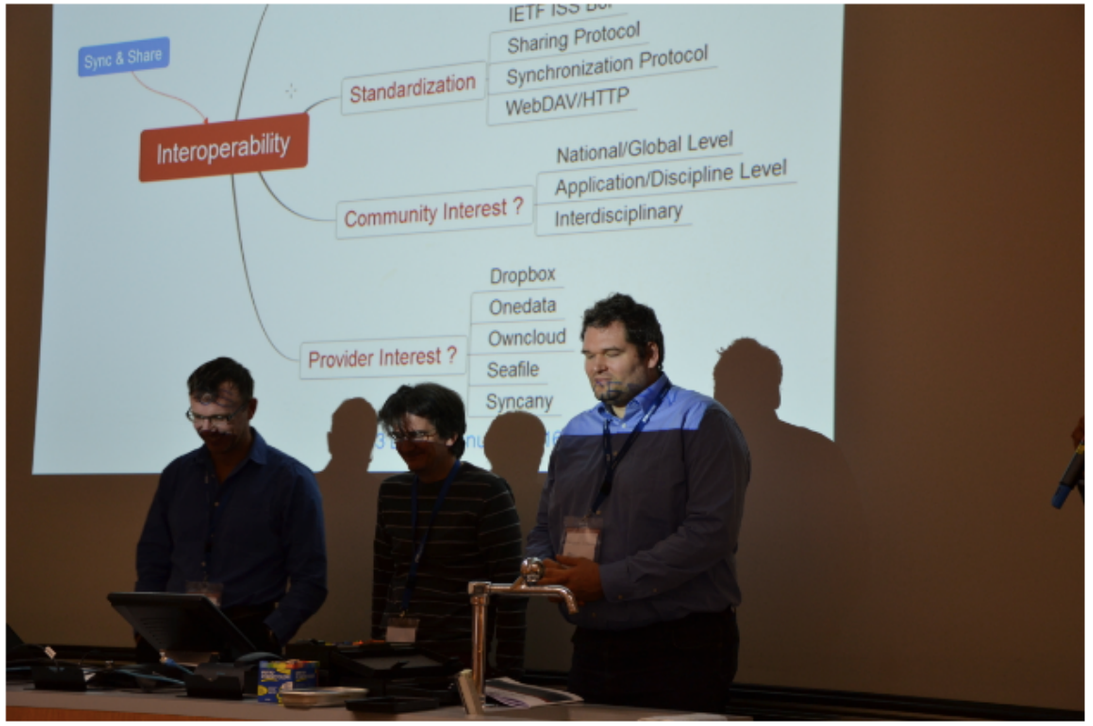
breaks

Inclusive community

- multiple initiatives and projects joining, *some examples*
  - **sync/share federation**
    - CS3 became the “home” of Open Cloud Mesh (federated sharing)
  - **integrated user environments for Data Science**
    - bottom-up introduction of Jupyter notebooks via user/application session
    - connection to compute and interplay with open source cloud stacks (“CISS”)
  - **interaction with HPC**

Inclusive community

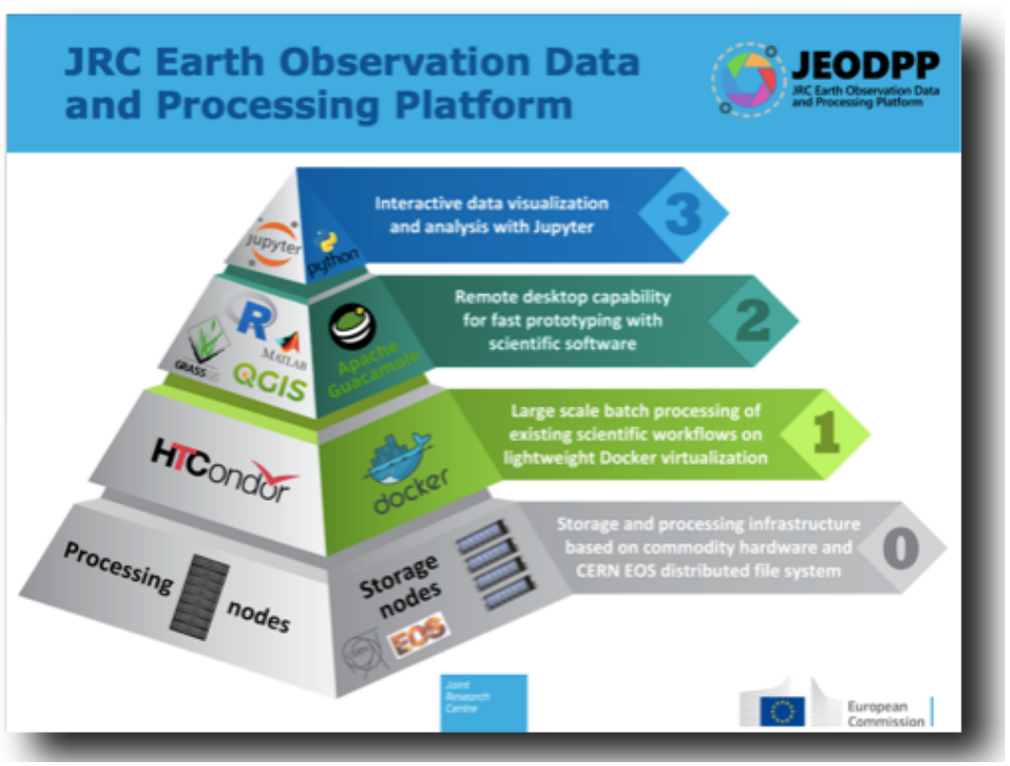
- multiple initiatives and projects joining, *some examples*
  - **sync/share federation**
    - CS3 became the “home” of Open Cloud Mesh

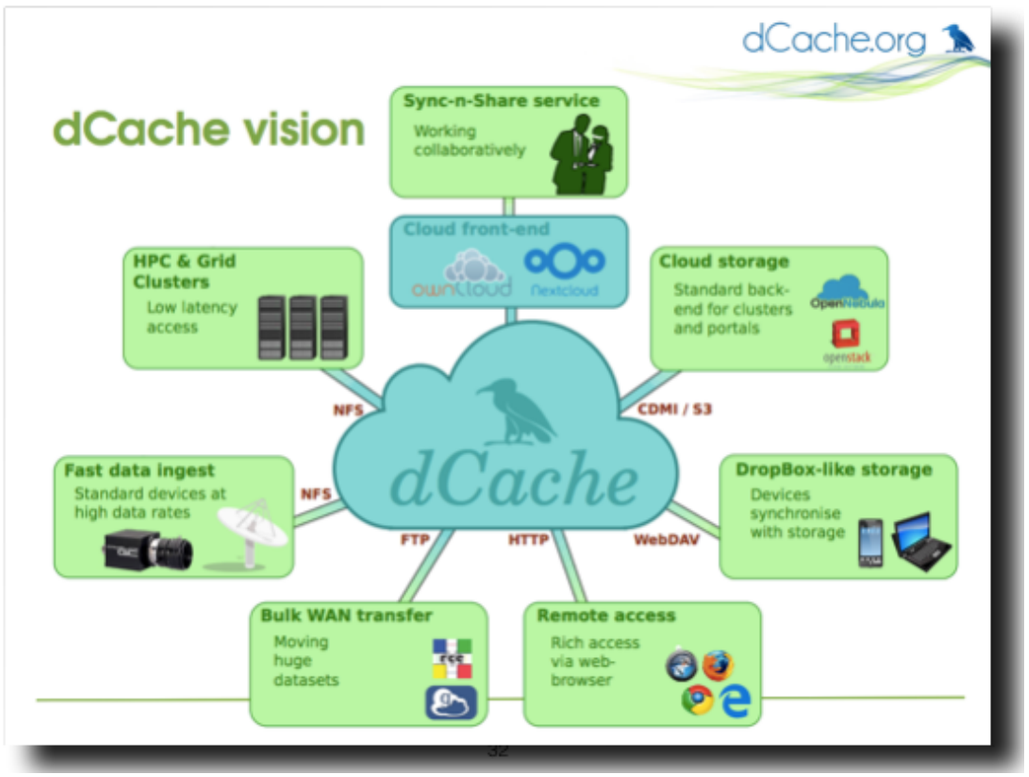


Praying for OCM to work... **ETHZ 2016**

Inclusive community (2)

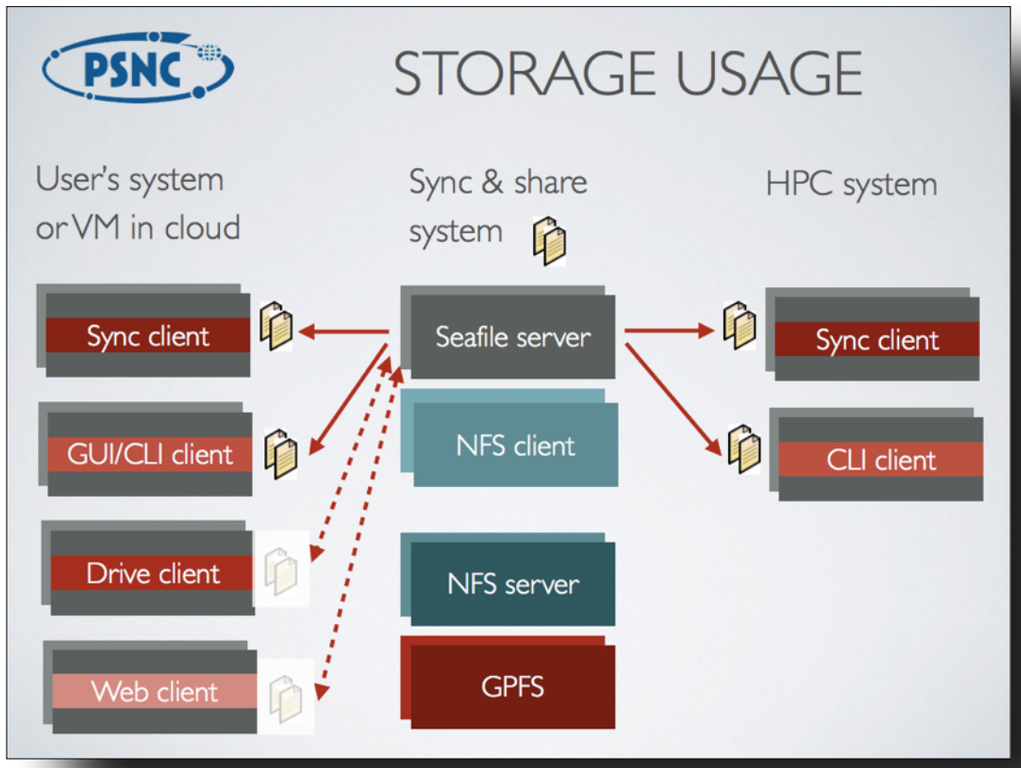
- multiple initiatives and projects joining, *some examples*
  - **integrated user environments for Data Science**
    - bottom-up introduction of Jupyter notebooks via user/application session
    - connection to compute and interplay with open source cloud stacks (CISS session)

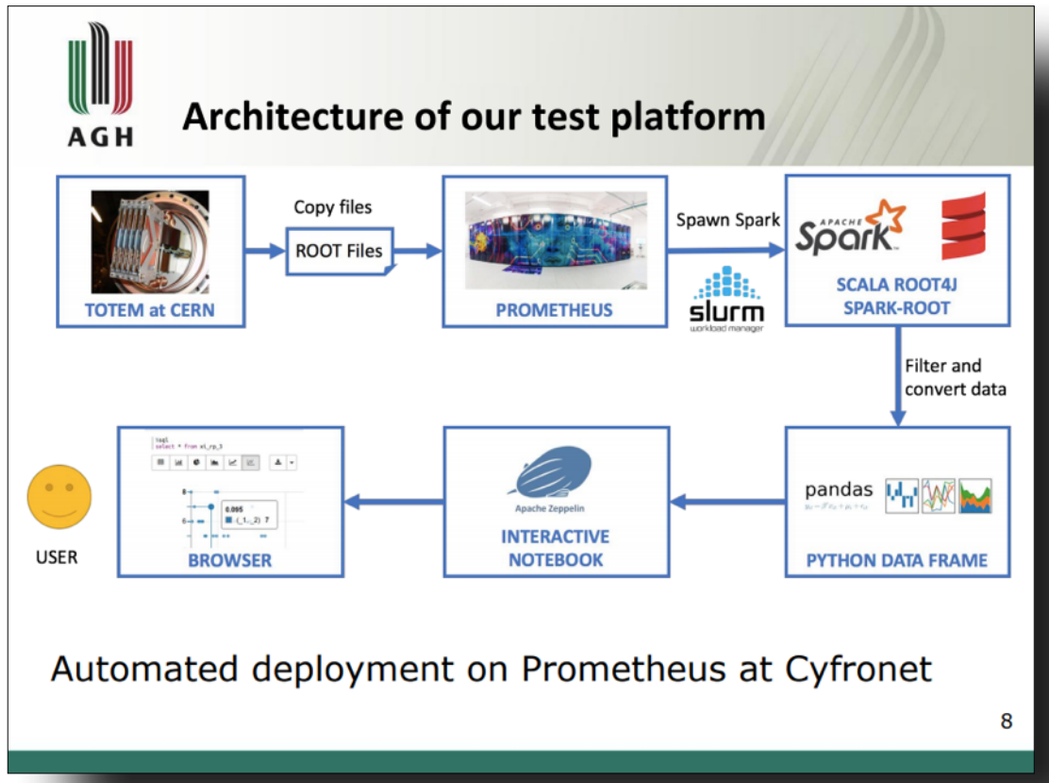




Inclusive community (3)

- multiple initiatives and projects joining, *some examples*
  - **interaction with HPC**





Commercial interest: keeping high

- Last year included new market segment with collaborative tools (OnlyOffice, Collabora)

- 2019 may see some big names again
  - interesting talks in progress!

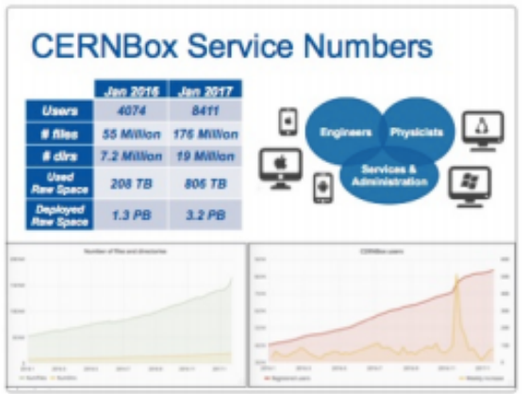
- We do invite inspiring **technical** talks from commercial partners

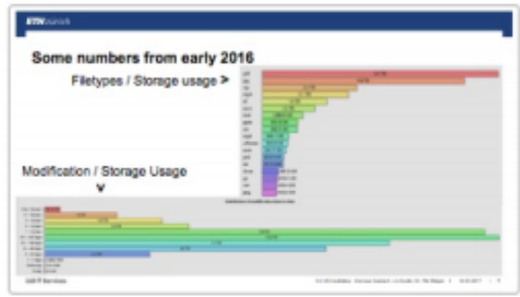
- We don’t invite sales talks

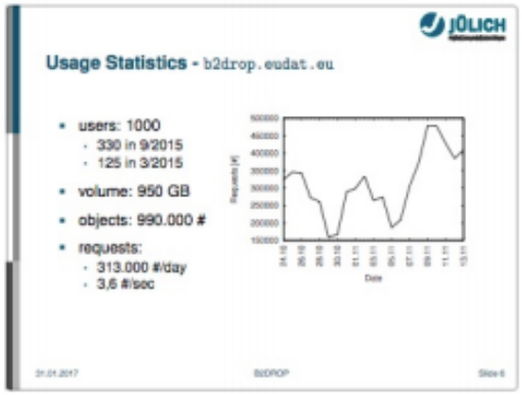
Community building & Site reports

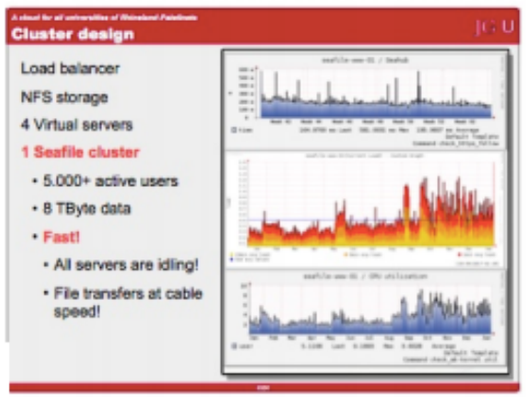
ETH zürich

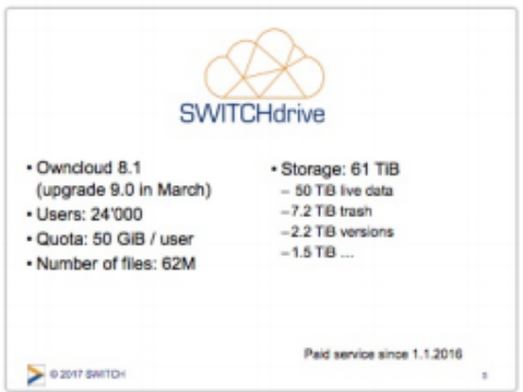
Summary of site report session – Tue. Jan. 31st

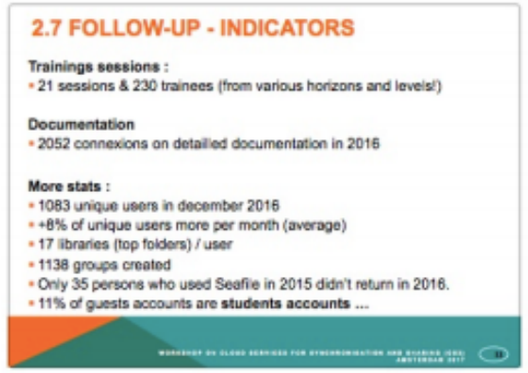


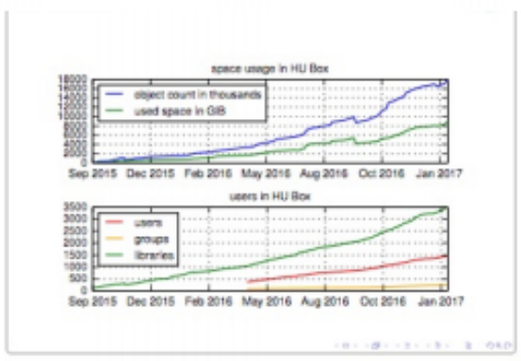


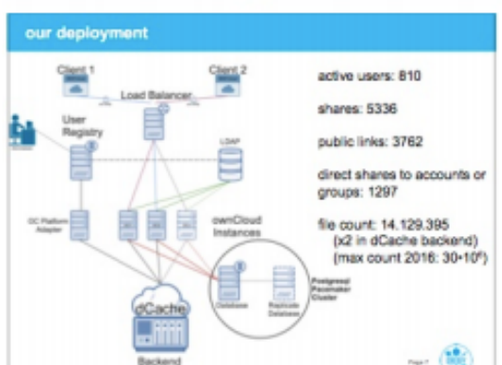


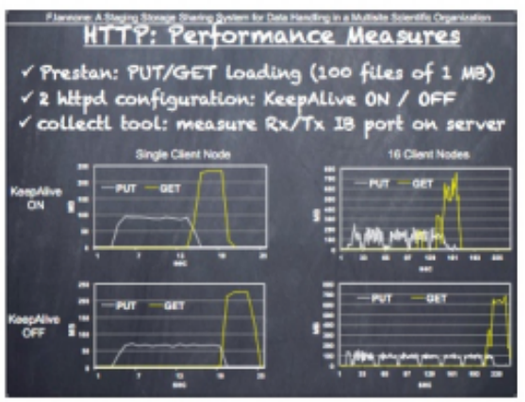












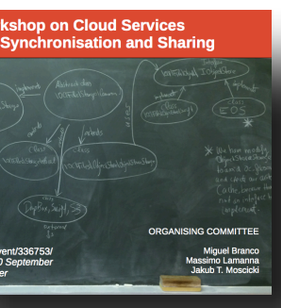
site report survey summary

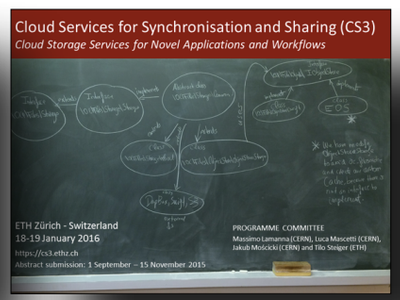
ETH zürich

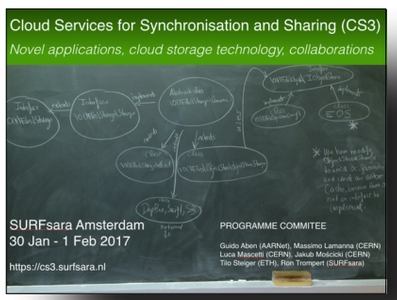
CS3 2018

Summed up user accounts:

owncloud : 266646
nextcloud : 24654
seafile : 15246
powerfolder : 87737
openstack swift : 1400







Discussions last year

- **Future** panel: how to face public clouds?
  - From “raw” storage to integrated application-aware storage
  - Proximity to the end users and applications
  - Consolidation as a community, federated collaboration
- EU recognition and role in EOSC
- Increasing role of Open Data => impact of FAIR

Sessions 2019

- **https://indico.cern.ch/event/726040/program**
- **Cloud infrastructure and software stacks for data science**
  - integration of CS3 services in modern cloud infrastructure and software stacks (CISS).
  - uniform environments shared across different researchers
  - access to computing facilities or workflow engine
    - **batch facilities, OpenStack&Container services, Spark clusters, Cloud-based resources, GPU hardware**
- **Open Data Ecosystems and CS3 (NEW)**
  - coordinated system between CS3 and OpenData services: **RDA, ORCiD, DataCite, GO-FAIR, the Open Science Foundation**
  - where a CS3 type service acts as the live data fulcrum

Cloud Services for Synchronisation and Sharing

28 - 30 January 2019, Roma

CS3

cs3.infn.it

International Programme Committee

Guido Aben	(AARNet)
Łukasz Dutka	(Cyfronet)
Fabio Farina	(GARR)
Agostino Funel	(ENEA)
Massimo Lamanna	(CERN)
Jakub T. Mościcki	(CERN)
Tilo Steiger	(ETH)
Stefano Stalio	(INFN)
Ron Trompert	(SURFsara)

Local Organising Committee

Francesco Failla	(INFN Roma)
Piero Gatta	(INFN Roma)
Claudio Grandi	(INFN Bologna)
Mauro Mancini	(INFN Roma)
Enrico Pasqualucci	(INFN Roma)
Stefano Stalio	(INFN LNGS)

Sponsors: INFN, ENEA, Consortium GARR

Organizers: CYFRONET, aarnet, CERN, ETH zürich, SURF SARA

Parnters: C-O, cynny space, DDN STORAGE, Nextcloud, ONEDATA, ONLYOFFICE, ownCloud, pydio, Seafile

Cloud Services for Synchronisation and Sharing

28 - 30 January 2019, Roma

CS3

cs3.infn.it

International Programme Committee

Guido Aben	(AARNet)
Łukasz Dutka	(Cyfronet)
Fabio Farina	(GARR)
Agostino Funel	(ENEA)
Massimo Lamanna	(CERN)
Jakub T. Mościcki	(CERN)
Tilo Steiger	(ETH)

Local Organising Committee

Francesco Failla	(INFN Roma)
Piero Gatta	(INFN Roma)
Claudio Grandi	(INFN Bologna)
	(INFN Roma)
ucci	(INFN Roma)
	(INFN LNGS)

Registration already open

Abstract submission deadline — end of November

https://cs3.infn.it

Sponsors: INFN, ENEA, Consortium GARR

Organizers: CYFRONET, aarnet, CERN, ETH zürich, SURF SARA

Parnters: CO, cynny space, DDN STORAGE, Nextcloud, ONEDATA, ONLYOFFICE, ownCloud, pydio, Seafile

INFRA-EOSC-02

Can CS3 help?

- It seems to make sense to capitalise on existing CS3 community
  - successful track record
  - exploitation plan already there: services exists and are used
  - community is large and varied (beyond HEP, beyond NRENs,…)
  - CS3 sustained itself for 5 years via bottom-up interest of participants
  - open, inclusive environment already proven => a necessary component for wide adoption of the outcomes of the INFRA-EOSC-02 call

Service areas?

- Sync/Share services as a first-class citizen in data science
    - federated FAIR-compliant infrastructure for sharing open data with easy access from end-user devices
    - “federation” of trust and quality
    - bridge to compute and integrated user environments
    - infrastructure-agnostic service-layer
        - flexible deployment on premise or in hybrid clouds

Would help to further evolve CS3 community itself…

- For users/sites: FAIR-compatible by definition
- For companies: new paths to market solutions
- …

Considerations (1)

- Must benefit everyone in the CS3 community
  - consortium must be relatively small number of parters to be efficient
  - but initial nucleus of partners is just to lay out first few bricks
    - everyone should have a possibility to adopt and contribute as early as possible
- Adopt similar principles as commercial-grade services: added value for the users
  - ease of access
  - match user needs
  - QoS

Considerations (2)

- Very short time to prepare the proposal
- What would it produce, provide and actually do?
- Relation to other ongoing projects
  - OCRE, EUDAT,…
- Long-term integrated into EOSC Catalog/Hub