

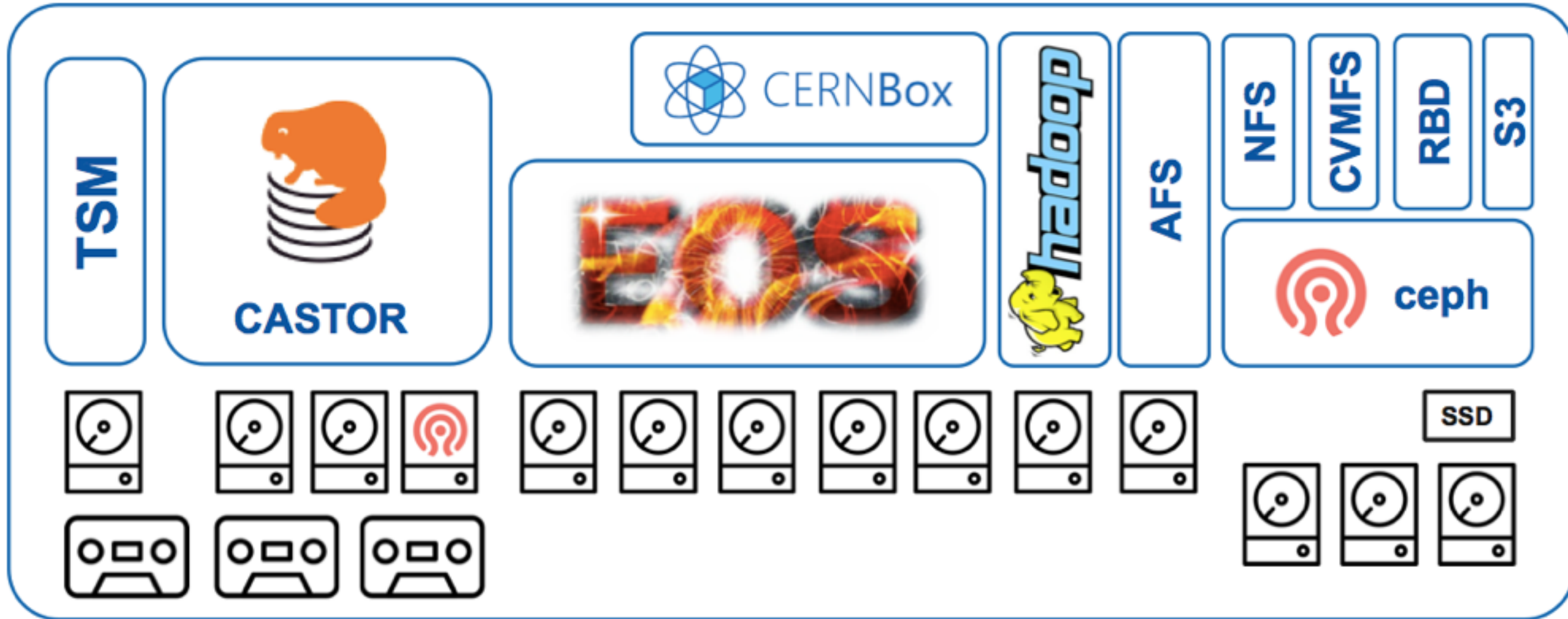
EOS/CERNBox for data analysis

Jakub T. Moscicki
CERN IT-ST

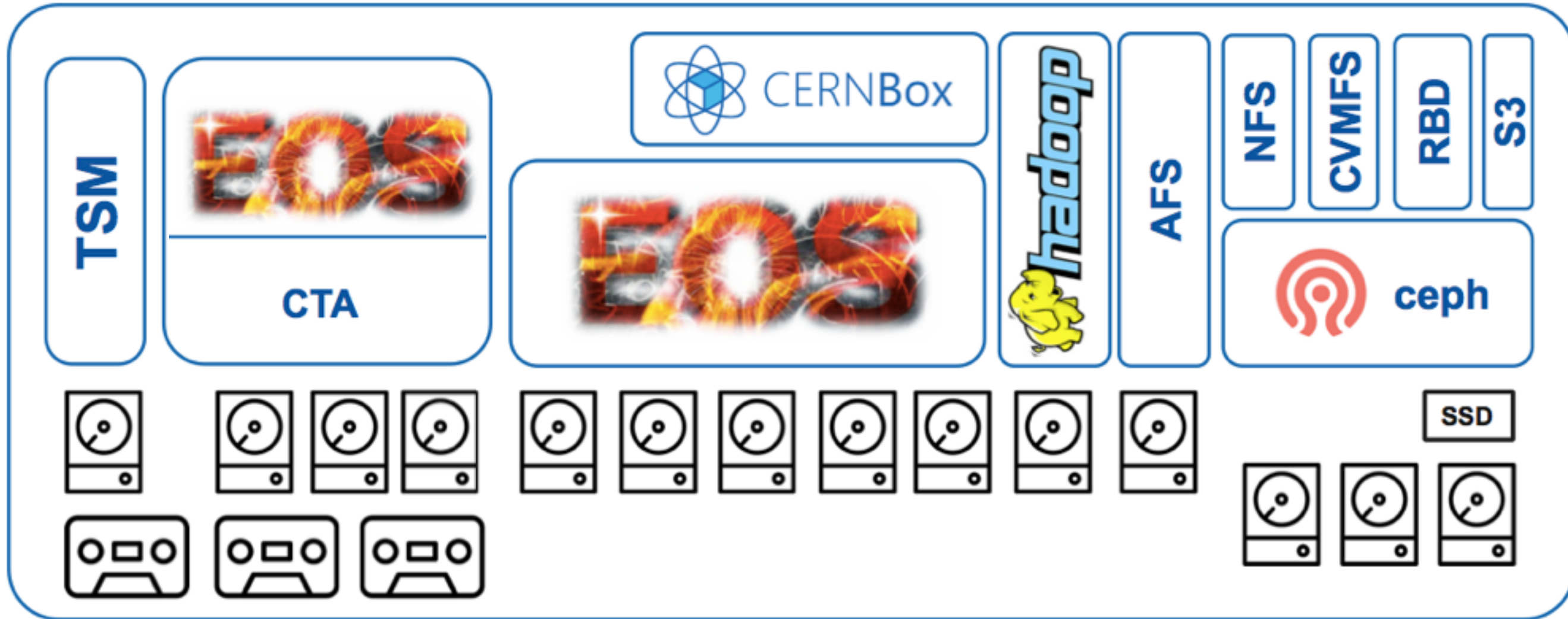
(on behalf of CERNBox and EOS teams)

9th LHCb Computing Week, May 2017
Data Preservation Track

Storage Services



Storage Services



EOS @ CERN

- Disk storage for all LHC/physics data and now also user files (CERNBox)



EOS Instances



Public

AMS
COMPASS
NA61/NA62
Others

CERNBox



CERNBox data: **560 TiB**

CERNBox #files: **229 Mil**

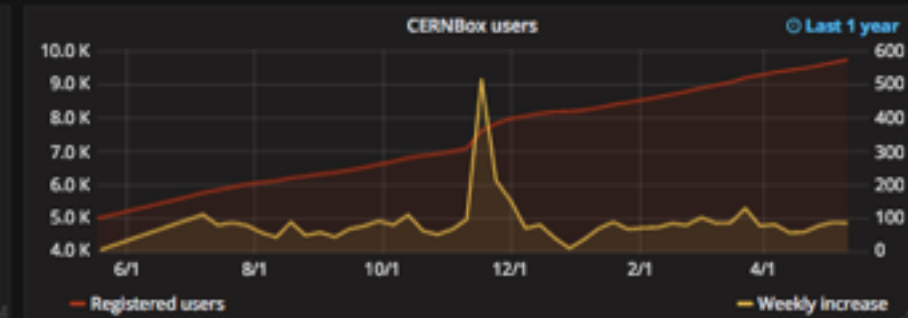
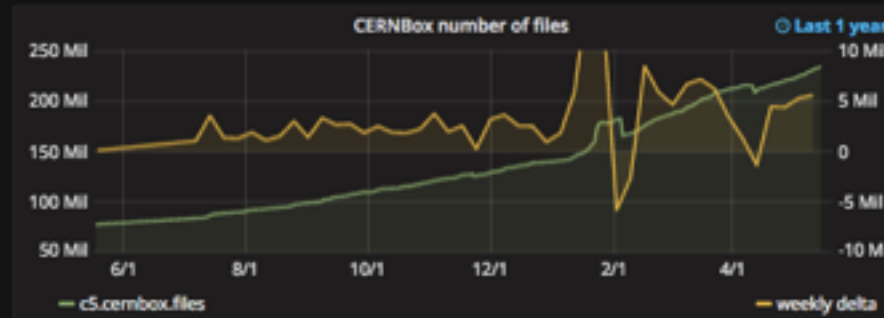
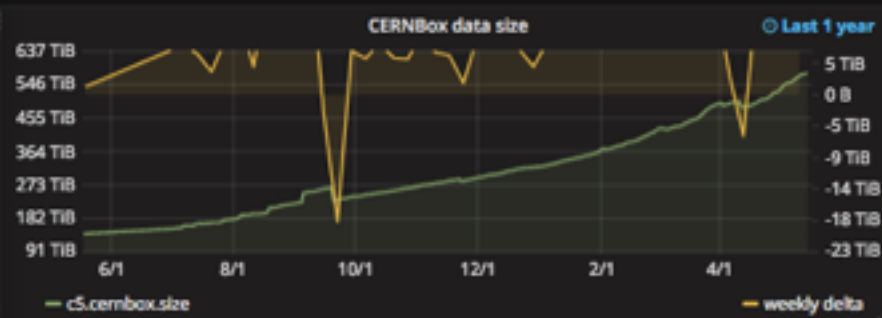
CERNBox #users: **9672**

Δ data **19.00 TiB** (1 week)

Δ #files: **5568962** (1 week)

Δ(#users): **88** (1 week)

▼ CERNBox additional info



▼ EOS

ALICE

ATLAS

CMS

LHCb

PUBLIC

USER

#files **303 Mil**

#files **114.6 Mil**

#files **51.3 Mil**

#files **17.5 Mil**

#files **90.9 Mil**

#files **98.7 Mil**

Δfiles: **-789 K** (1 week)

Δfiles: **-520 K** (1 week)

Δfiles: **734 K** (1 week)

Δfiles: **-857 K** (1 week)

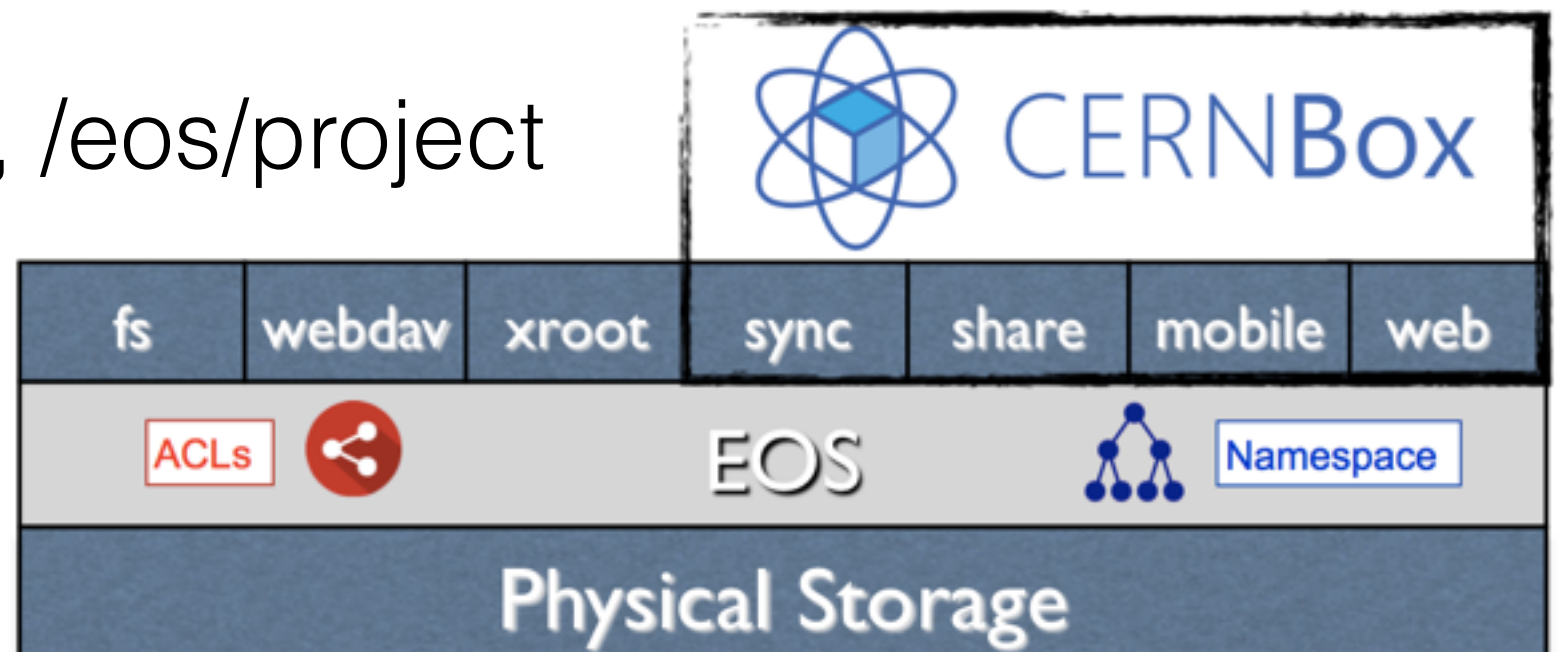
Δfiles: **2.420 Mil** (1 week)

Δfiles: **6.40 Mil** (1 week)

CERNBox

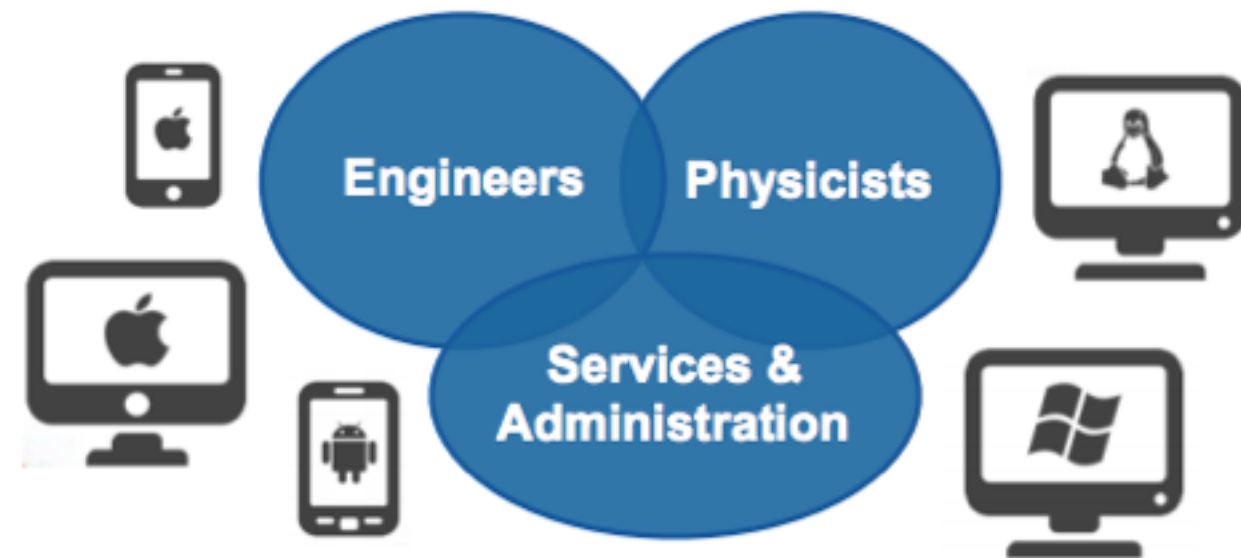


- Makes it very easy to get in and out small datasets from EOS, offline work, share/collaborate with peers, independently of your OS (and even mobile)
 - “Dropbox for Science”
- You can get hold of your data also online
 - Ixplus: /eos/user, /eos/project

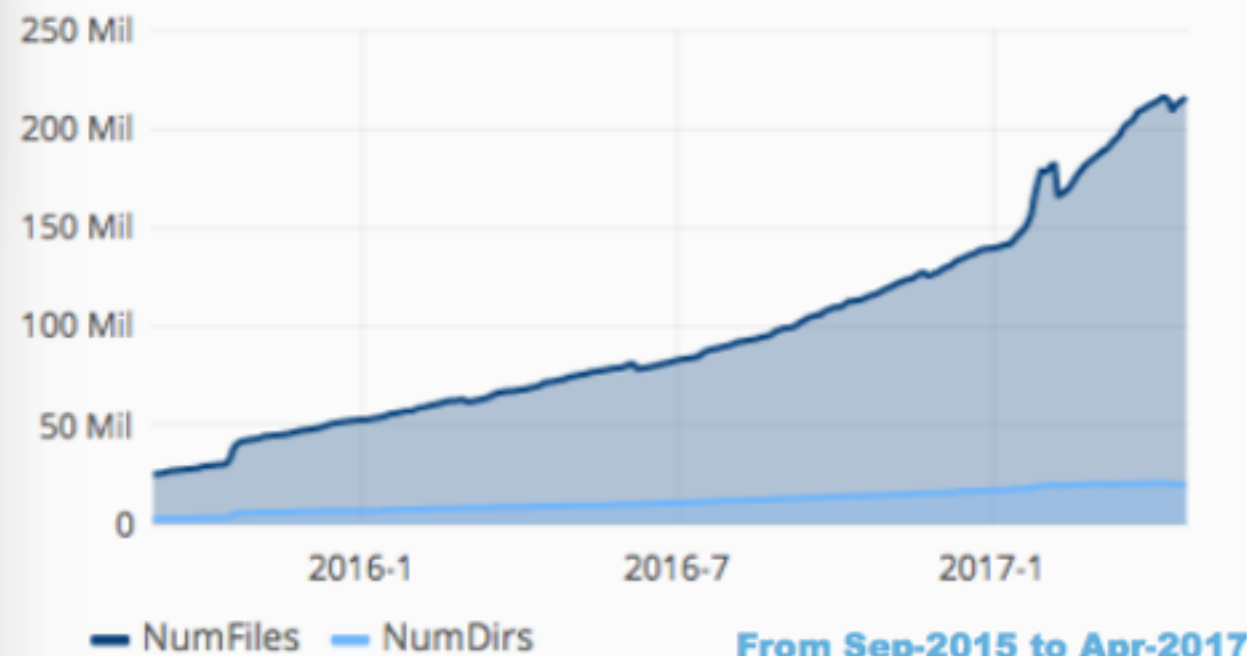


CERNBox Service Numbers

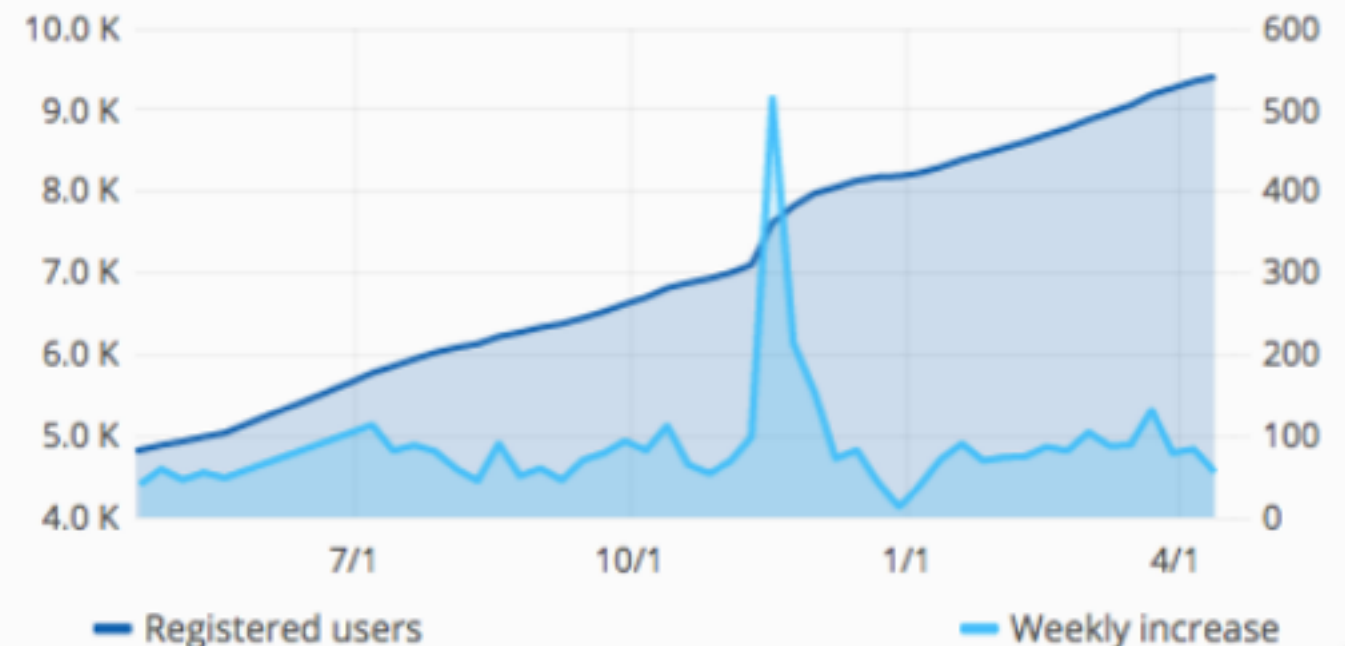
	Jan 2016	Jan 2017	Apr 2017
Users	4074	8411	9501
# files	55 Mil	176 Mil	218 Mil
# dirs	7.2 Mil	19 Mil	20.6 Mil
Used Raw	208 TB	806 TB	1110 TB
Deployed Raw	1.3 PB	3.2 PB	3.3 PB



Number of files and directories

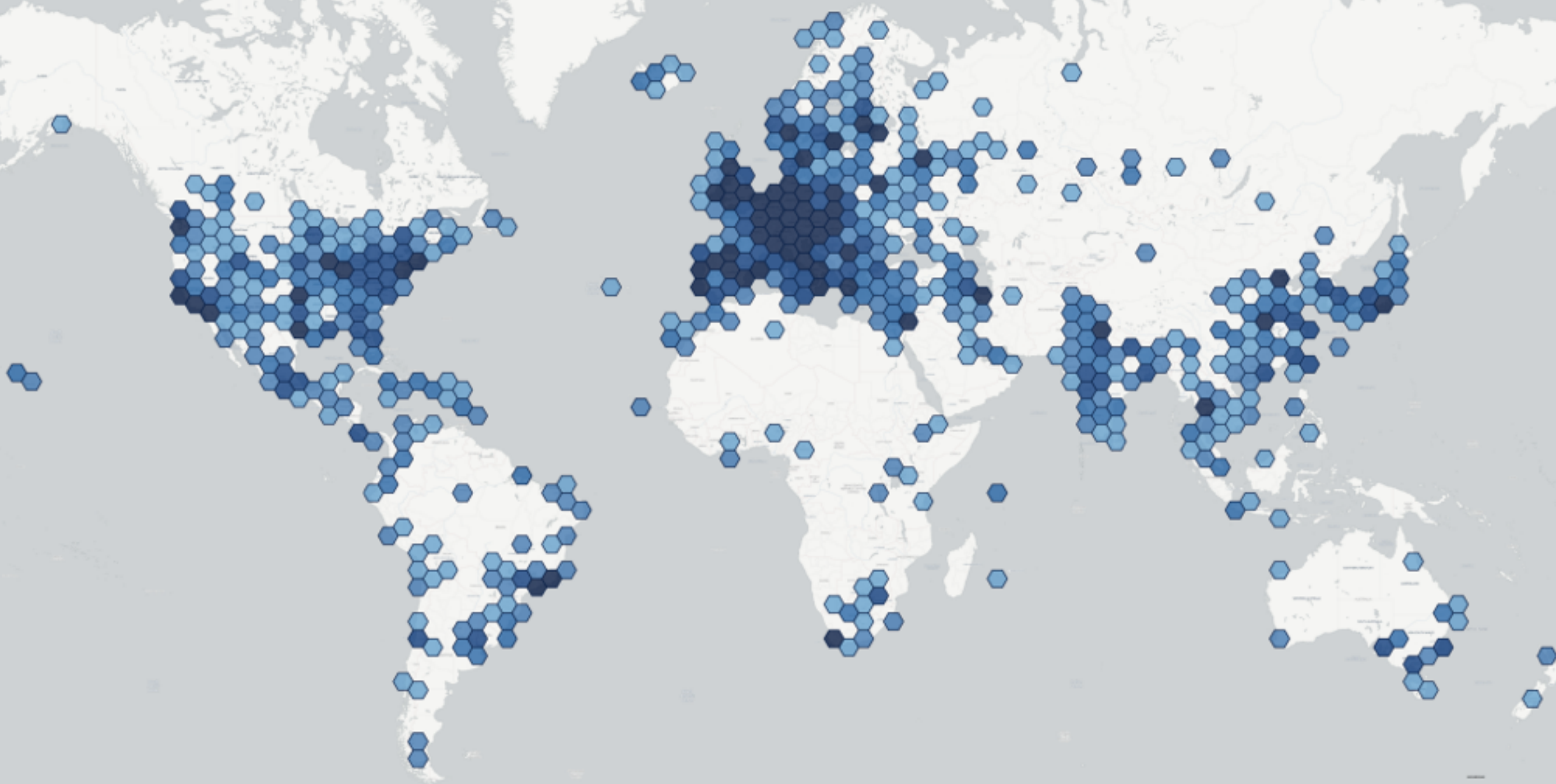


CERNBox users



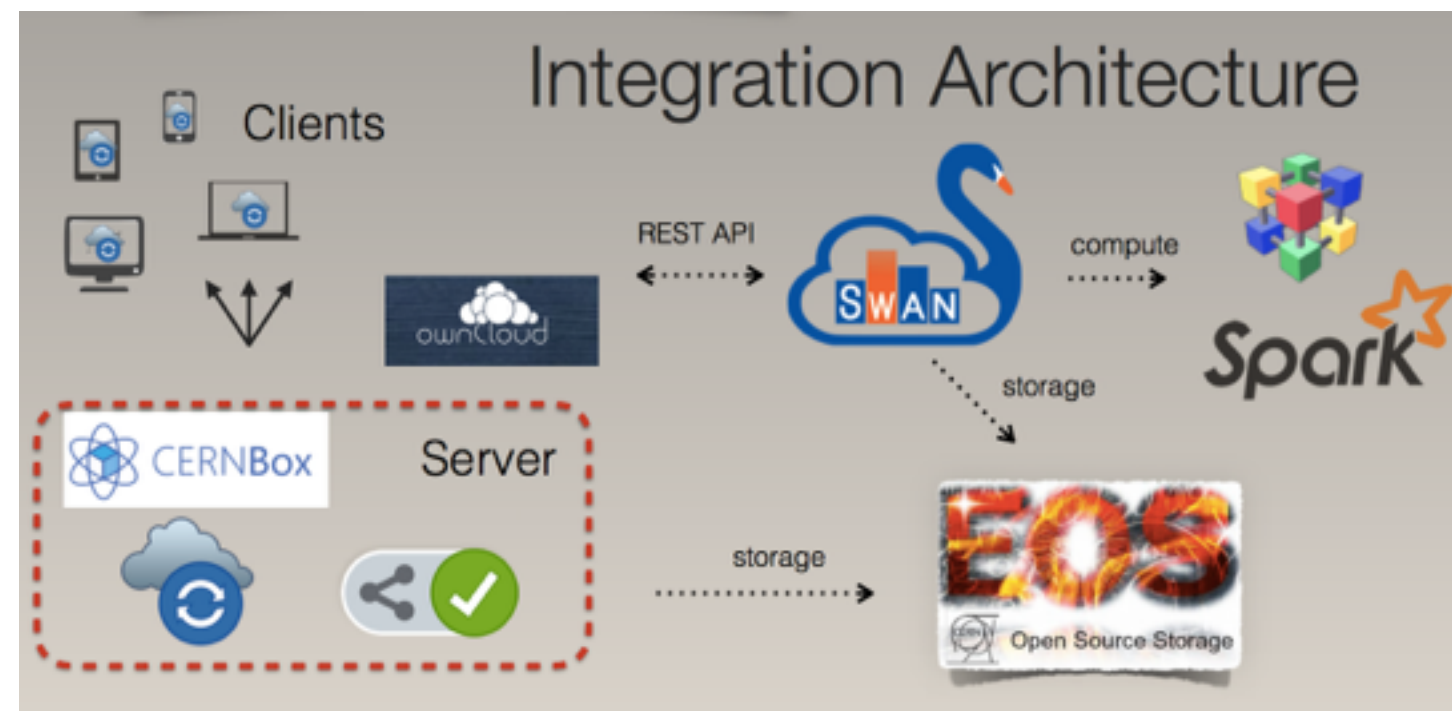


CERNBox



Storage for Jupyter Notebooks and Data Files

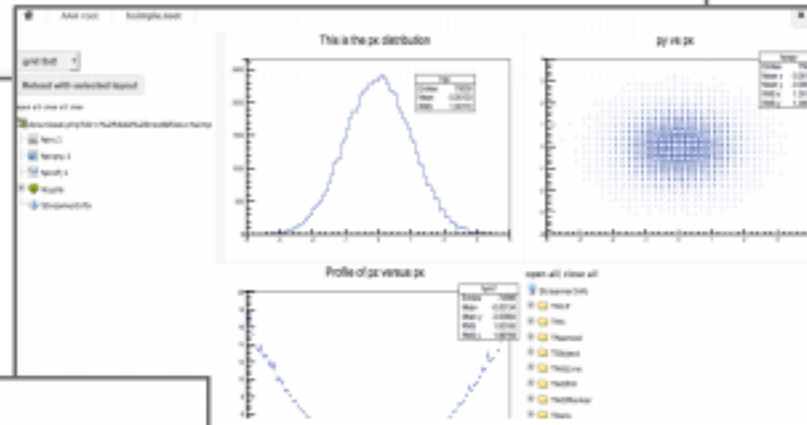
- eosuser is a storage backend for swan.cern.ch
 - CERNBox sync/share features available
- Support for end-user analysis
 - Open in SWAN
 - Visualisation of ROOT files
 - Share Notebooks and Code



Name	Size	Modified
CERNBox	0 kB	3 minutes ago
CERN visit.doc	98 kB	a year ago
CERN-2016-EN.pdf	204 kB	a year ago
Machine Retirement.xls	16 kB	6 months ago
Storage-Discussion.ppt		

Excel Online - WOPerc-int...
 https://oos.cern.ch/...
 Excel Online
 Edit in Browser | Print | Data | Find

PowerPoint Online
 https://oos.cern.ch/...
 FILE | HOME | INSERT | DESIGN | TRANSITIONS | ANIMATIONS | VIEW
 Welcome to OWA
 Try to test Test Test



Open in SWAN

LHC Page1

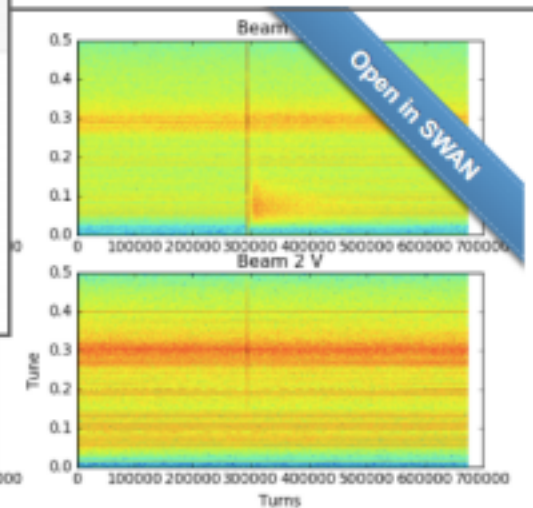
Plotting the data

We can now build a plot of the intensity and beams energy starting from the data we retrieved. The `pyROOT` library can be used to generate ROOT based data visualizations.

```

In [1]: plot.figure(figsize=(10,10))
fig, ax = plt.subplots(2)
ax[0].plot(intensity)
ax[0].set_xlabel('Time')
ax[0].set_ylabel('Intensity')
ax[1].plot(energy)
ax[1].set_xlabel('Time')
ax[1].set_ylabel('Energy [GeV]')
plt.show()
  
```

Intensity vs Time (2010-06-25 to 2010-06-26)
 Energy [GeV] vs Time (2010-06-25 to 2010-06-26)



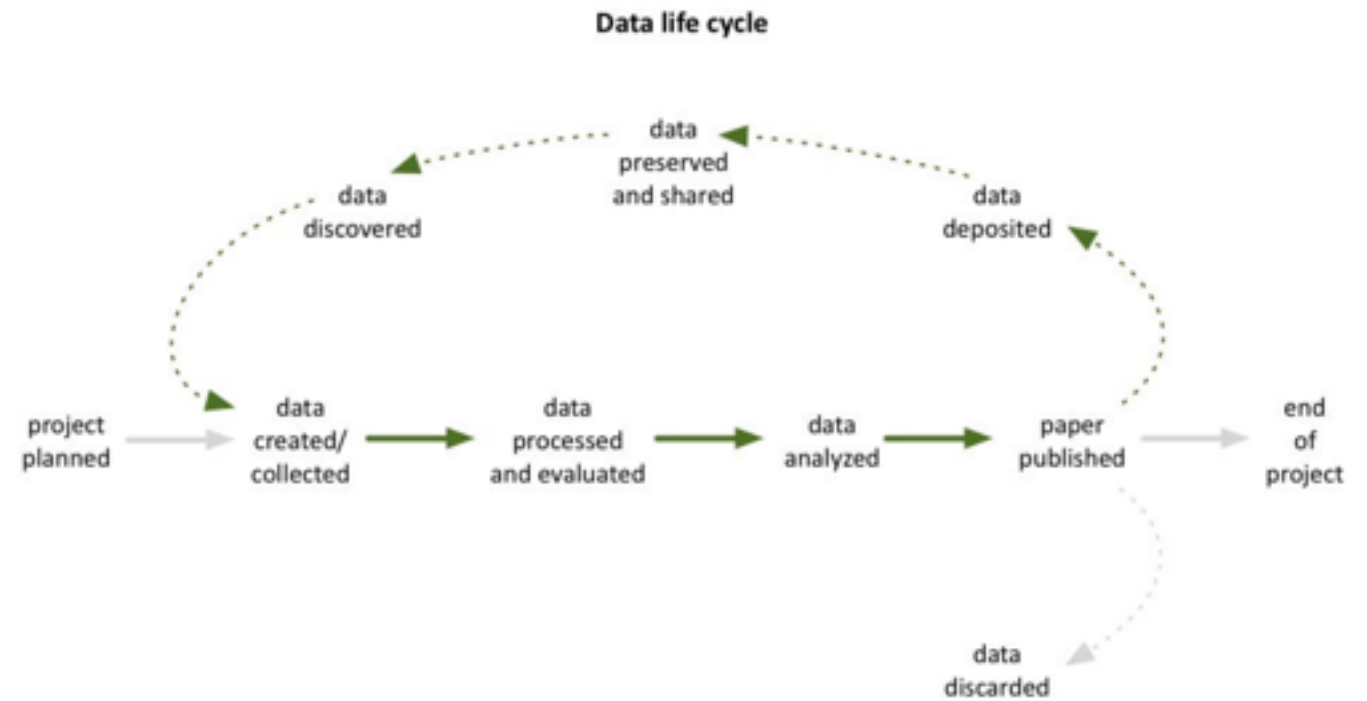
Open in SWAN

- Integration with multiple services
- ROOTJS
- Jupiter notebooks (SWAN)
- Microsoft Office 365
- and others to come...



Publishing to ORCID and Zenodo

- Developed in Community CS3 Workshops
<https://cs3.surfsara.nl>



Places

- Home
- Shared by me
- Shared with me
- Trashbin

Apps

- Metadata
- Pictures
- Groups
- Importer
- Uploader

documents

Upload New

Name	Size	Modified
example.odt DUBLIN_CORE_ORIG +1 more	23 kB	9 days ago

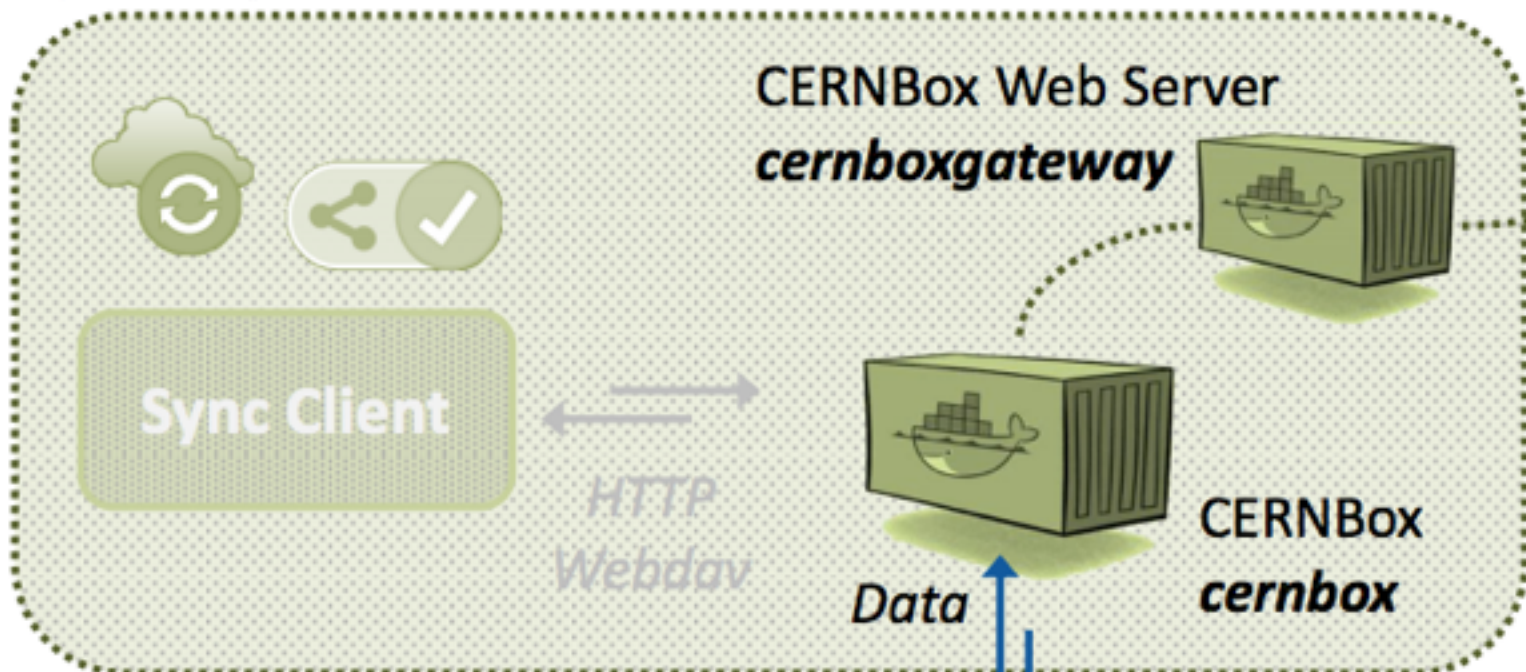
1 file 23 kB

Deposit to Zenodo

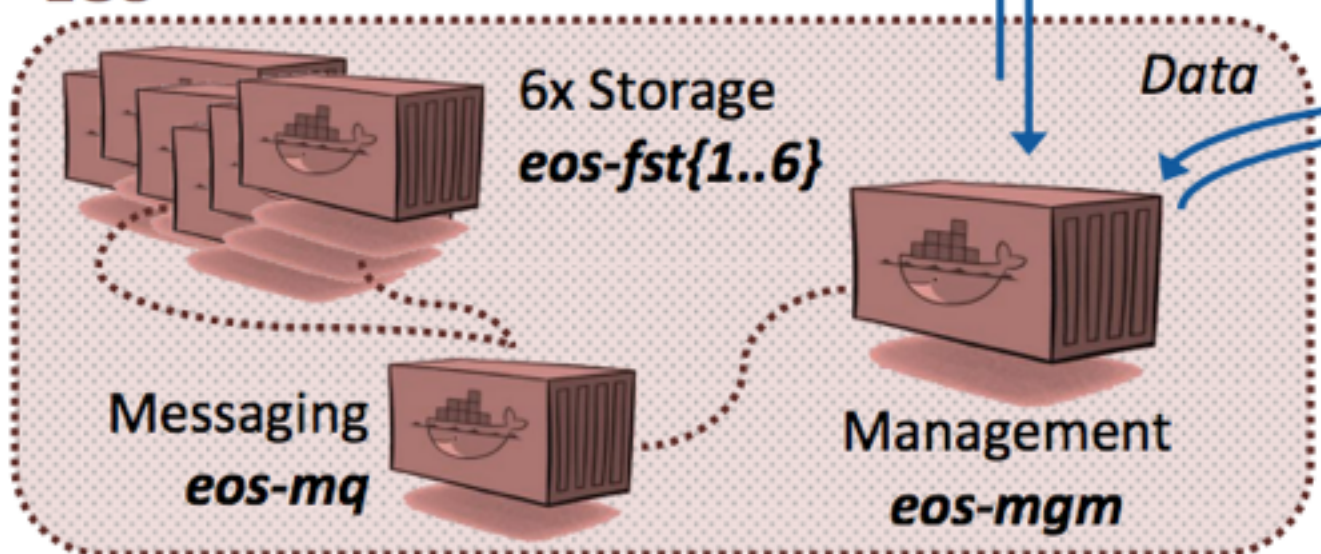
Docker-based Service – Architecture



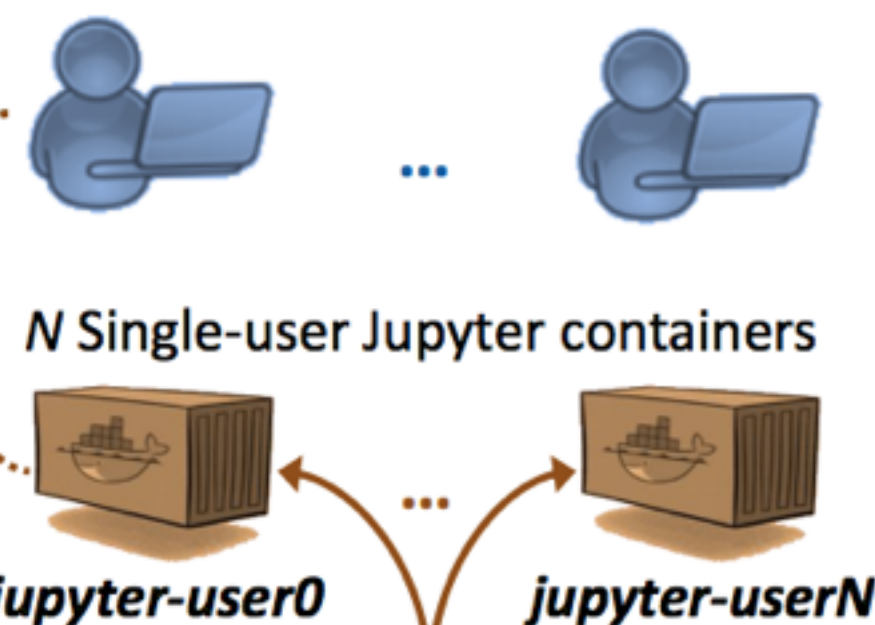
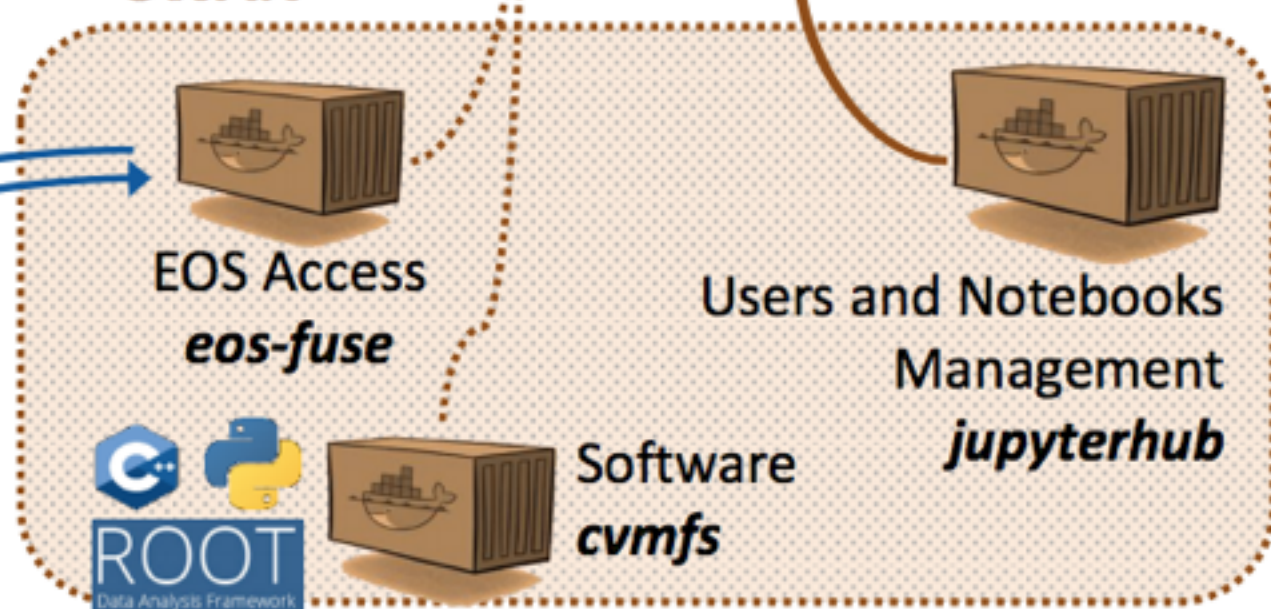
CERNBox



EOS



SWAN



Managing user credentials in containers

- A container needs have access to a mount point /eos/...
- A container must be privileged to mount EOS fuse
- A container needs krb5 ticket to authenticate
- The access may be via the web interface (SWAN)
 - SSO -> krb5 ???
- Possible mitigations:
 - Constrained delegation

Need to manage access credentials

See talk by Luca&Massimo

Talk back

- What are the questions from Data Preservation perspective for storage services we provide?