



FINAL EVENT

HYBRID EVENT

22 June 2023

11:00 - 17:45 CEST



Location: Poznan (Poland),
co-located with *EGI Conference*

Science Mesh - Unlocking Open Science & Collaborative Research Landscape



Marcin Sieprawski

Head of Big Data Lab
Software Mind



Ron Trompert

Senior Consultant
SURF



Enric Tejedor Saavedra

Software Engineer
CERN



Lennart Hofeditz

Research Associate
Universität Potsdam

Science Mesh Interoperable Research Workflows and Research Infrastructures

Room Paris | 11:20-12:15 CEST

[illegible]

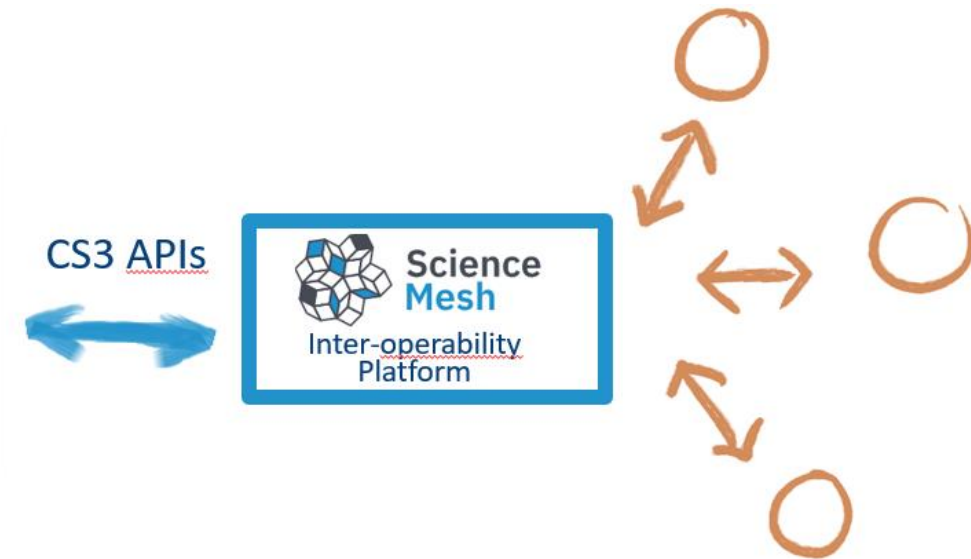
Marcin Sieprawski

Head of Big Data Lab

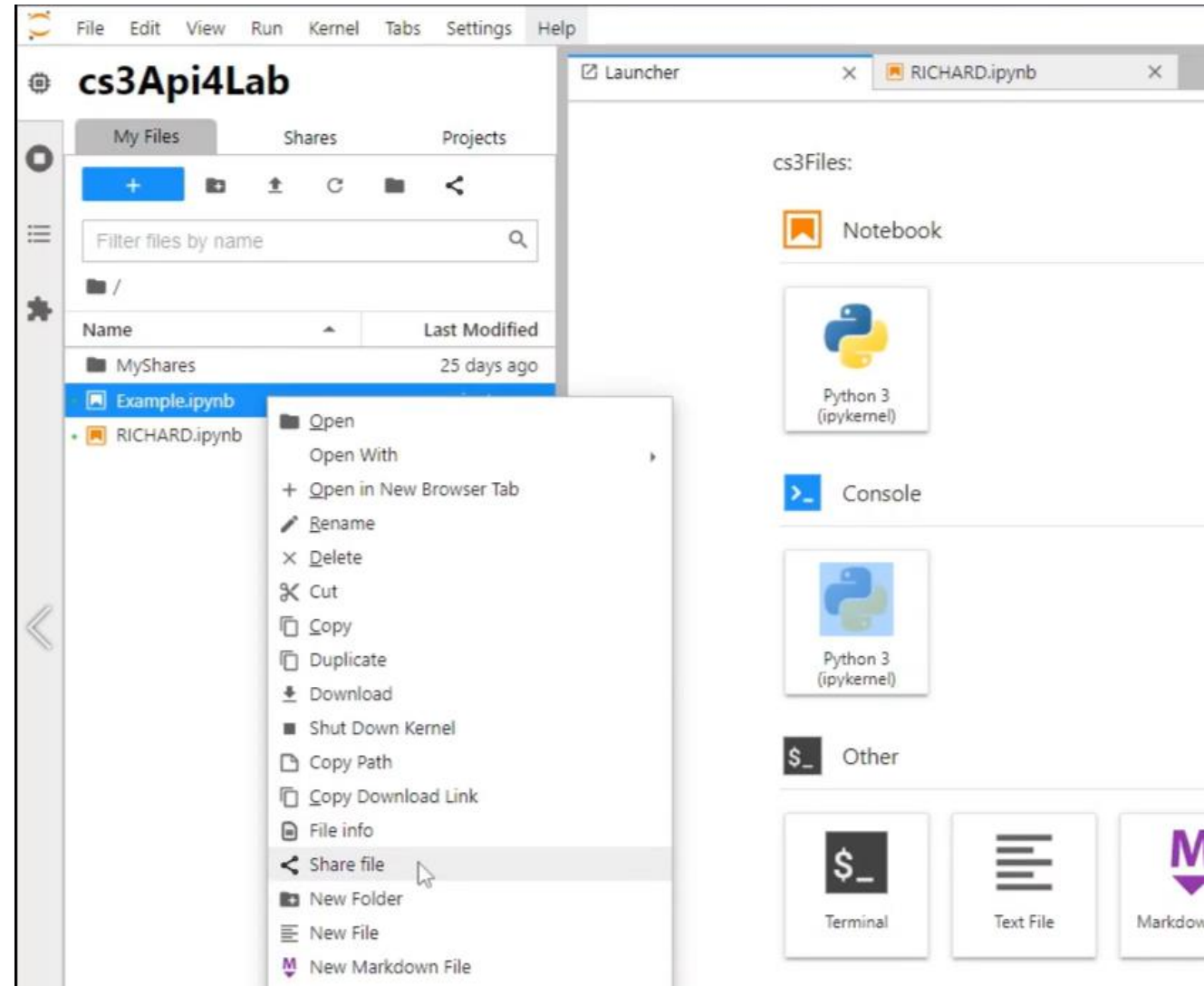


CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.

- # Collaboration on Jupyter notebooks in a sync-and-share environment
- # JupyterLab extension (**cs3api4lab**)
 - # Integration with ScienceMesh IOP (CS3 APIs)
 - # Sharing and collaboration features in JupyterLab
- # Unified shares interface (CS3 / OCM)
- # Collaborative workflows
 - # Locking mechanism
 - # Overwrite protection
 - # Prepared: merge support
- # Easy config



- # Sharing buttons
- # Entries in the context menus
- # SHARES tab
 - # Shared by me
 - # Shared with me
- # Sharing workflow
 - # Accepting shares
- # Pop-up windows:
 - # File information
 - Shares window

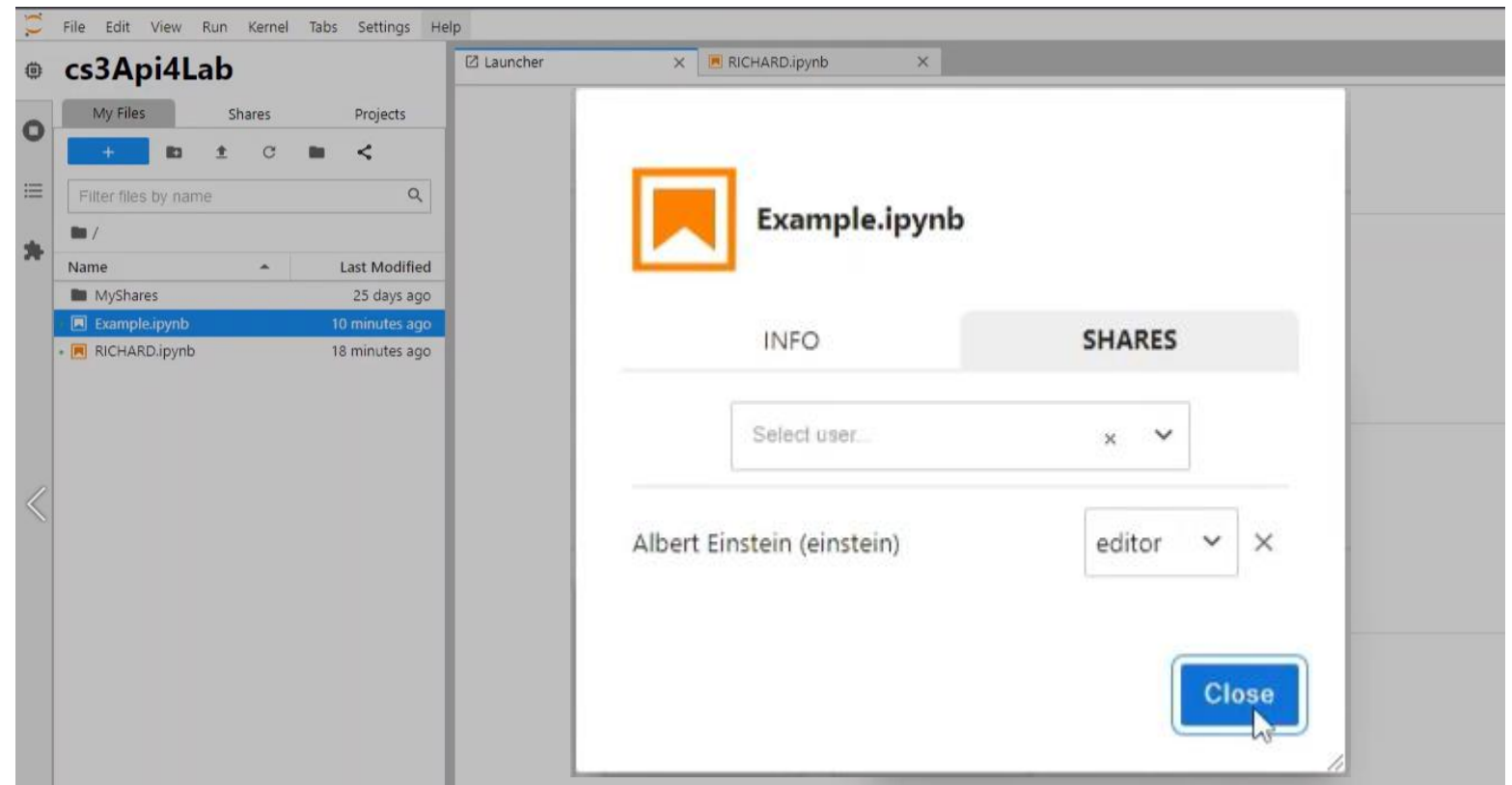


User search

- # Part of: **first name, last name**
- # Full: **username, email**

Role:

- # **Viewer**
- # **Editor**





The screenshot displays the **cs3Api4Lab** web application interface. The top navigation bar includes **File**, **Edit**, **View**, **Run**, **Kernel**, **Tabs**, **Settings**, and **Help**. The main interface is divided into two panels.

Left Panel (cs3Api4Lab):

- Navigation tabs: **My Files**, **Shares** (selected), **Projects**.
- Section: **Pending shares** / **Shared by me**.
- Search bar: **Filter files by name**.
- File list table:

Name	Last Modified
Example.ipynb	seconds ago

A tooltip for the selected file **Example.ipynb** displays the following details:

- Name: Example.ipynb
- Size: 228.8 KB
- Path: cs3driveShareByMe:reva/richard
- Created: 2023-03-03 18:15:47
- Modified: 2023-03-03 18:15:47
- Writable: false

Right Panel (Launcher):

- Tab: **RICHARD.ipynb**.
- Section: **cs3Files:**
- Available actions:

- Notebook
- Python 3 (ipykernel)
- Console
- Python 3 (ipykernel)
- Other

The **Other** section includes icons for:

- Terminal
- Text File
- Markdown File
- Show Contextual Help



The screenshot displays the **cs3Api4Lab** web interface. On the left, a sidebar shows file management options: 'My Files', 'Shares', and 'Projects'. Under 'Shares', there are sections for 'Pending shares' and 'Shared by me', each with a search bar and a table of files. The 'Shared with me' section at the bottom also has a search bar and a table. The main area shows a Jupyter notebook with three tabs: 'Launcher', 'Example.ipynb', and 'EINSTEIN.ipynb'. The 'Example.ipynb' tab is active, showing code cells. The first cell imports `matplotlib`, `numpy`, and `plt`. The second cell shows the output of a contextlib.ExitStack. The third cell sets a random seed and generates data. The fourth cell imports `Line2D` and creates custom lines. The fifth cell creates a subplot and plots the data. A modal dialog box is overlaid on the plot, stating 'This file is currently locked by another person' and offering two options: 'Stay in preview mode' and 'Create a copy'. The plot shows three lines: 'Cold' (blue), 'Medium' (grey), and 'Hot' (red), with the y-axis ranging from 15 to 20.

```
[3]: from matplotlib import rcParams,ycler
import matplotlib.pyplot as plt
import numpy as np
plt.ion()

[3]: <contextlib.ExitStack at 0x2a537b51460>

[4]: np.random.seed(19680801)

N = 10
data = [np.logspace(0, 1, 100) + np.random.randn(100) + ii for ii in range(N)]
data = np.array(data).T
cmap = plt.cm.coolwarm
rcParams['axes.prop_cycle'] =ycler(color=cmap(np.linspace(0, 1, N)))

from matplotlib.lines import Line2D
custom_lines = [Line2D([0], [0], color=cmap(0.),
                      Line2D([0], [0], color=cmap(.5),
                      Line2D([0], [0], color=cmap(1.),

fig, ax = plt.subplots(figsize=(10, 5))
lines = ax.plot(data)
ax.legend(custom_lines, ['Cold', 'Medium', 'Hot'])
```

This file is currently locked by another person

Stay in preview mode Create a copy

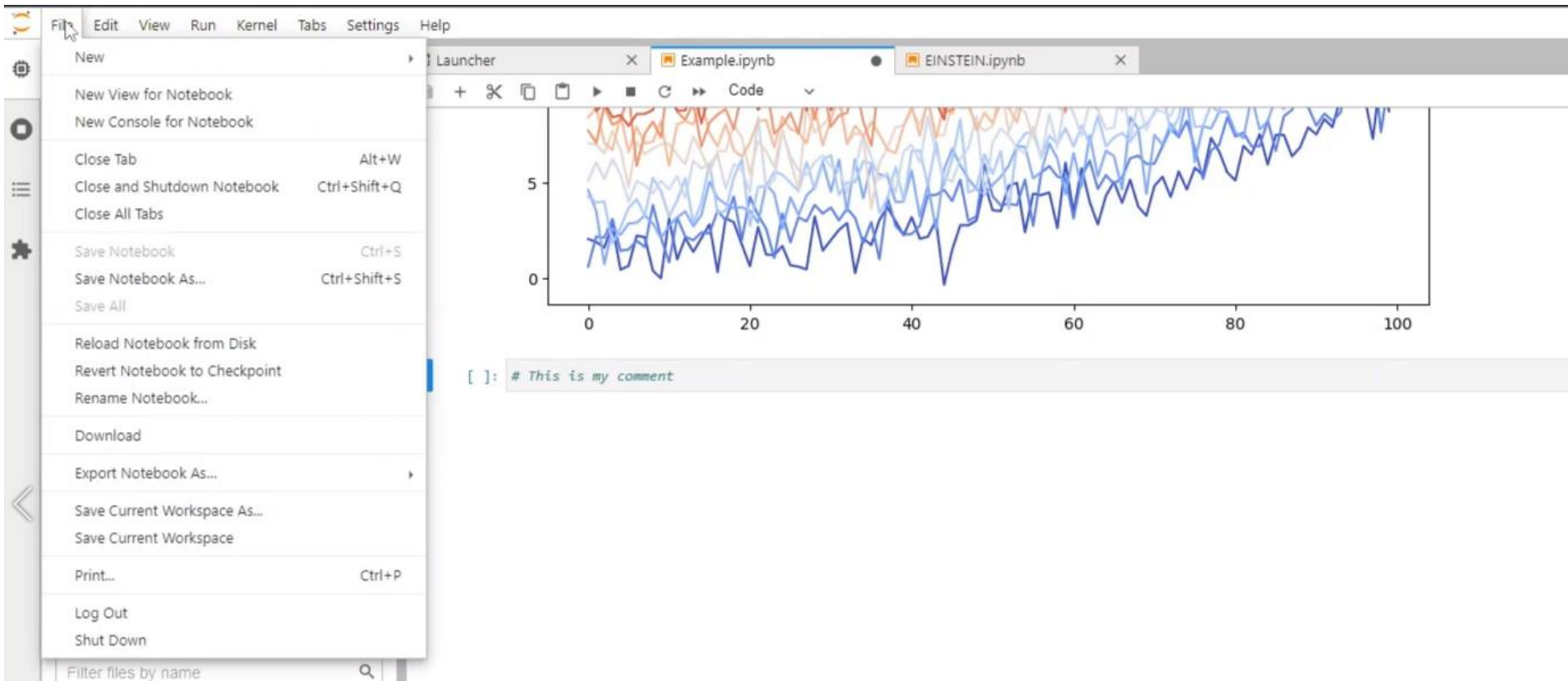
20
15

Cold
Medium
Hot

Collaboration workflows demo: High Energy Physics use case



Viewer role, Preview mode



The screenshot displays the JupyterLab application interface. The 'File' menu is open, showing options such as 'New', 'New View for Notebook', 'Close Tab', 'Save Notebook', 'Export Notebook As...', and 'Print...'. The main workspace contains a line plot with multiple data series in blue and orange, plotted against an x-axis ranging from 0 to 100 and a y-axis ranging from 0 to 5. Below the plot, a code cell is visible with the text `[]: # This is my comment`.



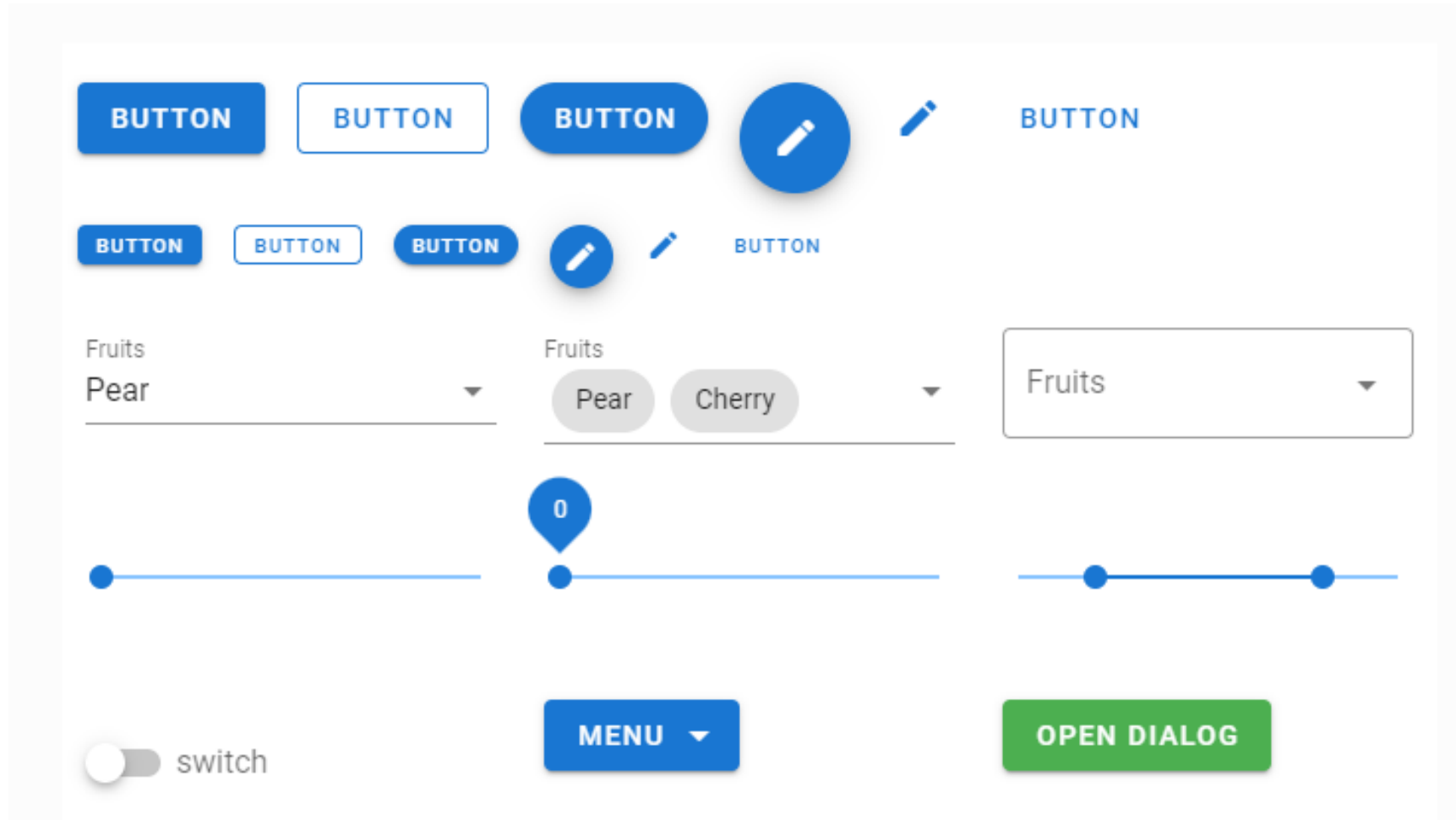
voilà

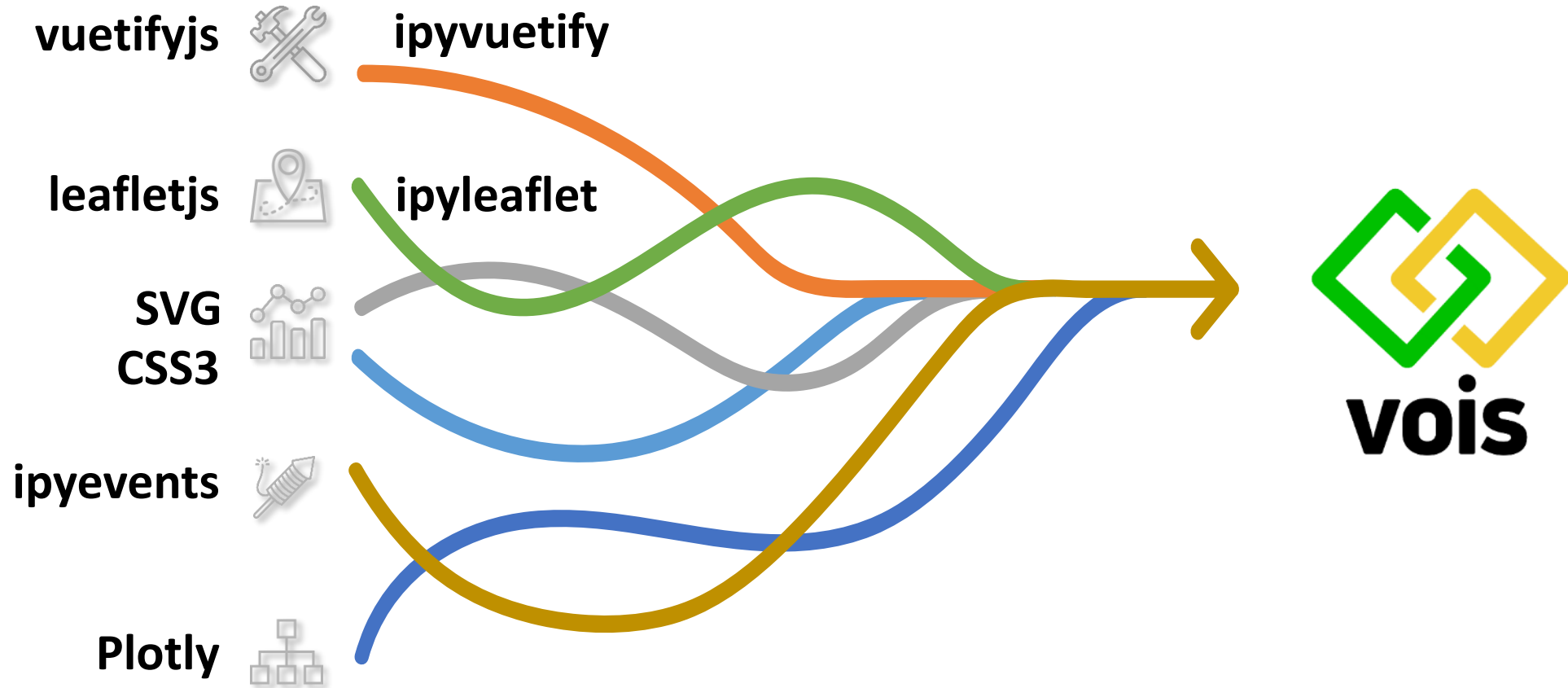
- # A Jupyter notebook extension to automatically create standalone applications and dashboards.
- # Notebooks are rendered by showing only the output of the cells, while the code is hidden.
- # Suitable for non-technical experts for communicating insights and foresight to a wider audience.
- # Single environment for full data analytics workflows from research and innovation to outreach engaging policy makers and citizens.



- # **VOilà** Simplification library
- # Its goal is to make life easier for the creation of impactful Voilà dashboards
- # Developed by **JRC**: Earth Observation use case of Science Mesh Data Science environments
- # VOIS library is used inside the European Commission BDAP Cloud Platform to communicate scientific results to a wider audience.
- # After having passed the IP clearance and security checks phases, the library is ready for publication as open source on the <https://code.europa.eu>

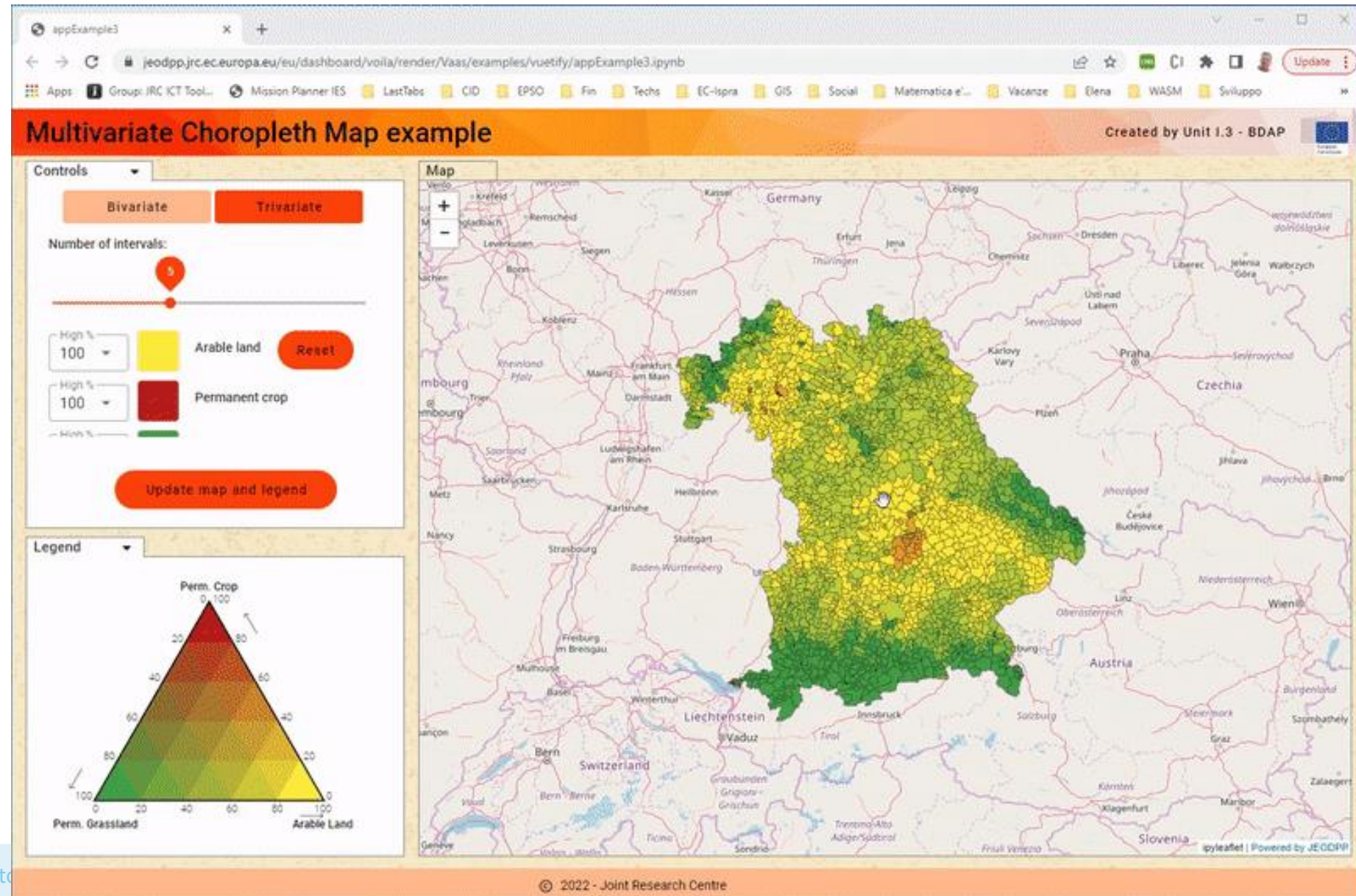






The VOIS library has some modules dedicated to geospatial data visualization

On the right an example dashboard to interactively create a bi-variate or tri-variate choropleth map, i.e. a map representation where the colors of the polygons depend on two or three data attributes



JRC Big Data Analytics Platform

Hardware

- Petabyte scale data hub
- Co-located with computing cluster

Interactivity

- Novel interactive data analysis
- Exploratory visualisation tools

Dissemination

- Web-based data dissemination
- Visualisation services
- Secure file transfer (ftps)

Security

- Encrypted protocol (https)
- Multi-factor authentication

Web-based access

- JupyterLab
- Remote data science desktop
- Distributed computing
- Development environments

Machine Learning

- Specialized hardware
- Artificial Intelligence and Deep Learning modelling



Thank you!
Discover more on...

 cs3mesh4eosc.eu

 [company/cs3mesh4eosc](https://company.linkedin.com/cs3mesh4eosc)

 [CS3org](https://twitter.com/CS3org)

 [CS3MESH4EOSC Project](https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw)

<https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw>



CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.



**CS³
MESH⁴
EOSC**

Connecting European Data



SWAN: CERN's service for web-based analysis

Enric Tejedor, CERN



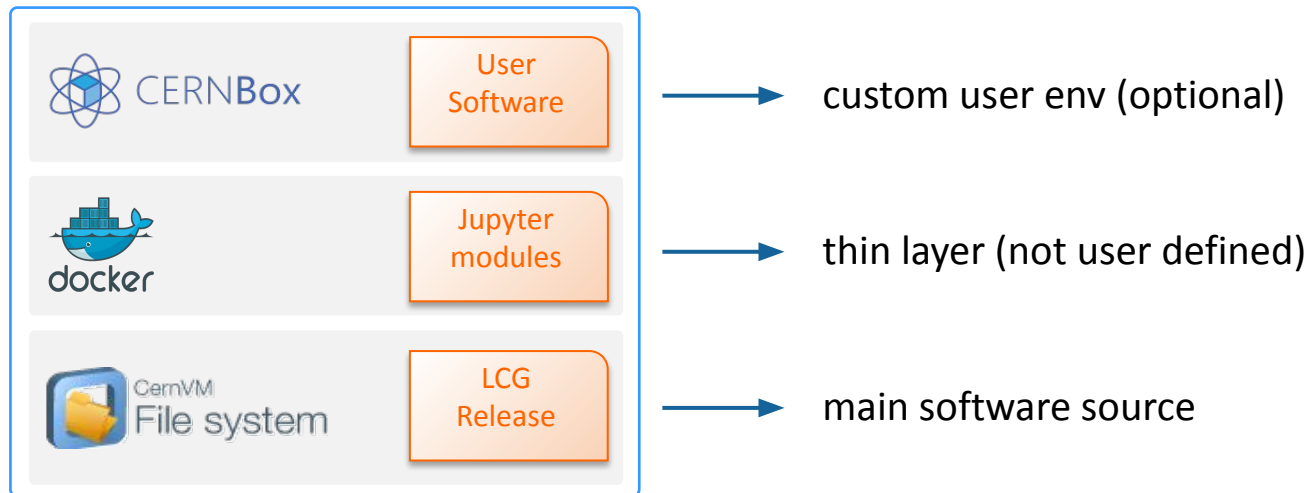
CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.

- > Interactive analysis with a web browser
 - No local installation needed
 - Based on Jupyter Notebooks
 - Calculations, input data and results “in the Cloud”
- > Easy sharing of scientific results: plots, data, code
- > Good for data analysis and exploration, and also teaching
- > Integration with CERN resources → added value!
 - Software (CVMFS)
 - Storage (EOS, CERNBox)
 - Computing (GPU, Spark, HTCondor)

SWAN's building blocks

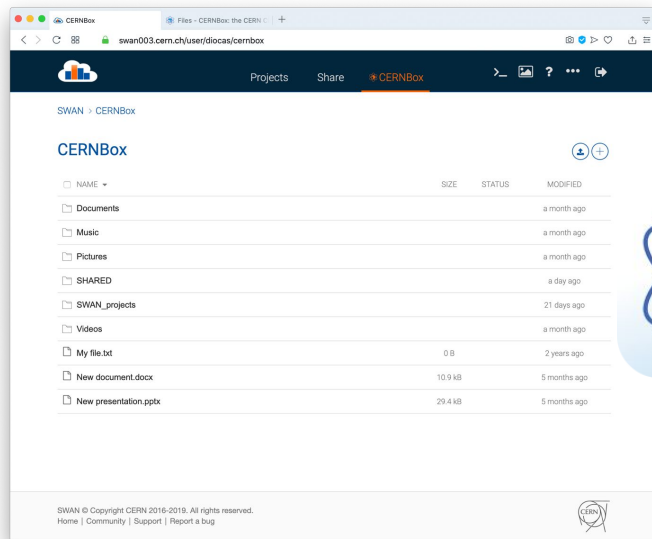


- > Find the software you need for your analysis
 - CVMFS: generic software stacks and experiment-specific ones
 - EOS: custom software environment



> Find the data you need for your analysis

- EOS: experiment repositories (/eos/atlas, /eos/cms, ...), projects, open data
- CERNBox as home directory, sync & share



share

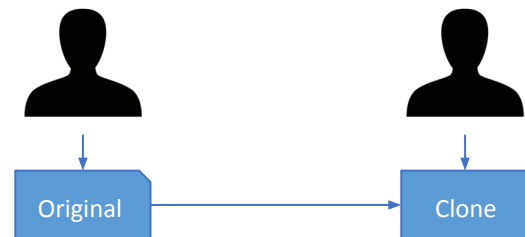


sync



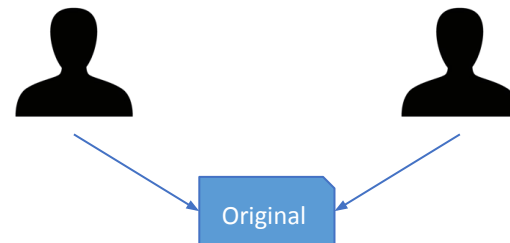
> Current sharing based on cloning

- Clones prevent overwrites if notebooks are edited concurrently
- Not ideal for collaboration: need to reclone every time some scientist produces a modification of the analysis



> Better model: **parallel editing**

- Proposed by CS3Mesh
- A user locks a shared notebook when editing it
- Still not concurrent editing but... do we really need it?



> Physics analysis

- Usually last stages of analysis
- Interactive, exploratory
- Collision event data, ntuple-like, columnar
- More and more with Machine Learning

> Non-physics analysis

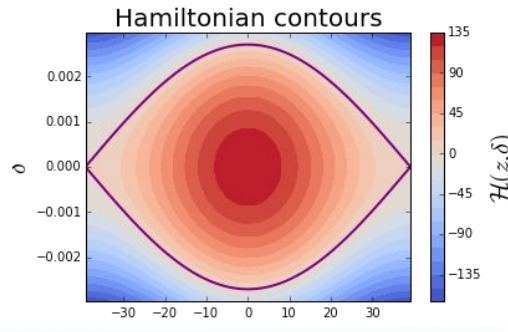
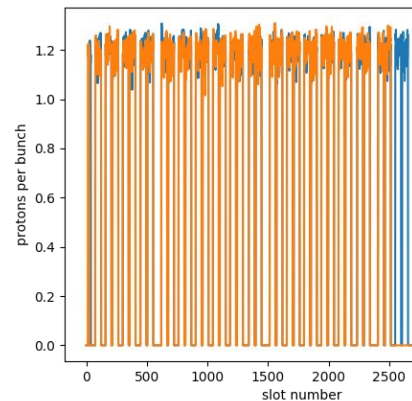
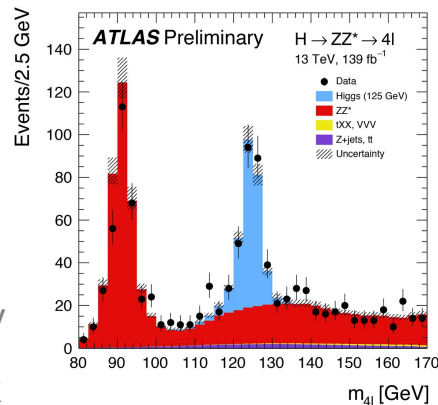
- LHC studies: extract machine measurements, query machine settings and beam dynamics simulations
- Query and process LHC logs distributedly via Spark
- Query and plot monitoring data in experiment DAQ systems

> Education

- Many schools/workshops use SWAN for teaching

> Outreach

- Analysis of high-energy-physics open data



- > CERN Open Science
 - CERN's research made available to other researchers and society
 - Publications, software, data

- > Open Data for education, training and outreach

- > **Voilà** as a tool for showcasing CERN's research
 - E.g. Open data analyses

Plot the dimuon spectrum

Now, the computation graph is set up. Next, we want to have a look at the result.

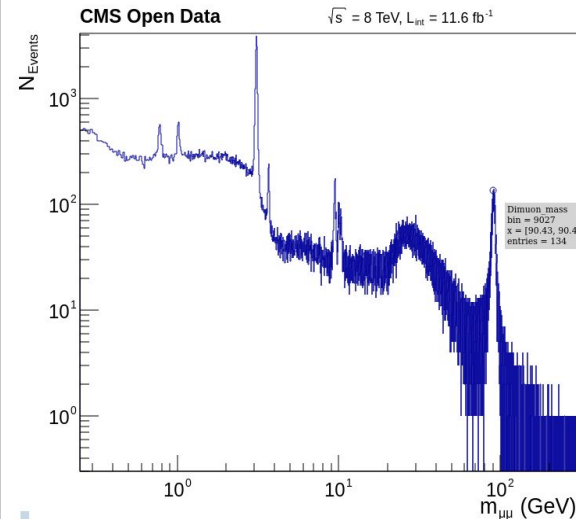
Note that the event loop actually runs the first time we try to access the histogram object (results of an `RDataFrame` graph are computed lazily).

`%time` measures the time spend in the full cell. You can compare it with the C++ equivalent of this notebook, it should be very similar since (almost) everything happens in C++ under the hood!

CPU times: user 5.72 s, sys: 266 ms, total: 5.99 s
Wall time: 6.17 s

ROOT provides interactive JavaScript graphics for Jupyter, which can be activated with the `%jsroot` magic. Click and drag on the axis to zoom in and double click to reset the view.

Don't forget that you can improve the statistics by increasing the number of events given to `Range`.





Connecting European Data

Thank you!
Discover more on...

 cs3mesh4eosc.eu

 company/c3mesh4eosc

 [CS3org](https://twitter.com/CS3org)

 [CS3MESH4EOSC Project](https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw)
<https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw>



CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.



**CS³
MESH⁴
EOSC**

Connecting European Data



On Demand Data Transfers

Connecting the ESCAPE Data Lake with the ScienceMesh

EGI 2023, Poznan, June 19-23 2023



CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.

Address the common needs of ESFRIs

Components

- # Data Lake (DIOS)
- # Science Platforms (ESAP)
- # Citizen Science (CS)
- # Software Repository (OSSR)
- # Virtual Observatory (VO)

(<https://projectescape.eu>)



(Slide stolen from G. Lamanna)

Two types of data transfers

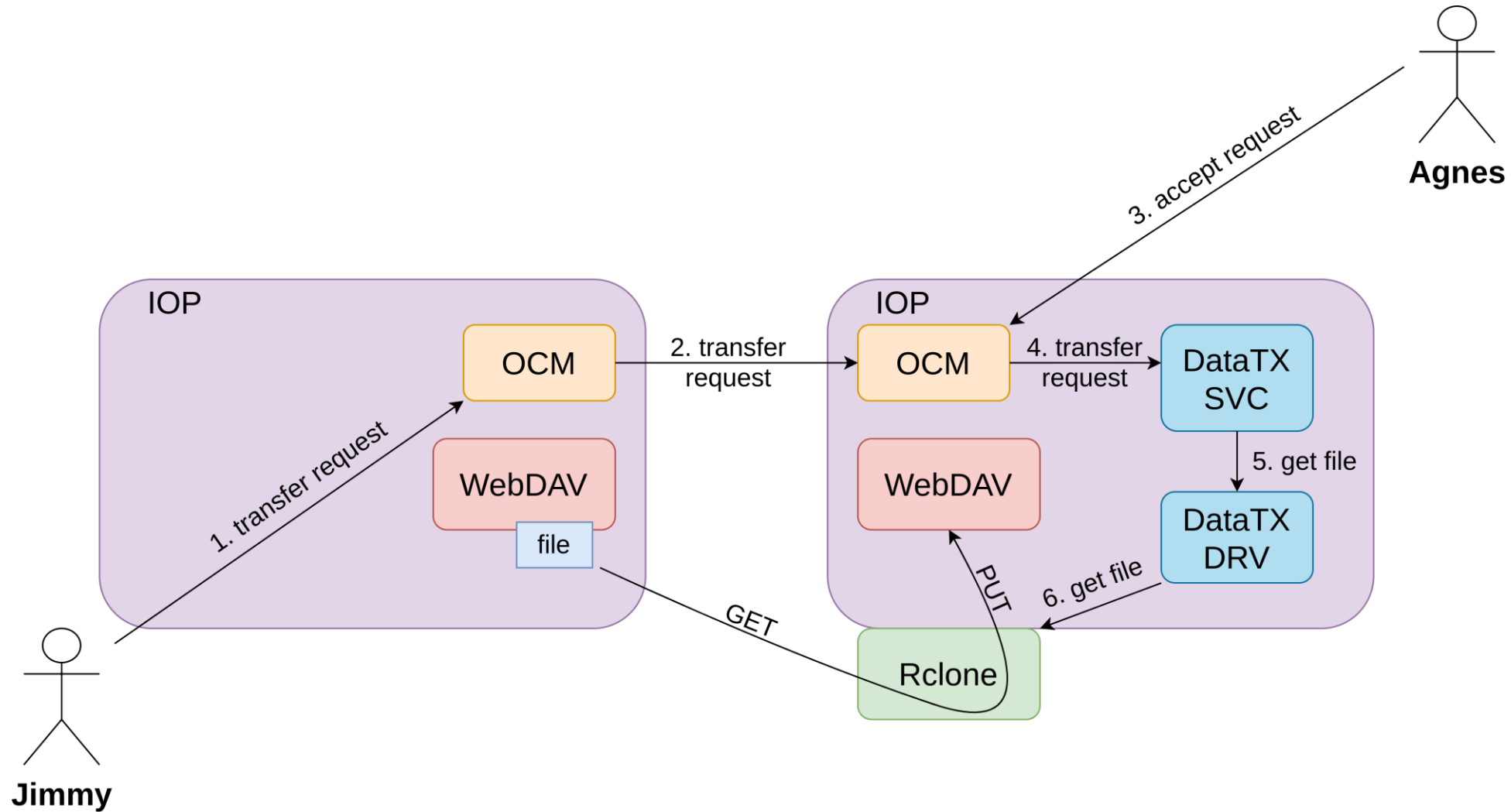
Ad-hoc

- # For small/medium sized data transfer needs within the ScienceMesh
- # Individuals or individual research groups
- # Rclone

Managed data transfers

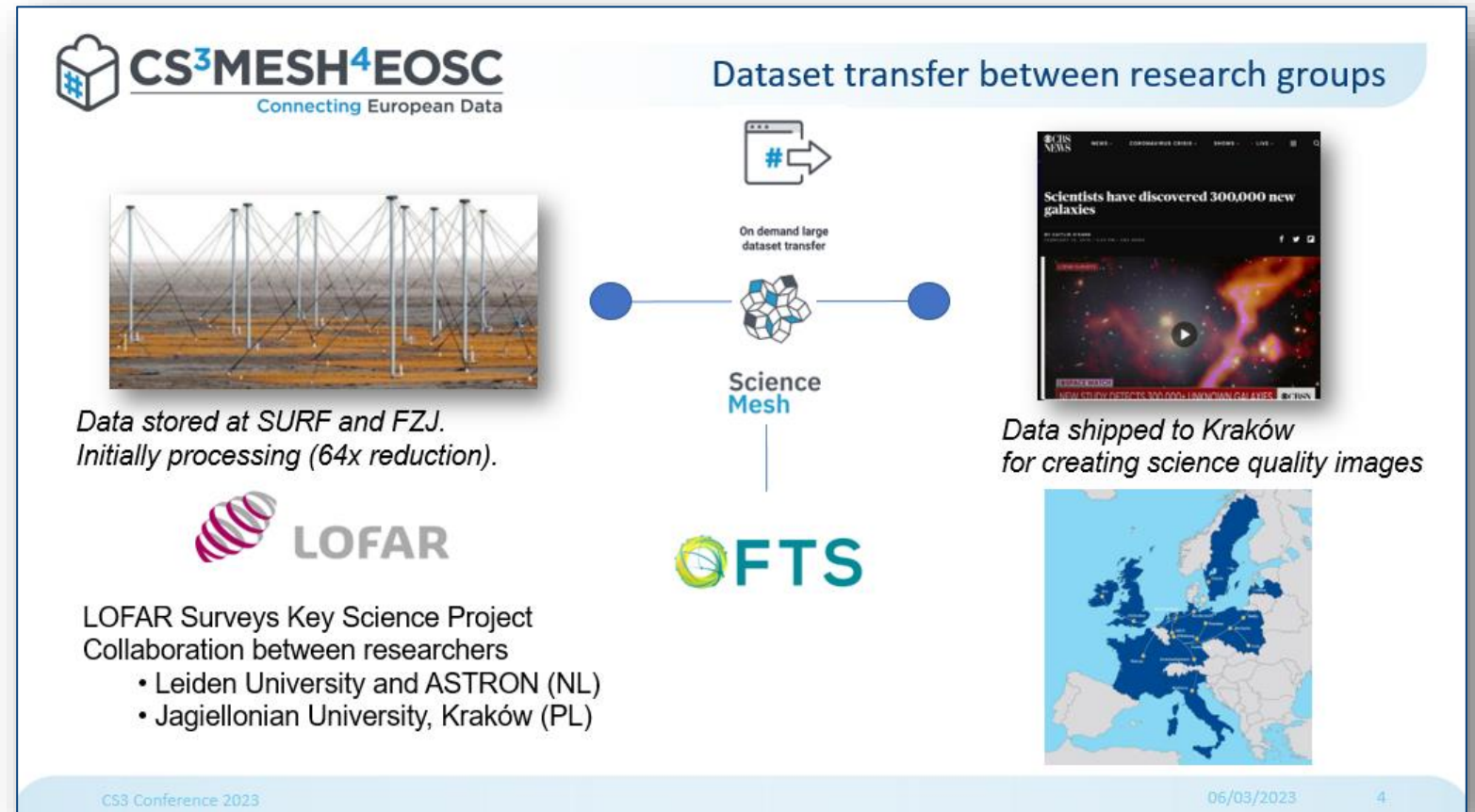
- # Connecting Big Science with sync-and-share (EFSS)
- # Communities
- # FTS
- # Federated AAI








Getting the data from the telescope to the end-user

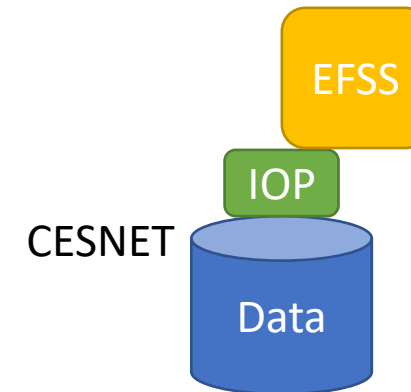
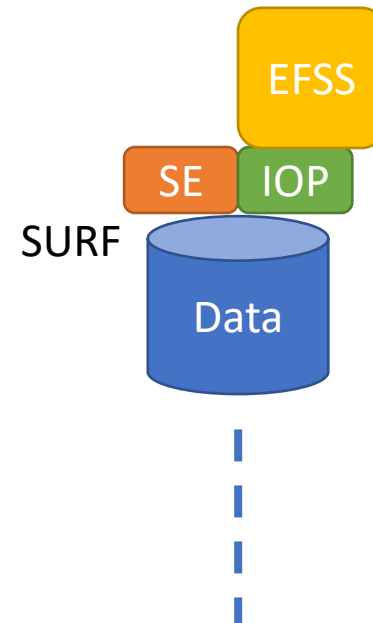
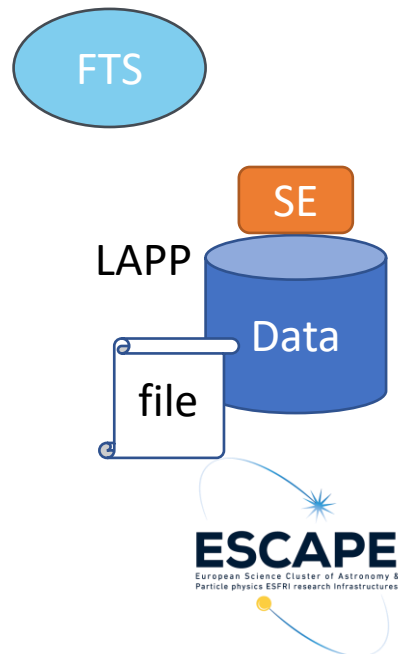


 Stella (data manager)

 Yan (PI)

 Mirek

FTS client





Stella (data manager)



Yan (PI)

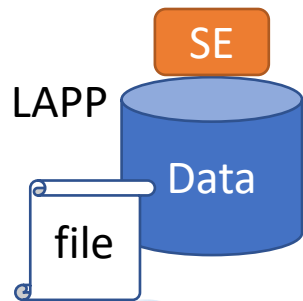


Mirek

Replicate data from LAPP to SURF

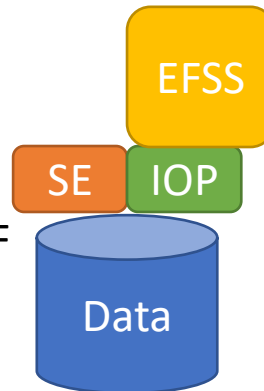
FTS client

FTS

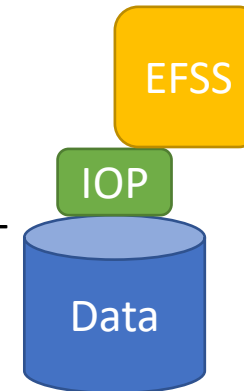



ESCAPE
European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

SURF



CESNET

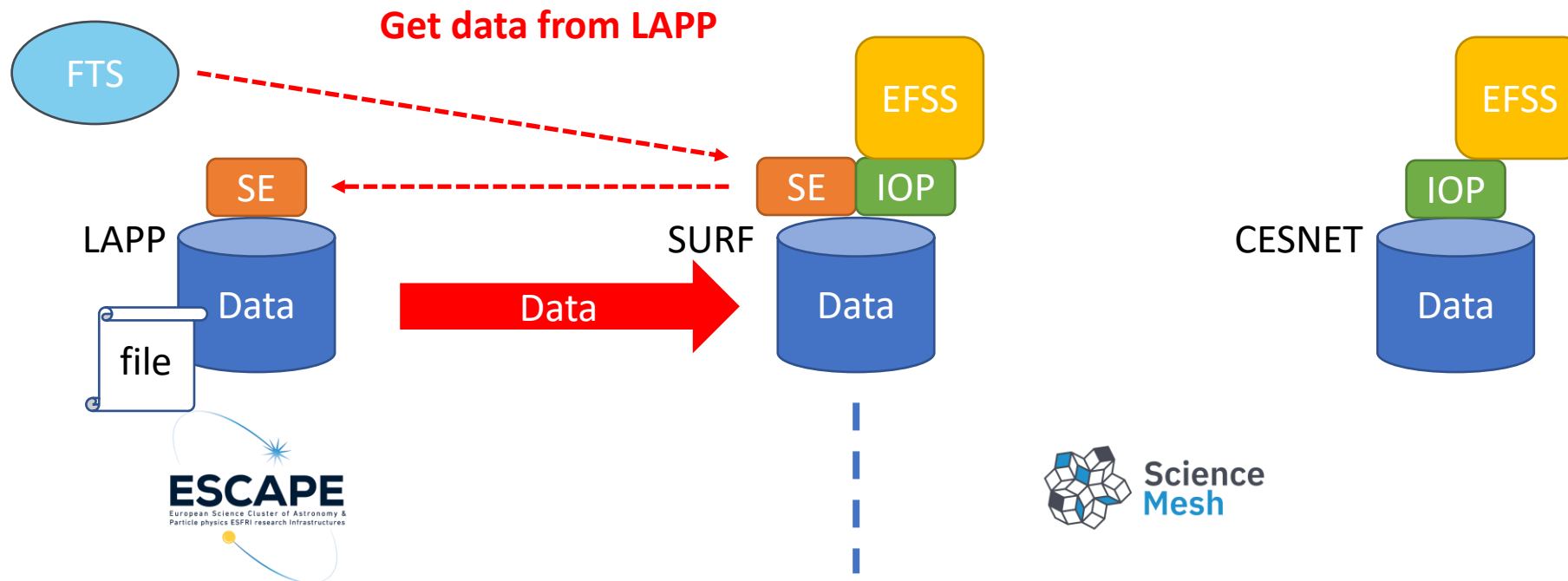



 Stella (data manager)

 Yan (PI)

 Mirek

FTS client



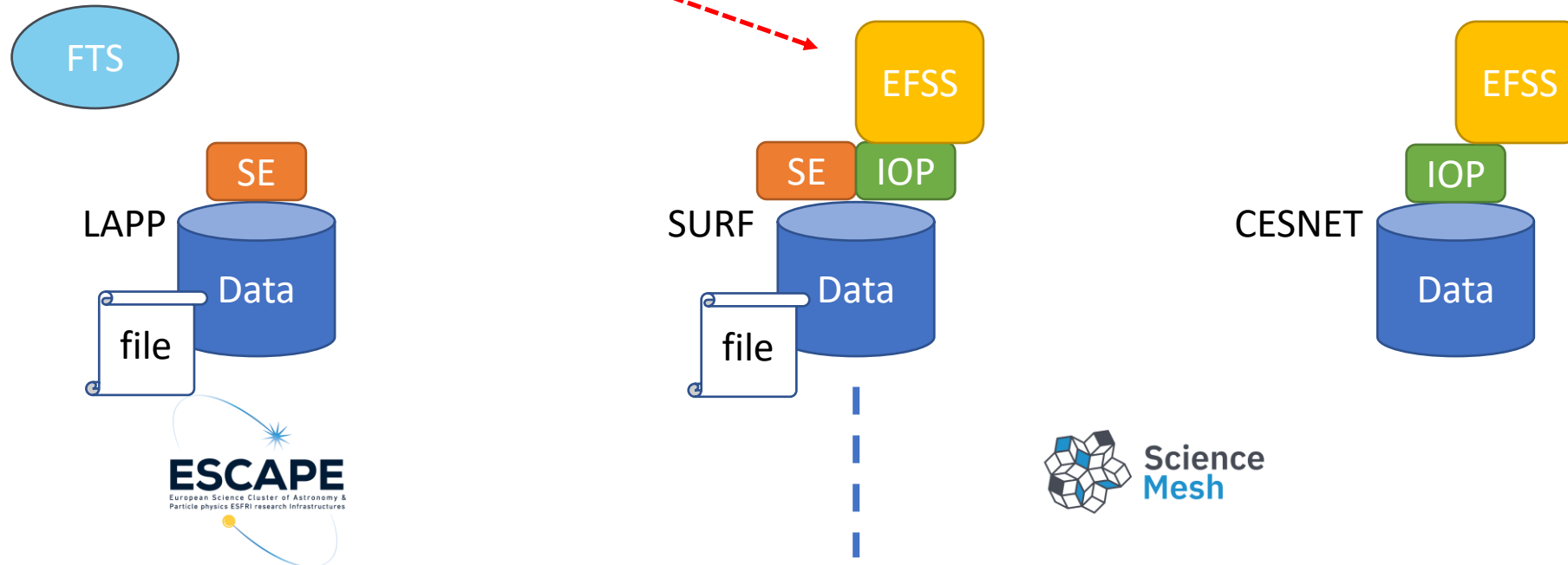
 Stella (data manager)


 Yan (PI)

 Mirek

Share data with Yan

FTS client



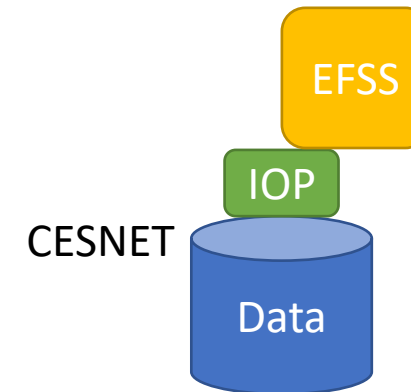
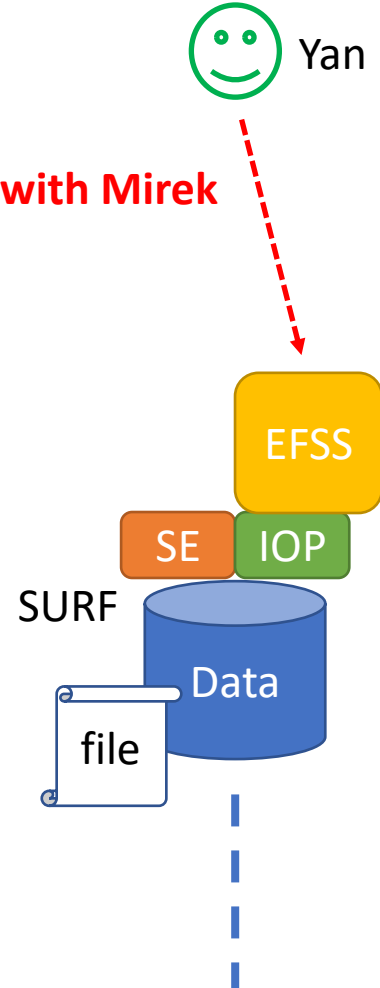
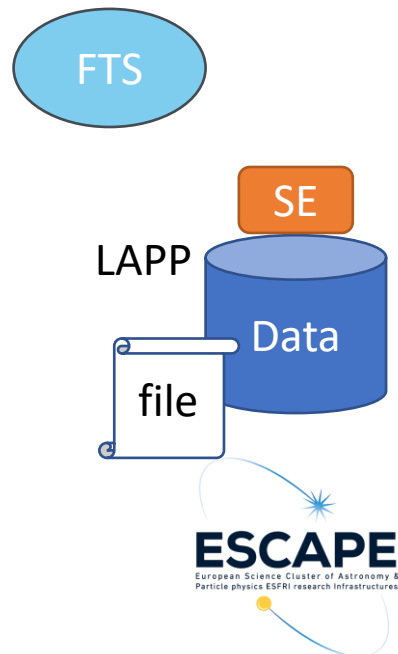
 Stella (data manager)


 Yan (PI)

 Mirek

Share data with Mirek

FTS client

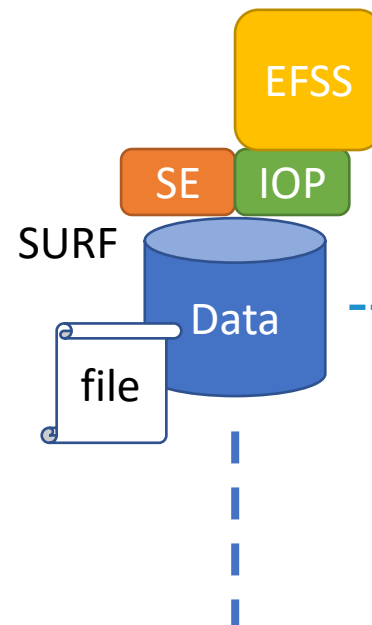
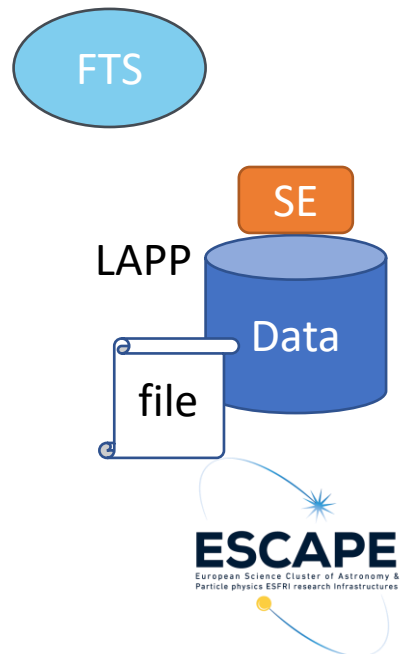


 Stella (data manager)

 Yan (PI)

 Mirek

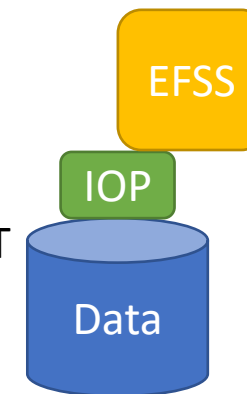
FTS client



Show me the data

CESNET

Federated share



Data Transfers

Connecting the  **ESCAPE** datalake
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

with the  **Science Mesh**

https://www.youtube.com/watch_popup?v=Nhv-btV8TUI



Thank you!
Discover more on...

 cs3mesh4eosc.eu

 company/c3mesh4eosc

 [CS3org](https://twitter.com/CS3org)

 [CS3MESH4EOSC Project](https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw)

<https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw>



CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.