



**CS<sup>3</sup>  
MESH<sup>4</sup>  
EOSC**

**Connecting European Data**



# Federating Sync-and-Share Services

Narges Zarrabi

CompBioMed2 - All Hands Meeting

22-06-2022



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

### # What is a sync and share service?

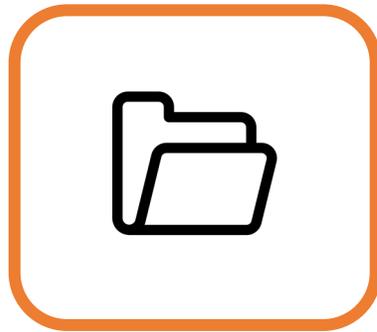
- # Access data on cloud
- # Sharing data and collaborative editing
- # Synchronise data between multiple devices
- # Built-in applications



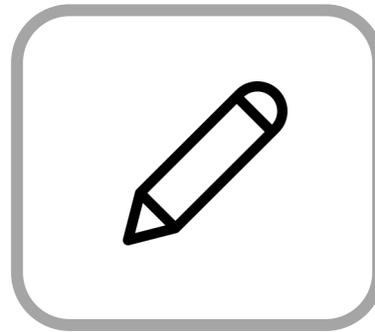
### # Enterprise File Synchronization and Sharing (EFSS)



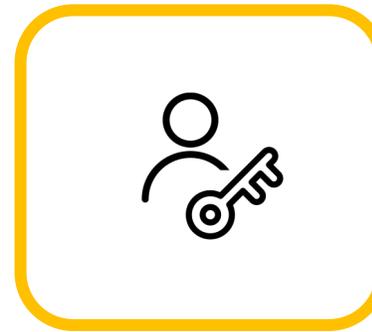
Cloud environment



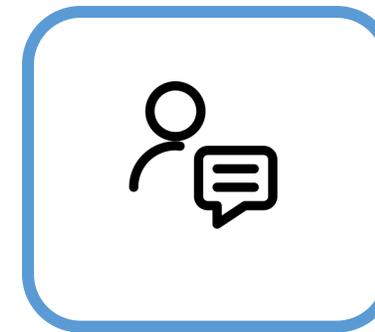
Sharing files & folders



Collaborative editing



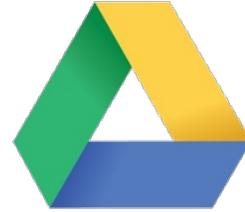
Roles and permissions



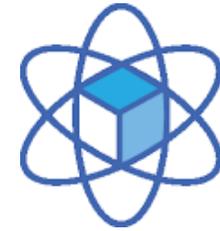
International access



Large storage



Google Drive



CERNBox



Nextcloud





**CS<sup>3</sup>  
MESH<sup>4</sup>  
EOSC**

Connecting European Data

# **C**loud **S**torage **S**ervices for **S**ynchronization and **S**haring (CS3)

# Science **M**esh (federated science cloud)

# **E**uropean **O**pen **S**cience **C**loud

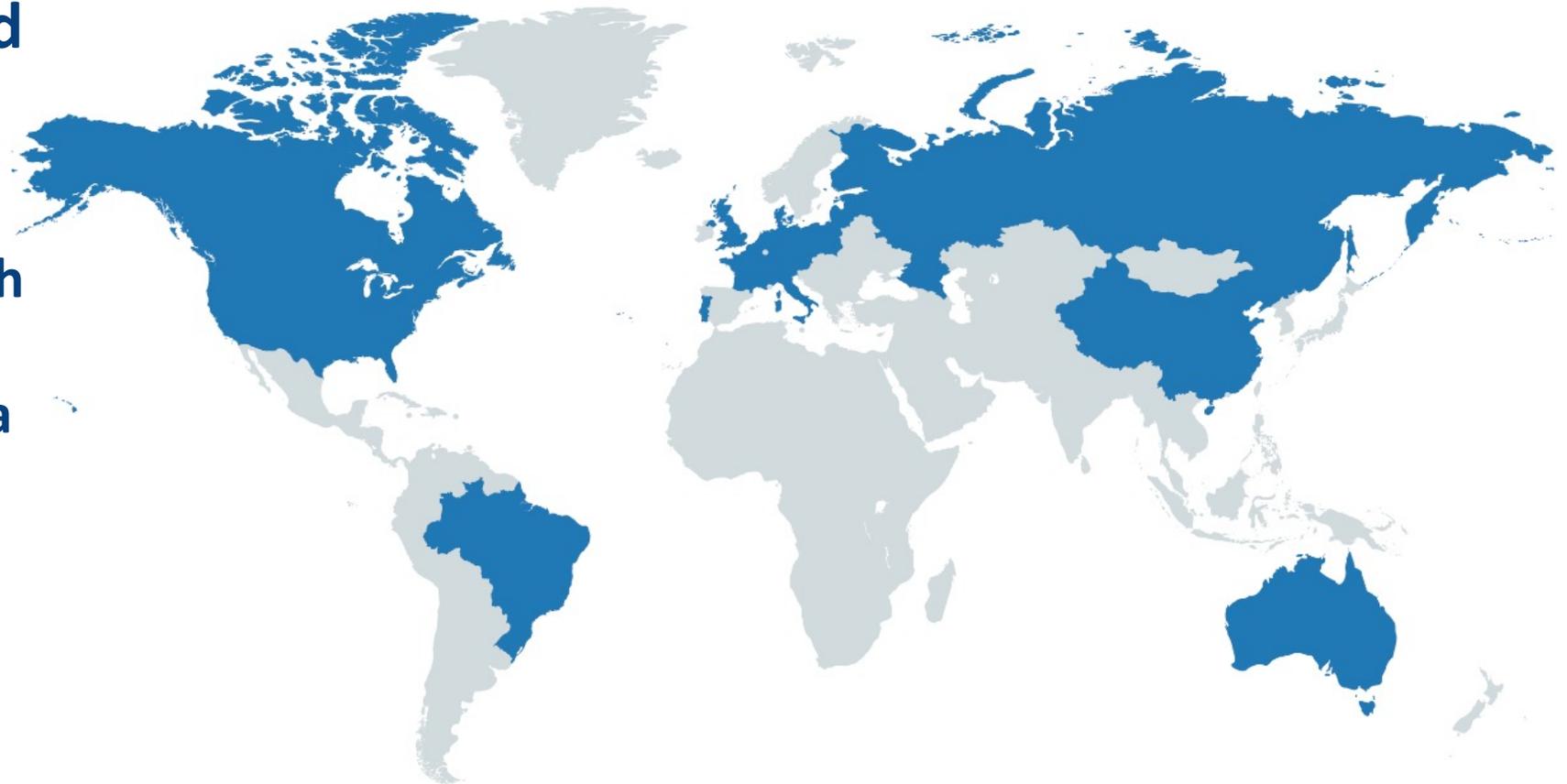


70 institutions 

25 countries 

147 participants 

*Participation from 2018*



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

- # Services integrated with user environments
- # Application for big-data analysis
- # Science outreach and education

## # CS3: community of practice since 2014

### # File Sync & Share (EFSS) services operated by...

- # Research labs, NRENs, Universities, companies,...
- # 130+ organisations, 30+ service nodes, 25 countries
- # > 400K users, >16 PB data, >3.5 billion objects

### # Research & Data Services...

- # Big Science, Education, long-tail,....

<http://www.cs3community.org>



Virtual 2021



København 2020



Roma 2019



Kraków 2018



Amsterdam 2017



Zürich 2016

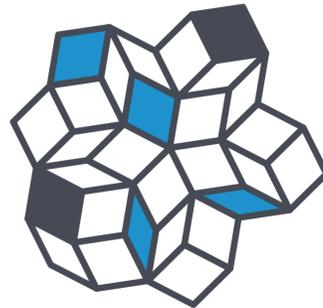


Genève 2014

- # Moved up in the stack by integration with more services
  - # Collaborative tools (editing)
  - # Research Interfaces
    - # Jupyter Notebooks
    - See Marcin's presentation:  
<https://indico.egeu.eu/event/5464/contributions/15704/>
  - # Integration with data repositories
  - # Open data publishing (FAIR)
- # From simple storage facilities to data science and research environments



- # CS3MESH4EOSC is a H2020 Project (2020-2022)
  - # Interconnects **existing sustainable storage and research services**
  - # Create an **innovative collaborative mesh** of application services
  - # Supports **open source** software and **FAIR, open data** principles
  - # Build an operational, **decentralised infrastructure** (with a lightweight central component) based on the protocols:
    - # CS3APIs (links apps with an EFSS)
    - # Open Cloud Mesh/OCM (links EFSS')



**Science  
Mesh**

# <https://cordis.europa.eu/project/id/863353>



# Science Mesh

- # Decentralized **Mesh of EFSS nodes**
- # Based on **Open Standards** and **Open Source Software**
- # **Federated** environment where researchers can collaborate
- # **Application Platform** for distributed collaborative tools

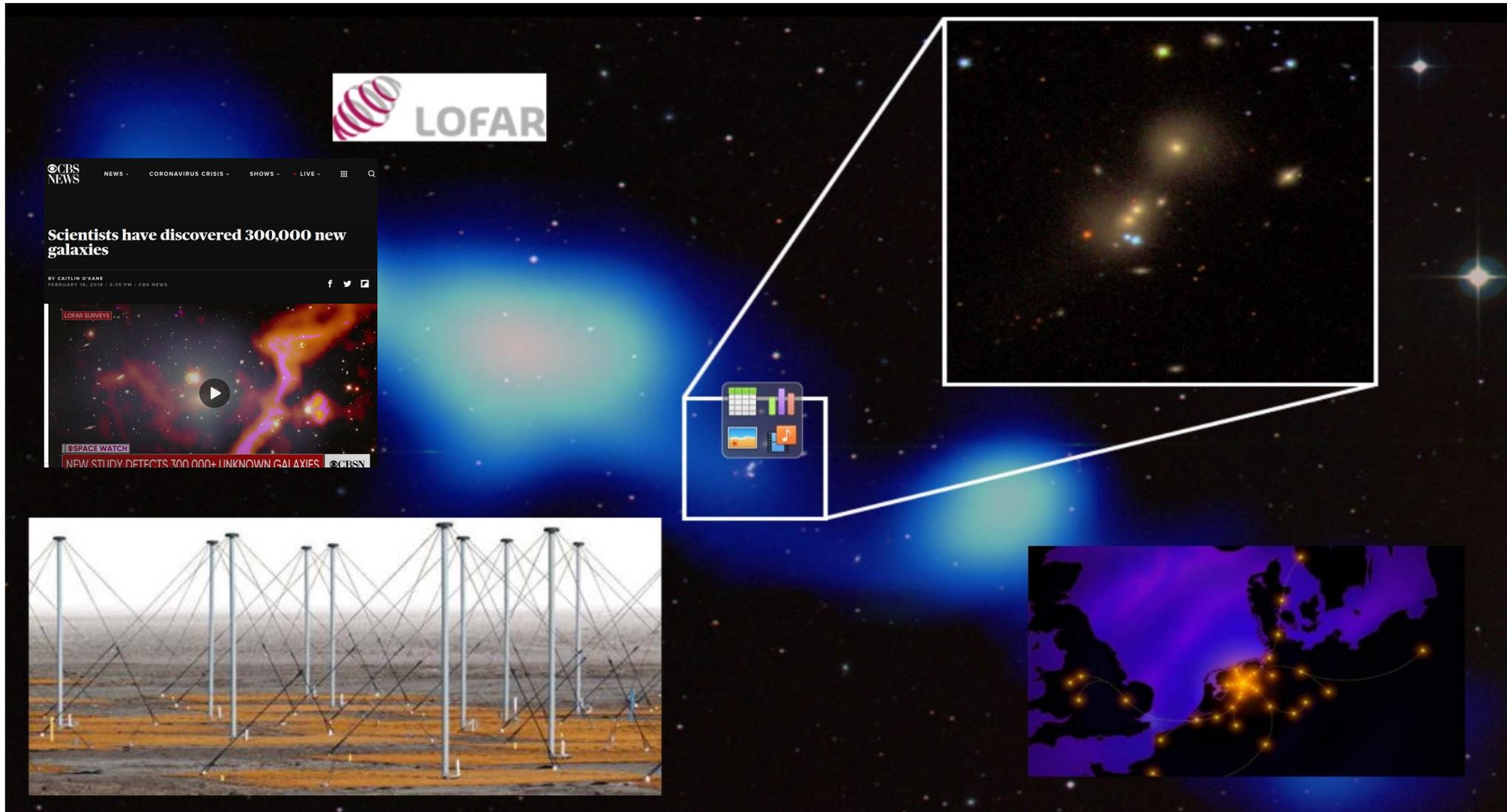


## # Collaborative Data Science:

- # Astronomy
- # Finance
- # IoT
- # Earth Observation<sup>1</sup>
- # High Energy Physics<sup>2</sup>
- # SmartCities
- # Pharma
- # ... In virtually every business.

[1] <https://cs3mesh4eosc.eu/use-cases/data-science-environments-monitoring-land-degradation>

[2] <https://cs3mesh4eosc.eu/use-cases/data-science-environments-high-energy-physics-cern>



The image is a composite illustrating the LOFAR use case. It features a central background of a starry night sky. On the left, there is a screenshot of a CBS News article titled "Scientists have discovered 300,000 new galaxies" with a video player. Above the article is the LOFAR logo. Below the article is a photograph of the LOFAR radio telescope array in a desert landscape. On the right, a white-bordered box shows a zoomed-in view of a galaxy cluster. Below this box is a small icon representing data analysis tools (grid, bar chart, pie chart, film strip, music note). At the bottom right, there is a map of Europe with glowing orange dots and lines representing the LOFAR telescope stations.

## Dataset transfer between research groups

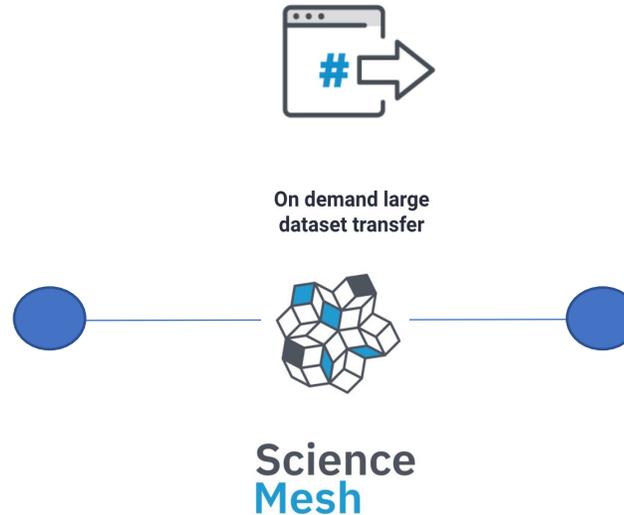


*Data stored at SURF and FZJ.  
Initially processing (64x reduction).*



LOFAR Surveys Key Science Project  
Collaboration between researchers

- Leiden University and ASTRON (NL)
- Jagiellonian University, Kraków (PL)



*Data shipped to Kraków  
for creating science quality images*



## # Study land degradation with Data Analytics using Jupyter

```

File Edit View Run Kernel Tabs Settings Help
Launcher x MISLAND.ipymb x reports.py
Code
    rgb2hex(0,60,0), # 7=Forest
    ]

LULC_LandCover_ColorMap = { 0: "black" }
for i in range(len(LULC_LandCover_colors)):
    LULC_LandCover_ColorMap[i+1] = LULC_LandCover_colors[i]

# Custom python function to be applied to ESA CCI Land Cover
def LandCover_MaskAndRecode(img):
    v, esa = img[0],img[1]

    # Mask-out what is outside the input polygon
    esa[v == 0] = 0

    # Recode
    esa[ (esa>=10) * (esa<=30) ] = 5 # Crop
    esa[ esa==40 ] = 6 # Grass
    esa[ (esa>=50) * (esa<=100) ] = 7 # Forest
    esa[ (esa>=110) * (esa<=153) ] = 6 # Grass
    esa[ (esa>=160) * (esa<=180) ] = 4 # Wetland
    esa[ esa==190 ] = 3 # Urban
    esa[ (esa>=200) * (esa<=202) ] = 2 # Bare
    esa[ esa==210 ] = 1 # Water
    esa[ esa > 7 ] = 0 # Nodata including snow

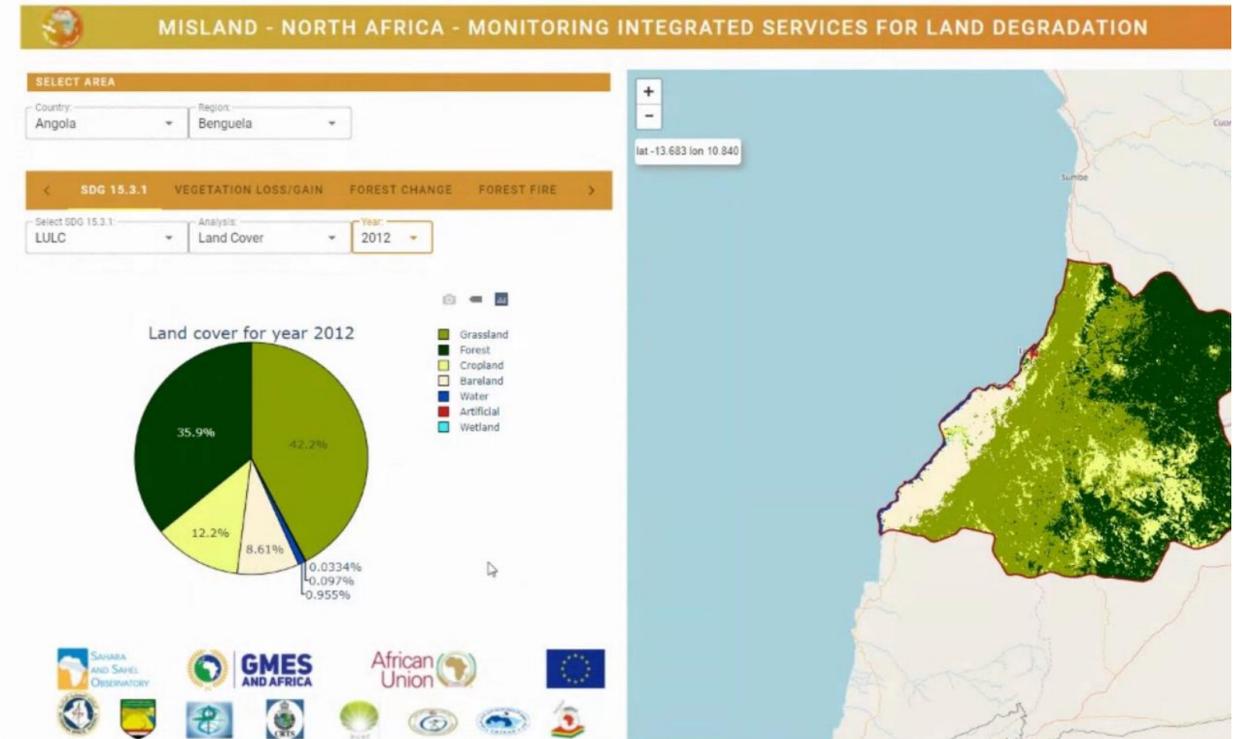
    return esa

# Process a polygon and display the output raster in the map
def LULC_LandCover():

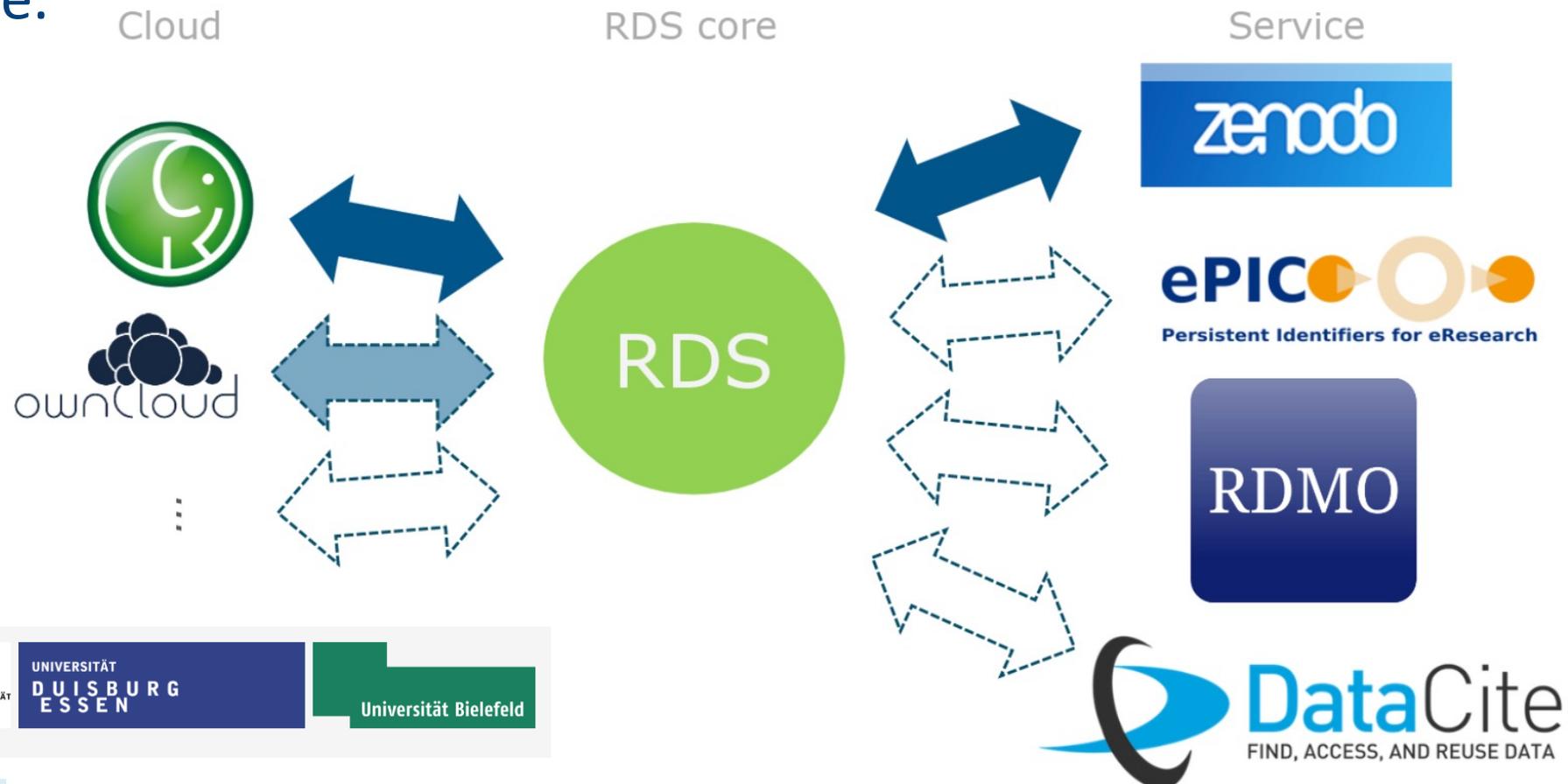
    if len(selected_geojson) > 0 and len(year.v_model) > 0:

        waitON()
        map.clear()

```



# RDS was designed as a middleware, which make it easy to interact with different internal or external services within an already existing user interface.



# Research Drive, a self-service sync-and-share solution for RDM

Power users



Reporting



Dashboard



Access to Research Drive service can be requested through the EOSC portal (B2DROP) <<< costs via DICE Virtual Access mechanism >>>

Users

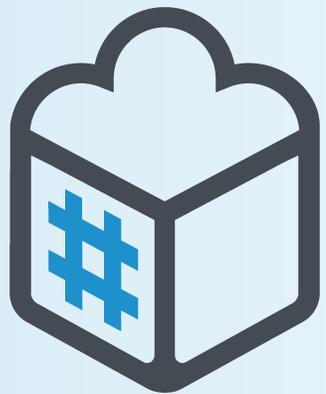


Integrated user apps



Storage backend





**CS<sup>3</sup>  
MESH<sup>4</sup>  
EOSC**

**Connecting European Data**

**Thank you!**  
Discover more on...

 [cs3mesh4eosc.eu](https://www.cs3mesh4eosc.eu)

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

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

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

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

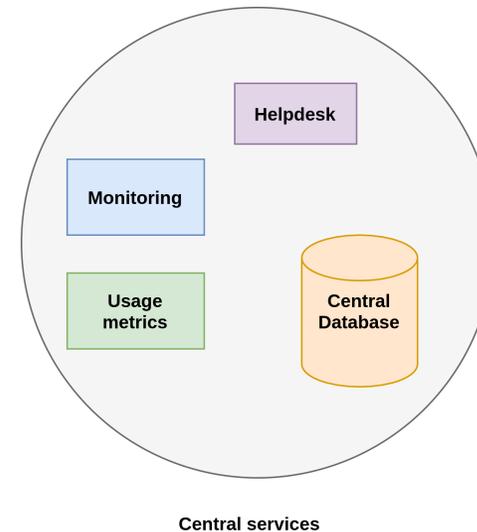
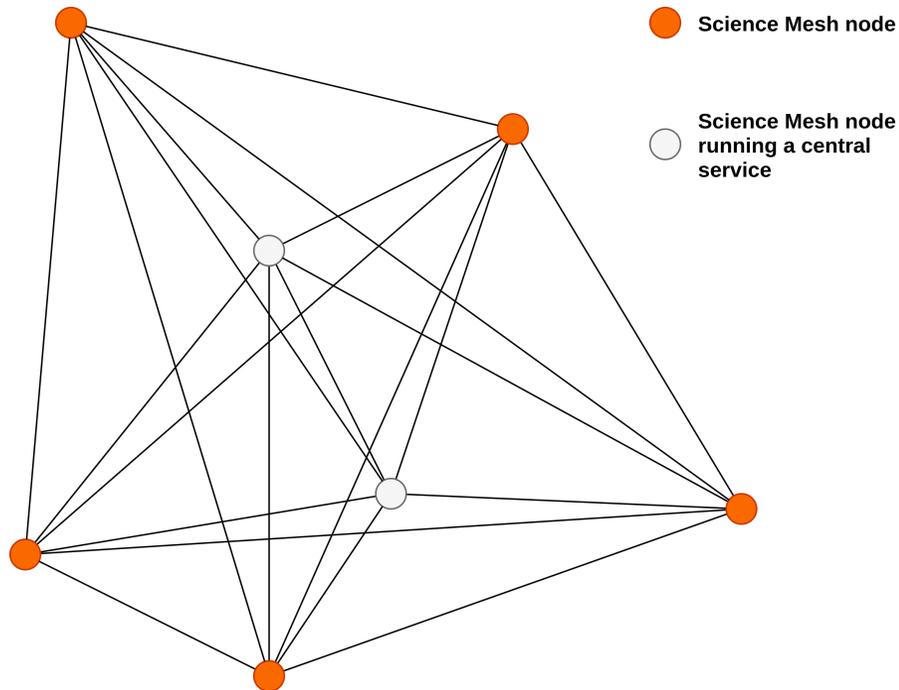


CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.



## # Starting point

- # Sites running **already sustainable** EFSS services
- # Joined together in an effort for further development of their services



## # What makes the Science Mesh the Science Mesh?

- # Policies

- # Protocols

  - # CS3APIs for application/service<-> EFSS

  - # Open Cloud Mesh (OCM) for EFSS<->EFSS

- # Local costs incurred at the Science Mesh nodes are already covered
  - # Science Mesh nodes themselves are already sustainable
- # Control costs of central overhead by minimising it.
  - # Under the assumption that obtaining funding would be easier this way
  - # Funding through:
    - # Local funding sources
    - # Membership fee
    - # .....



## # Why on-demand data transfers?

- # Data locality may be important for some use-cases
  - # Sharing of data is not good enough
    - # Remote processing may cause overload on the source EFSS system
  - # Larger data sets
  
- # Support for data transfers between EFSS systems and between EFSS systems and other storage systems

## # Different 3<sup>rd</sup> party data transfer mechanisms

### # Rclone

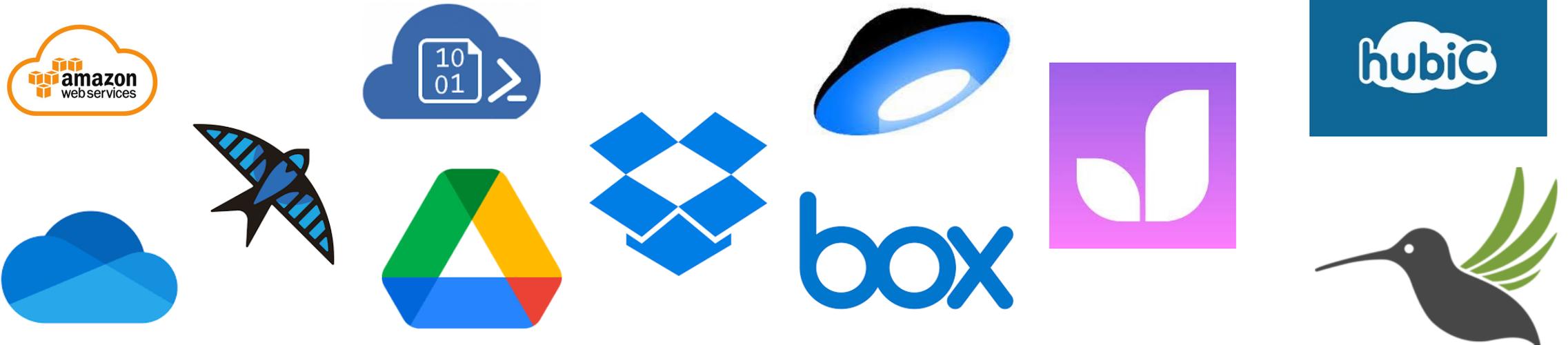
- # For small/medium sized data transfer needs
- # Individuals or individual research groups

### # Rucio/FTS

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



- # Commandline programme to manage files on cloud storages
- # It can interface with over 40 cloud storage products



- # Implements various unix commands like rsync, cp, mv, mkdir, mount etc. for cloud storages

- # Data transfer scheduler
- # Developed by CERN to distribute LHC data
- # Adopted by many other projects

## FTS



- Distributes majority of LHC data across WLCG infrastructure
- 7 WLCG and 13 non-WLCG instances
- ~25 Virtual Organisations
  - ATLAS, CMS, LHCb, AMS, NA62, Compass, ILC, Magic, Belle II, Mice, Xenon, Snoplus, Gridpp, Dune, LZ, Solidexperiment.org, SKA, Ligo, Icecube, Elixir, NP02, CAST, ESCAPE, Eiscat.se, Virgo
- Integrated with experiment frameworks: Rucio, PhEDEx, DIRAC
- Transferred in 2019 so far >750 PB (700 only for ATLAS, CMS and LHCb)



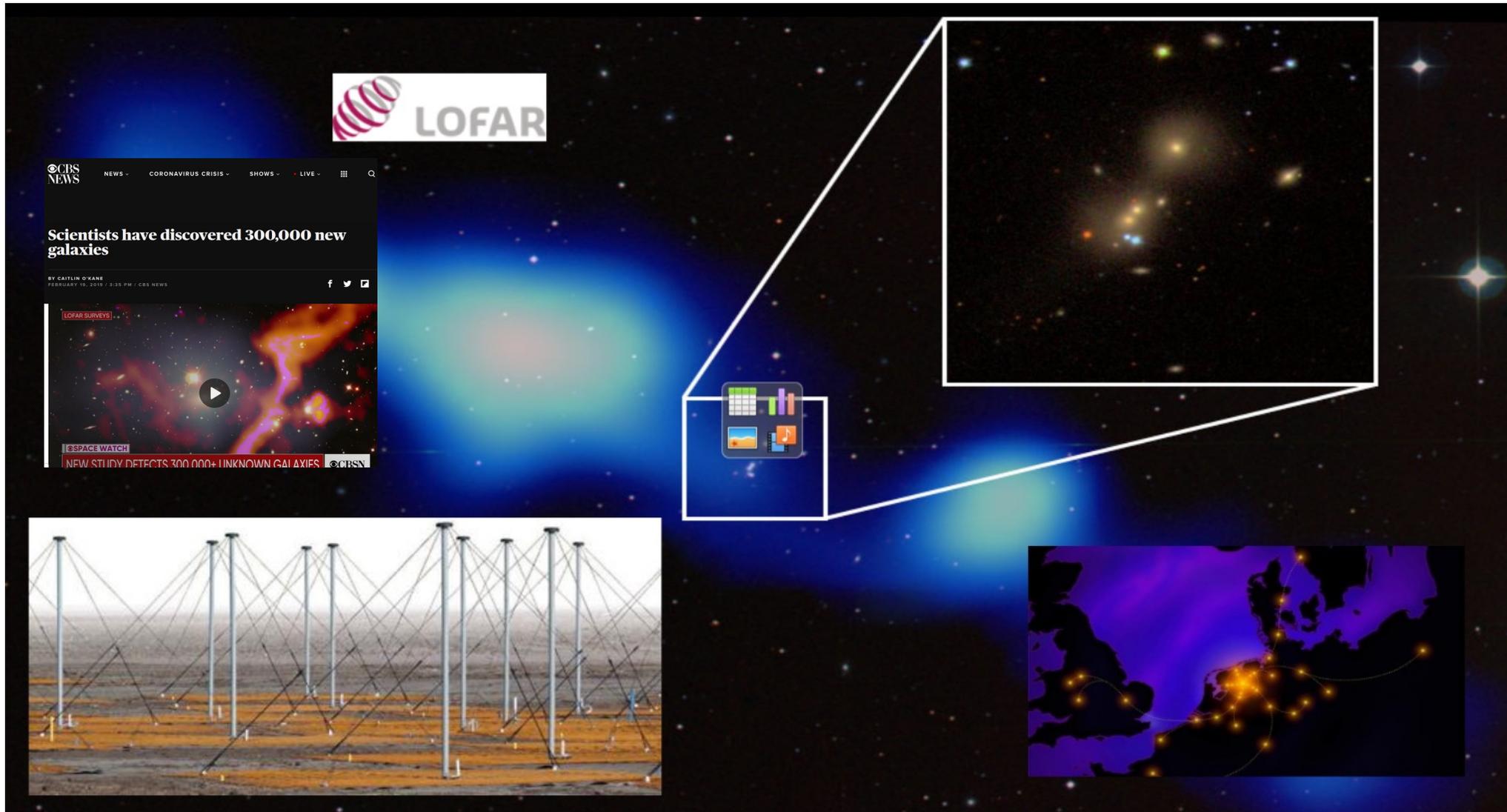
25	24PB	26M	20
Virtual Organizations	Volume/week	Transfers/week	FTS Instances


11/04/2019
FTS improvements for LHC Run-3 and beyond
3



- # Data management
- # Developed by ATLAS
- # What does it do?
  - # Managing data collections
  - # Metadata management
  - # Replicating data to registered storage elements
  - # .....
- # Uses the FTS for data transfers
- # Not only used by ATLAS, but also by CMS, SKA and many others





The image is a composite illustrating the LOFAR use case. It features a central background of a starry night sky. On the left, there is a screenshot of a CBS News article titled "Scientists have discovered 300,000 new galaxies" with a video player. Above the article is the LOFAR logo. Below the article is a photograph of the LOFAR radio telescope array in a desert landscape. On the right, a white-bordered box shows a zoomed-in view of a galaxy cluster. Below this box is a small icon representing data analysis tools (grid, bar chart, pie chart, film strip, music note). At the bottom right, there is a map of Europe with glowing orange dots and lines representing the LOFAR telescope stations.

## Dataset transfer between research groups

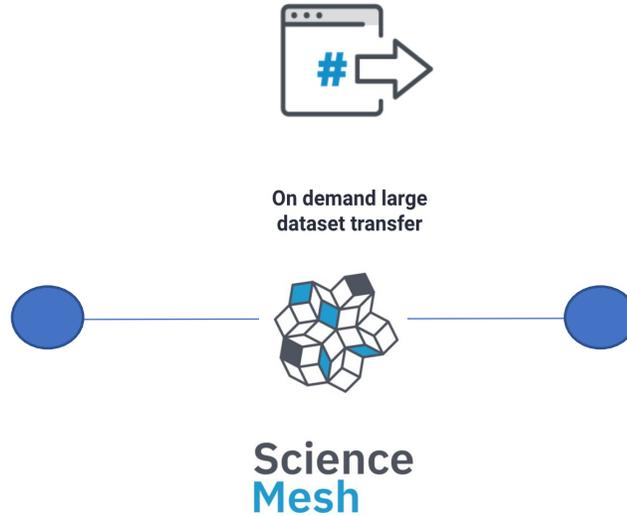


*Data stored at SURF and FZJ.  
Initially processing (64x reduction).*



LOFAR Surveys Key Science Project  
Collaboration between researchers

- Leiden University and ASTRON (NL)
- Jagiellonian University, Kraków (PL)



*Data shipped to Kraków  
for creating science quality images*





- # <https://sciencemesh.io>
- # <https://cs3mesh4eosc.eu>
- # <https://reva.link>
- # <https://github.com/cs3org/cs3apis>
- # <https://github.com/cs3org/OCM-API>
- # <https://www.cs3community.org>