CS3 MESH4 EOSC

Connecting European Data

Federating Sync-and-Share Services

Narges Zarrabi

CompBioMed2 - All Hands Meeting
22-06-2022

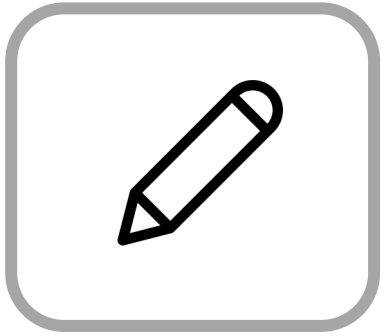
CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.

What is a Sync and Share Service?

- 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

Examples of Sync & Share Services

Dropbox

Google Drive

B2DROP

OneDrive

SURF DRIVE

CERNBox

RESEARCH DRIVE

box

ownCloud

Nextcloud

CITRIX ShareFile

- **Cloud Storage Services for Synchronization and Sharing (CS3)**
- **Science Mesh (federated science cloud)**
- **European Open Science Cloud**

CS3 Community

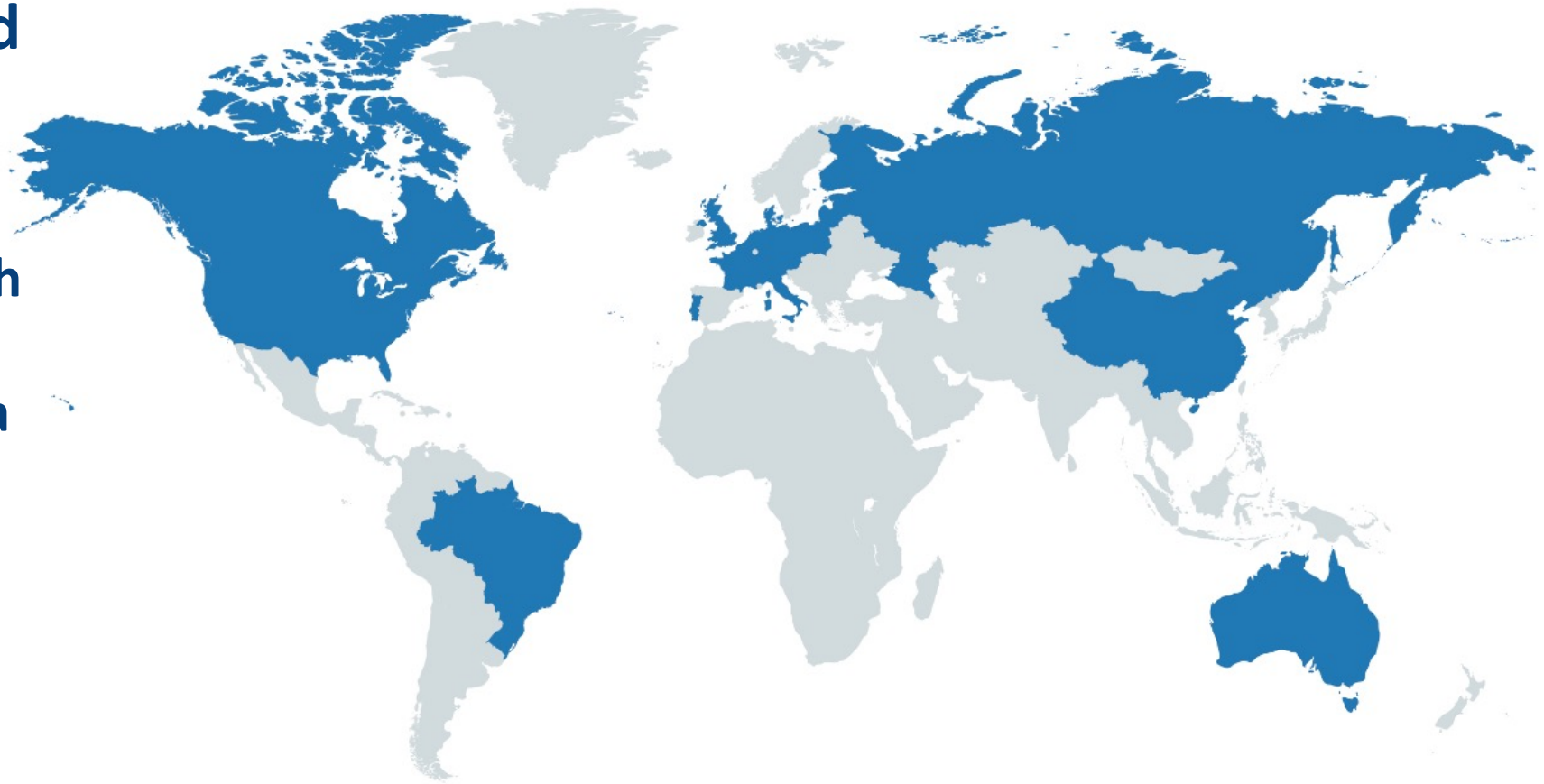
CS3

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**

From a grassroots community …

- **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

CS³

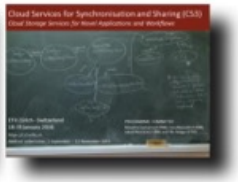
Virtual 2021

København 2020

Roma 2019

Kraków 2018

Amsterdam 2017


Zürich 2016

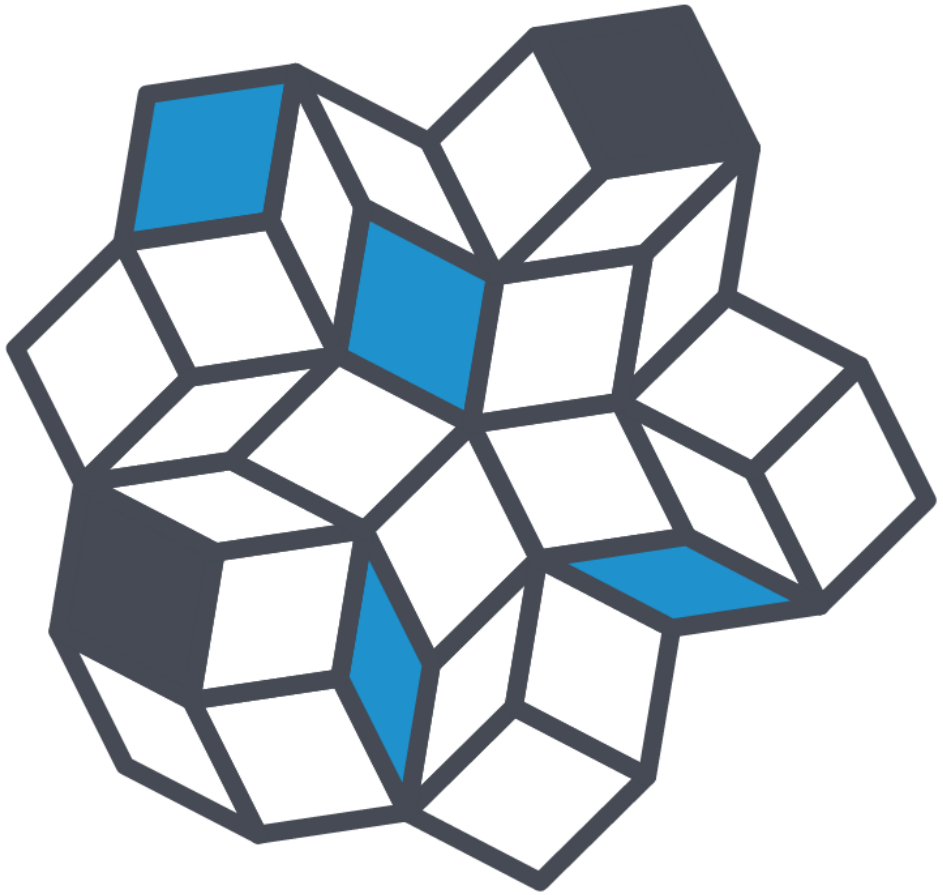
Genève 2014

- Moved up in the stack by integration with more services
  - Collaborative tools (editing)
  - Research Interfaces
    - Jupyter Notebooks
      See Marcin's presentation:
      https://indico.egi.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

What is ScienceMesh?

- 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

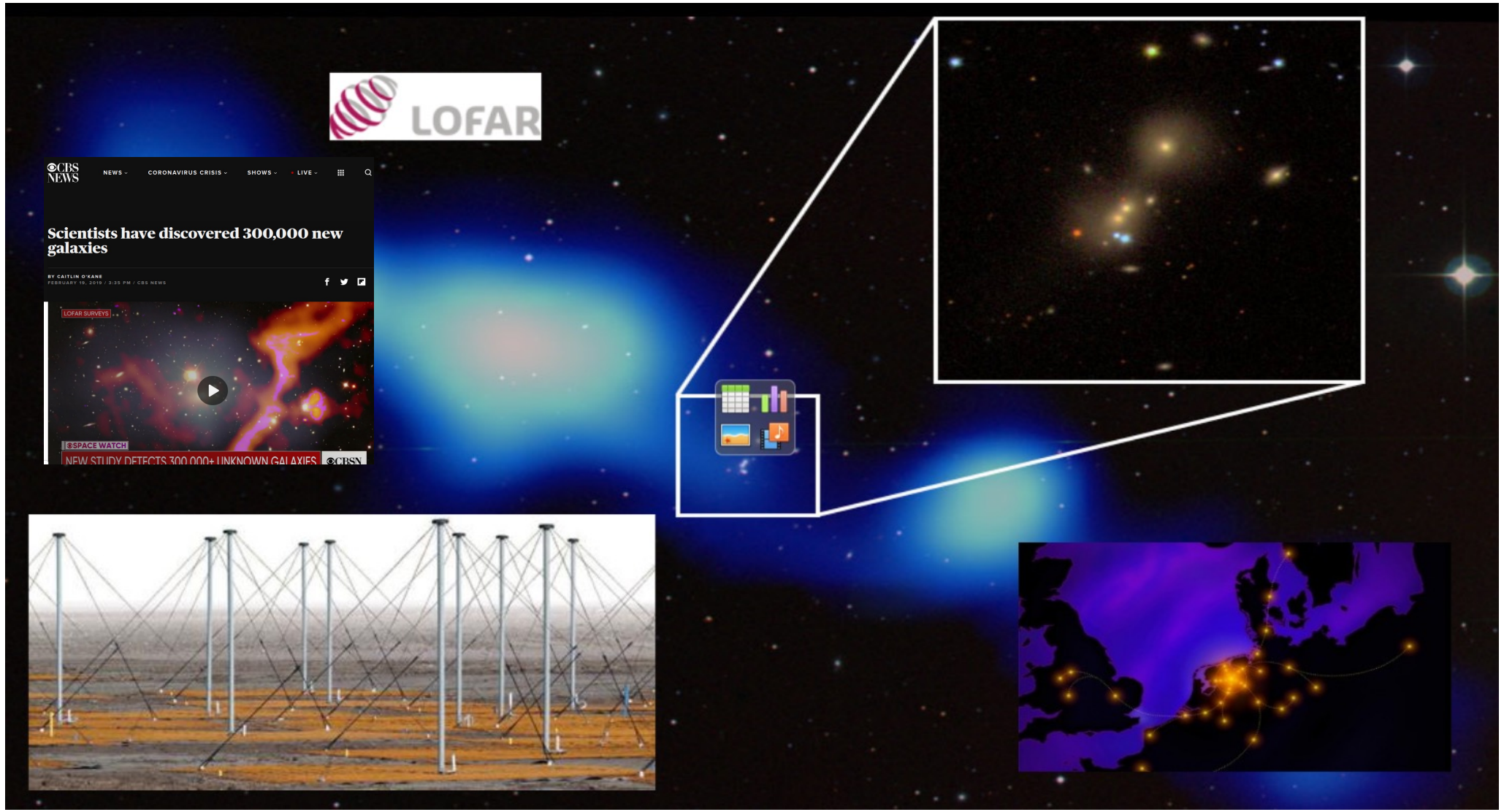
Science Mesh use cases

- Collaborative Data Science:
  - Astronomy
  - Finance
  - IoT
  - Earth Observation[1]
  - High Energy Physics[2]
  - 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

Use-case: LOFAR



Dataset transfer between research groups

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

LOFAR

LOFAR Surveys Key Science Project
Collaboration between researchers

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

On demand large dataset transfer

Science Mesh

RCLONE

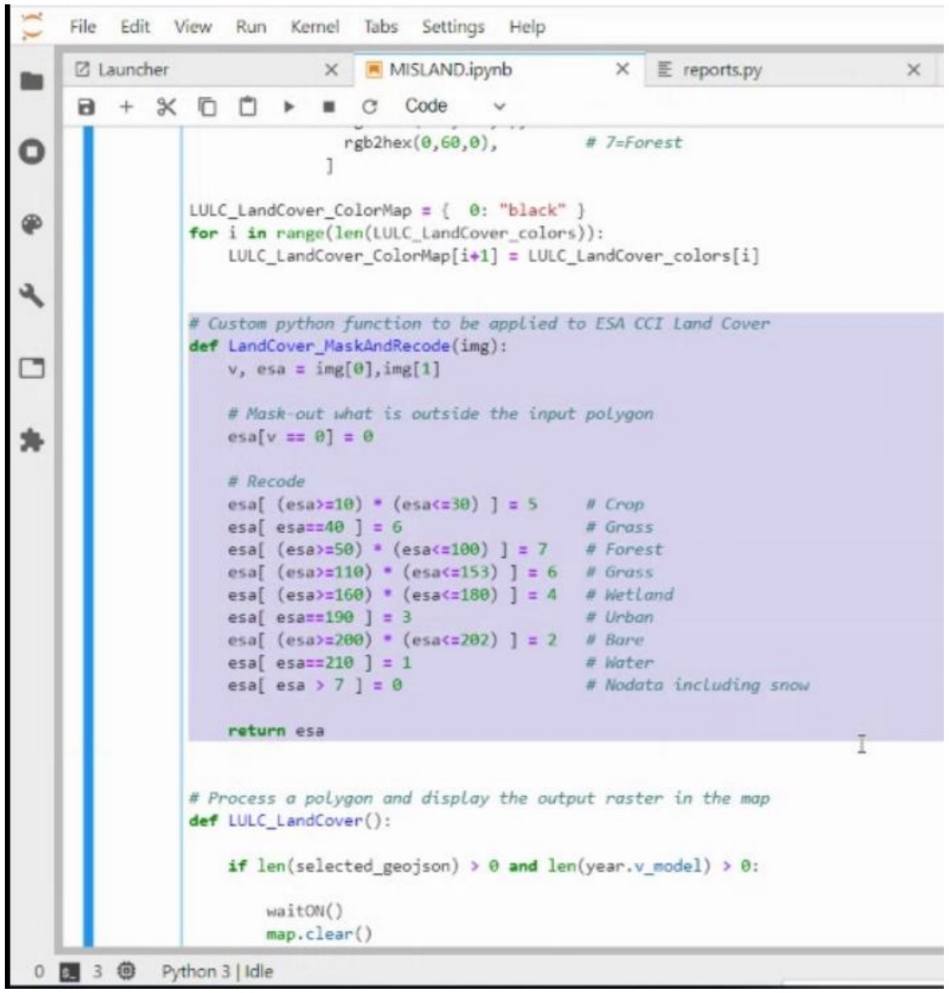
FTS

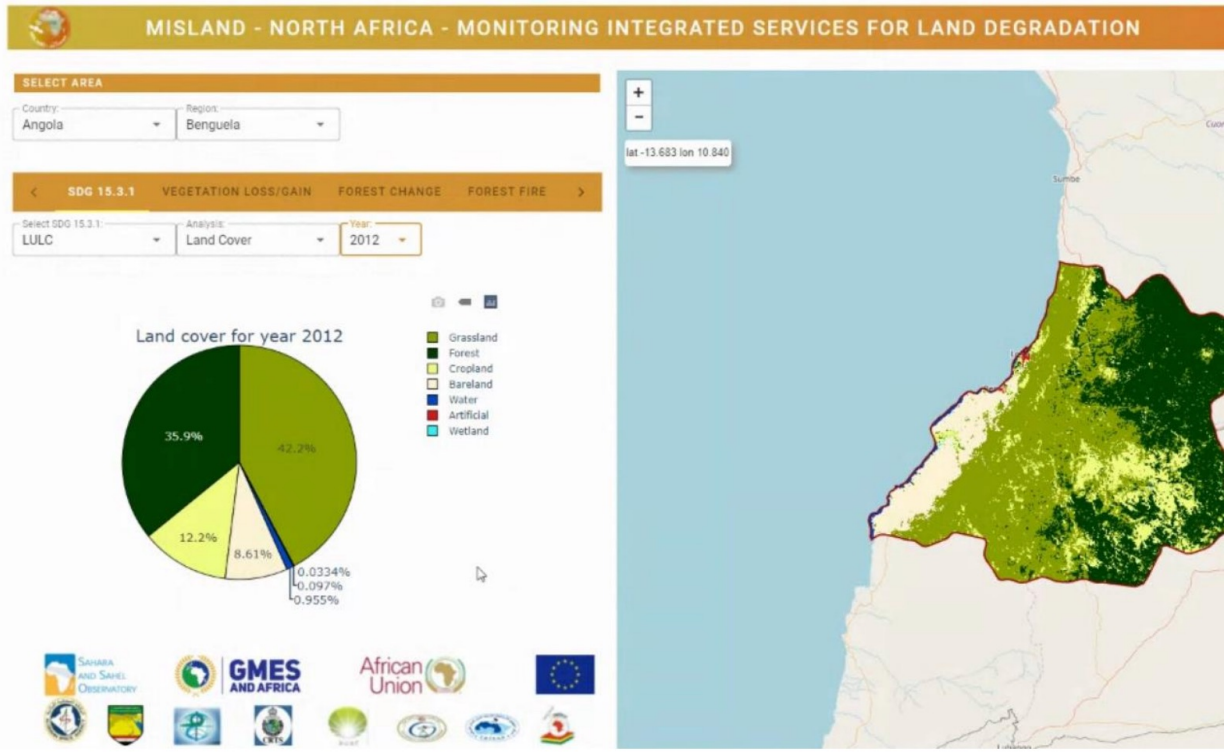
RUCIO

Data shipped to Kraków
for creating science quality images

Earth Observation use case: MISLAND

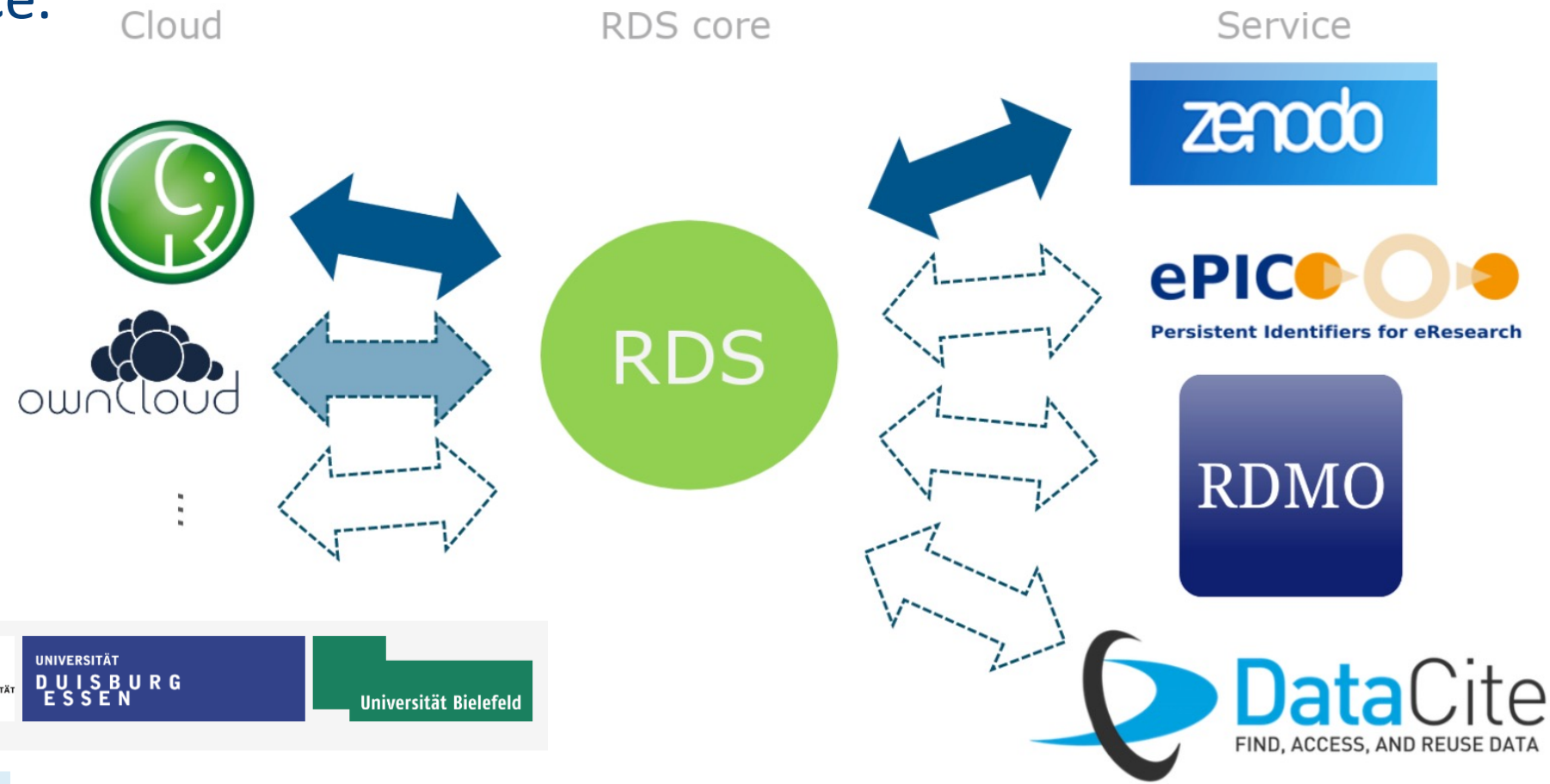
Study land degradation with Data Analytics using Jupyter





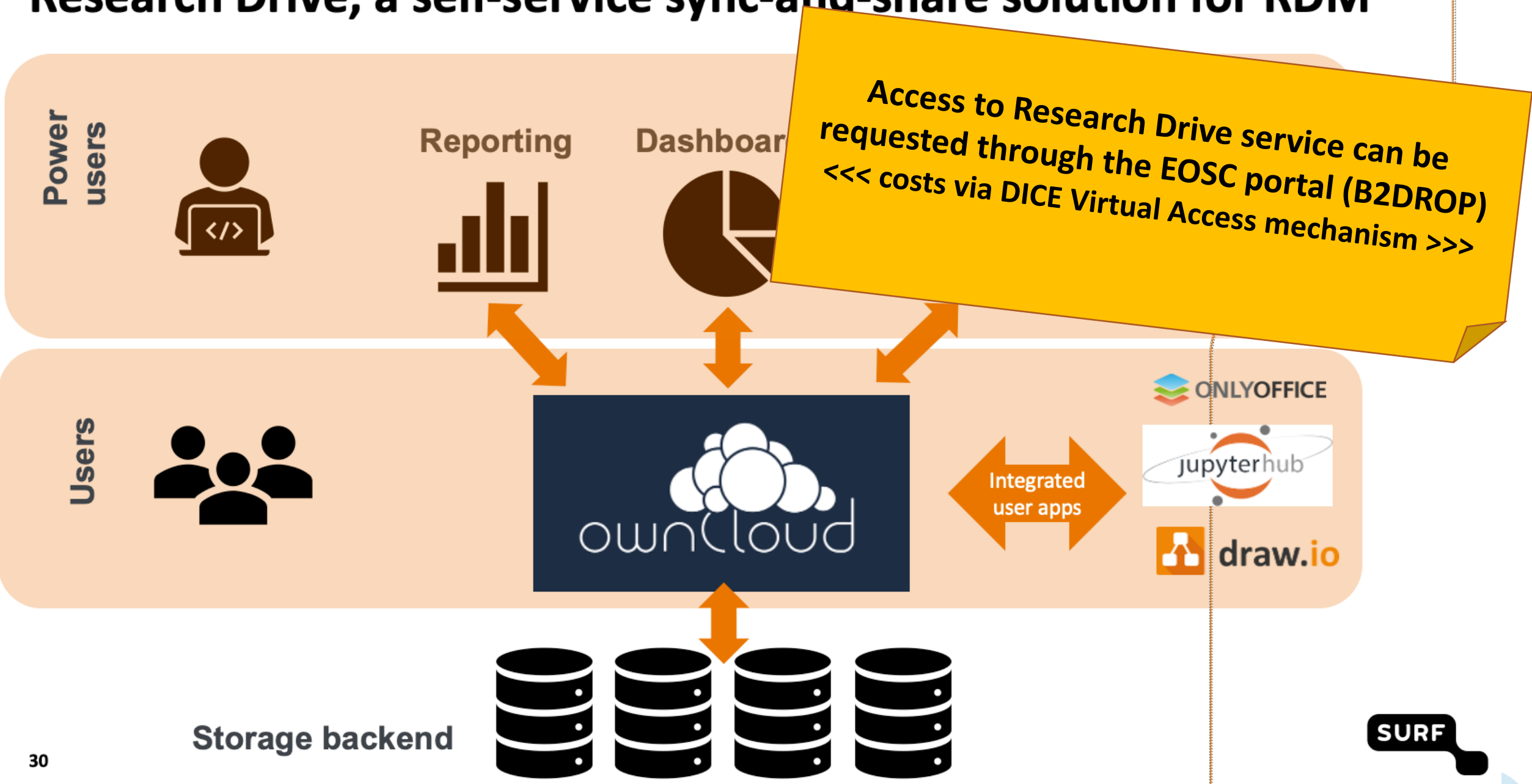
Sciebo RDS: FAIR Data

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



Westfälische Wilhelms-Universität Münster
Universität Duisburg Essen
Universität Bielefeld

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



CS3 MESH4 EOSC

Connecting European Data

Thank you!

Discover more on...

cs3mesh4eosc.eu

company/cs3mesh4eosc

CS3org

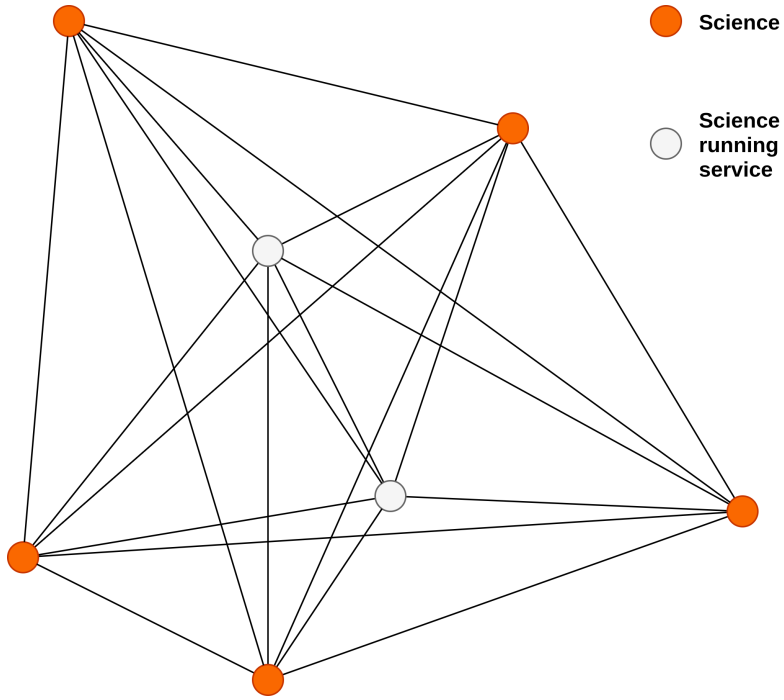
CS3MESH4EOSC Project

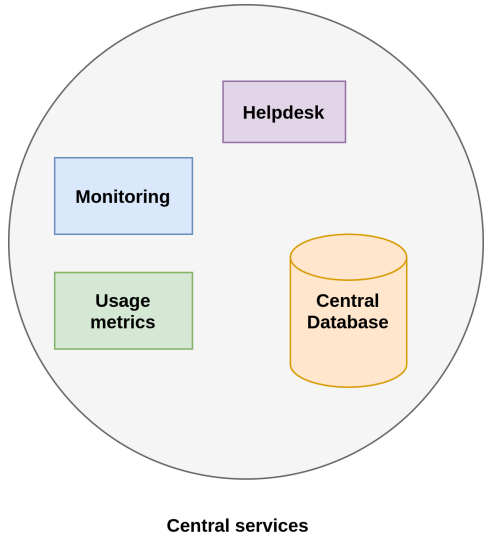
https://www.youtube.com/channel/UCHKcZEkMqXjCvc3MLFjFxbw

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

Sciebo RDS

- 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

Sustainability

- 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
    - ……

On Demand Data Transfer

On-demand data transfers

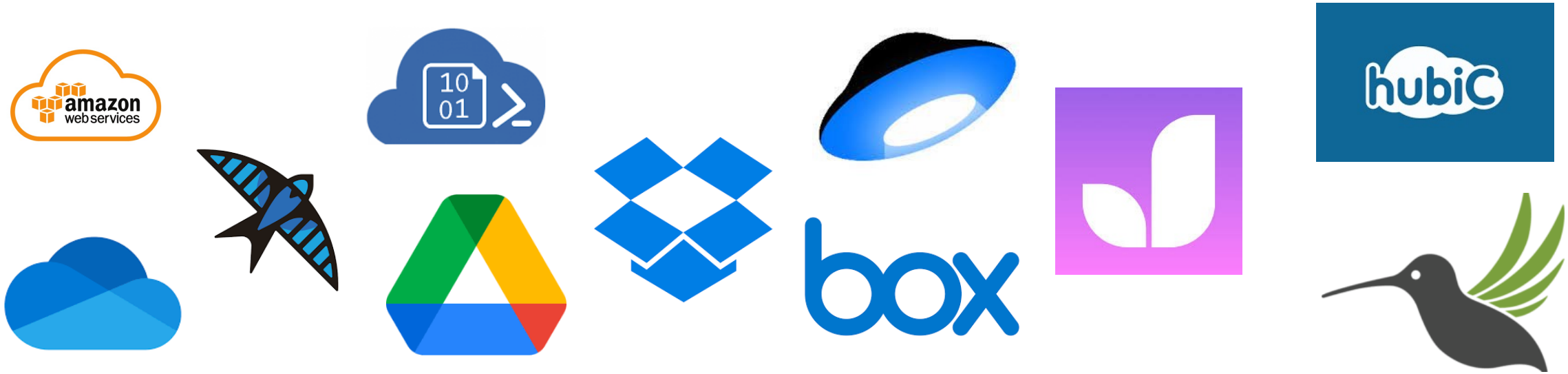
- Why on-demand data transfers?
  - Data locality may be important for some use-cases
    - Sharing of data is not good enough
      - Remote processing may cause overload on the source EFSS system
    - Larger data sets

- Support for data transfers between EFSS systems and between EFSS systems and other storage systems

On-demand data transfers

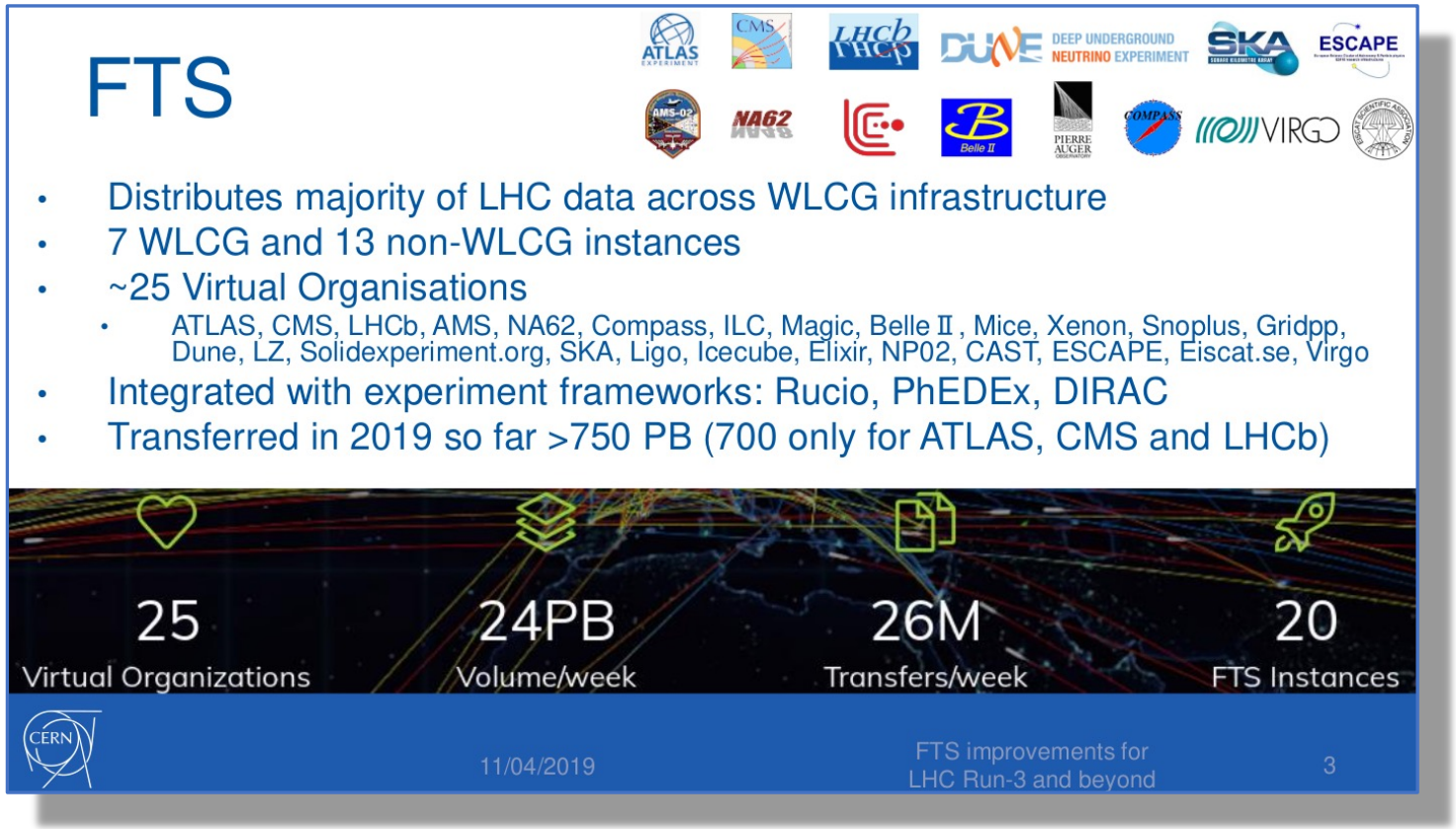
- Different 3rd party data transfer mechanisms
  - Rclone
    - For small/medium sized data transfer needs
    - Individuals or individual research groups
  - Rucio/FTS
    - Connecting Big Science with sync-and-share (EFSS)
    - Communities

- 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

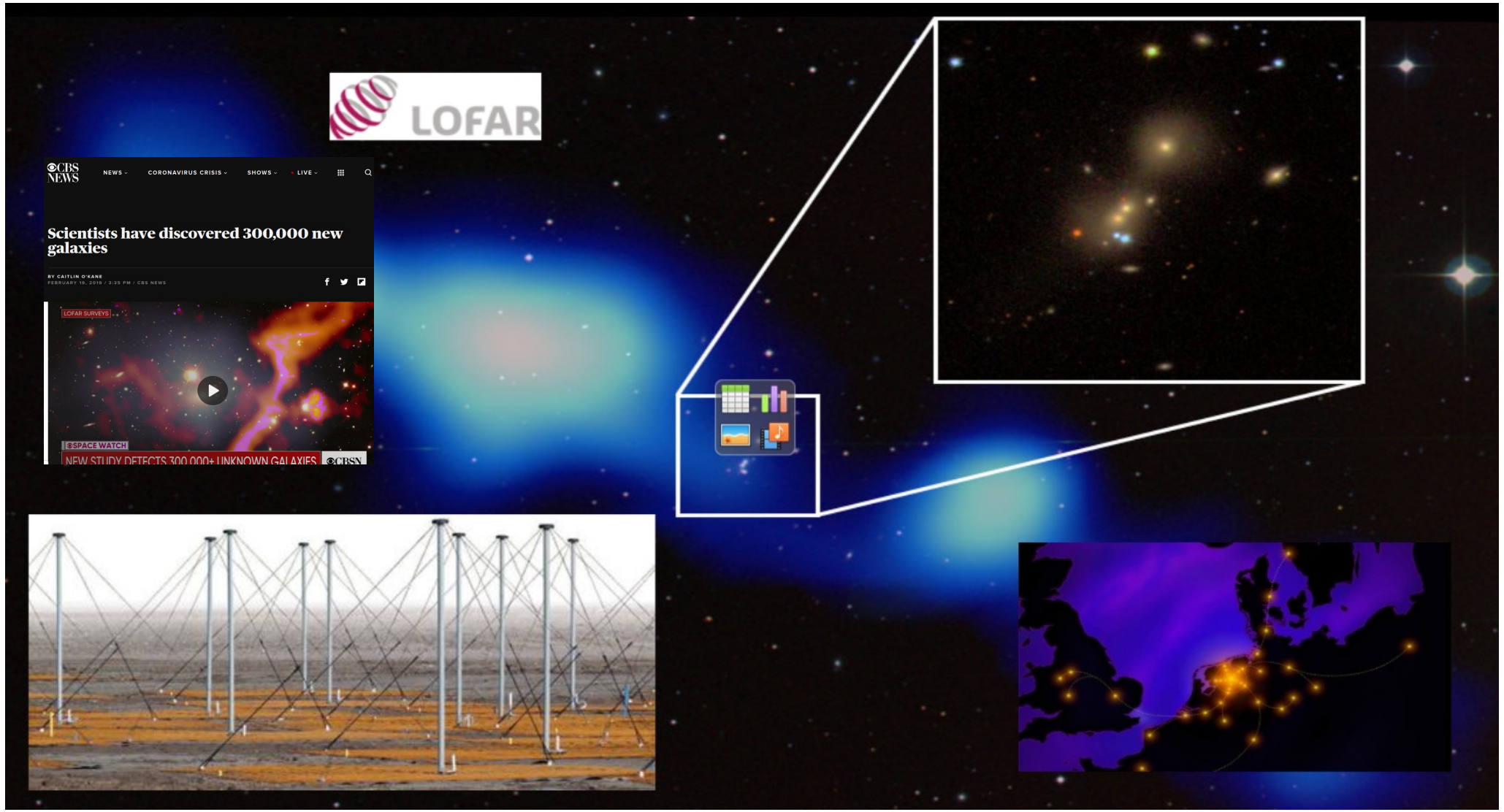
- Data transfer scheduler
- Developed by CERN to distribute LHC data
- Adopted by many other projects



- 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

RUCIO
SCIENTIFIC DATA MANAGEMENT

Use-case: LOFAR



Dataset transfer between research groups

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

LOFAR

LOFAR Surveys Key Science Project
Collaboration between researchers

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

On demand large dataset transfer

Science Mesh

RCLONE

FTS

RUCIO

Data shipped to Kraków
for creating science quality images

More information

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