CS3 MESH4 EOSC

Connecting European Data

Easy sharing through On-Demand Data Transfers

Ron Trompert and Antoon Prins

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

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

- Data transfers between sync-and-share (=EFSS) systems and between EFSS systems and other storage systems

On-demand data transfers

- On-demand data transfers in two flavours
  - Ad-hoc data transfers
    - For small/medium sized data transfer needs
    - Individual researchers or individual research groups having access to their own sync-and-share service
    - Typically local identity management system
    - Based on Rclone
  - Managed data transfers
    - Connecting Big(ger) Science with sync-and-share (EFSS)
    - Scientific communities having access to multiple storage systems and EFSS systems
    - Federated AAI
    - Based on Rucio/FTS

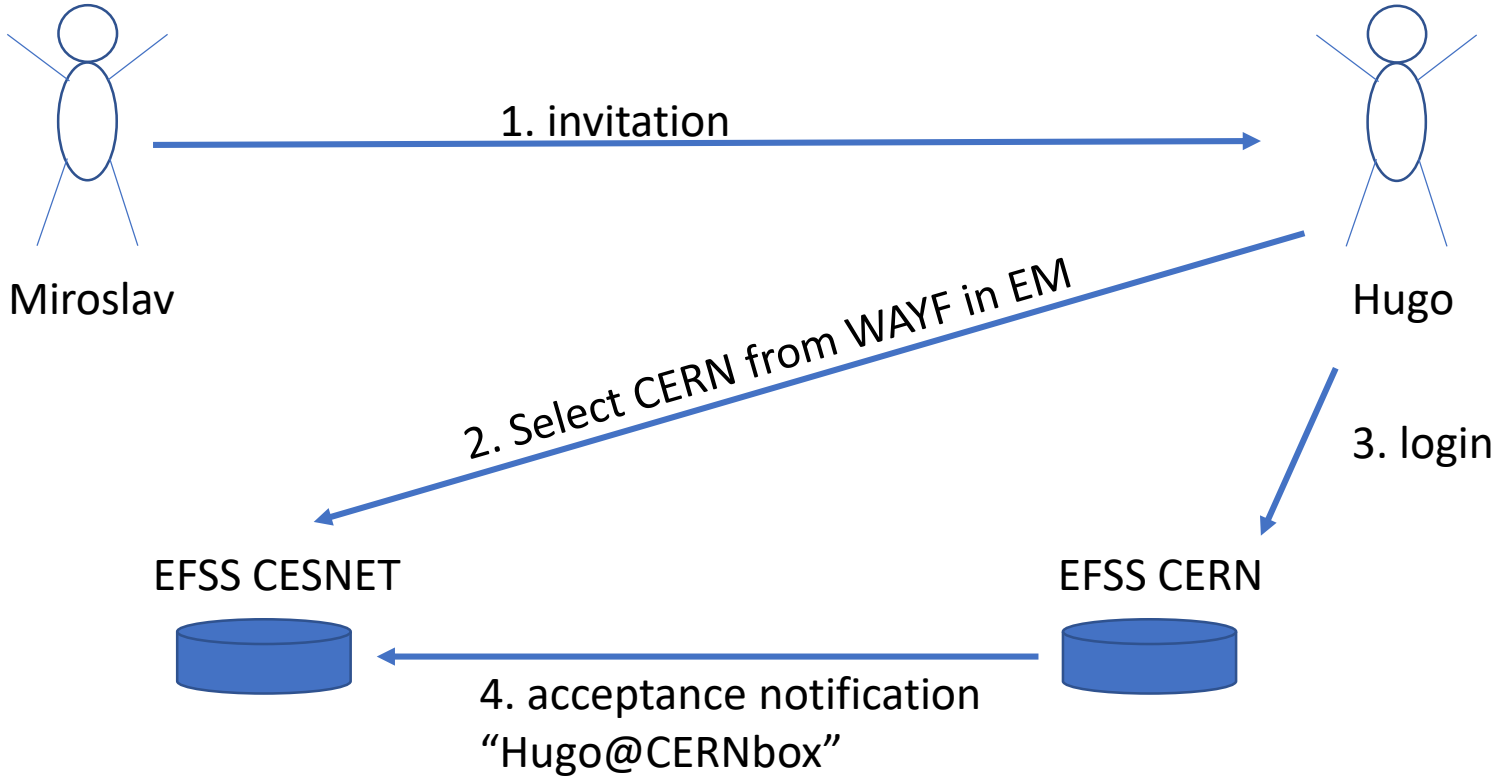
RCLONE

FTS
File Transfer Service

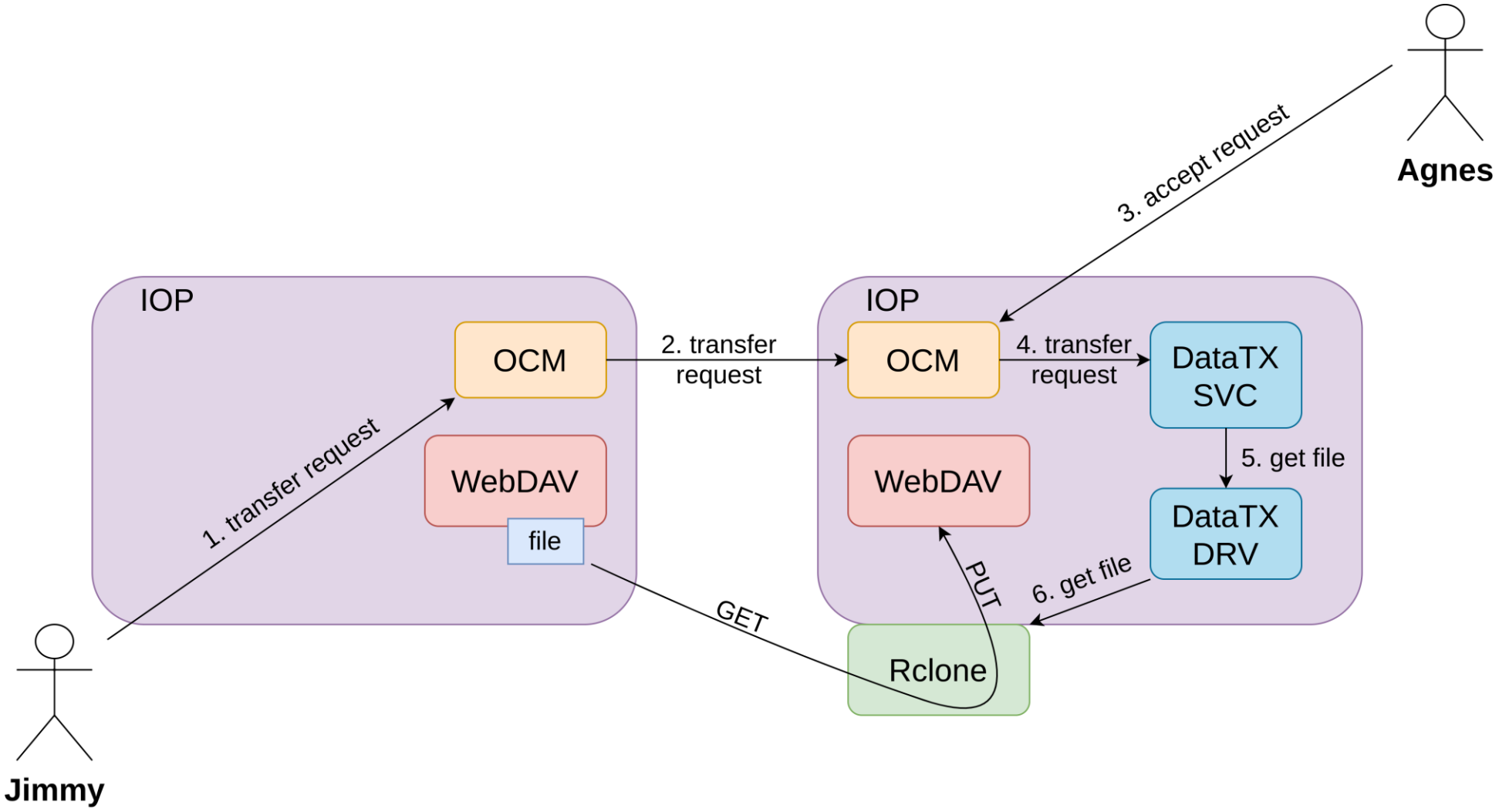
RUCIO
SCIENTIFIC DATA MANAGEMENT

Ad-hoc data transfers

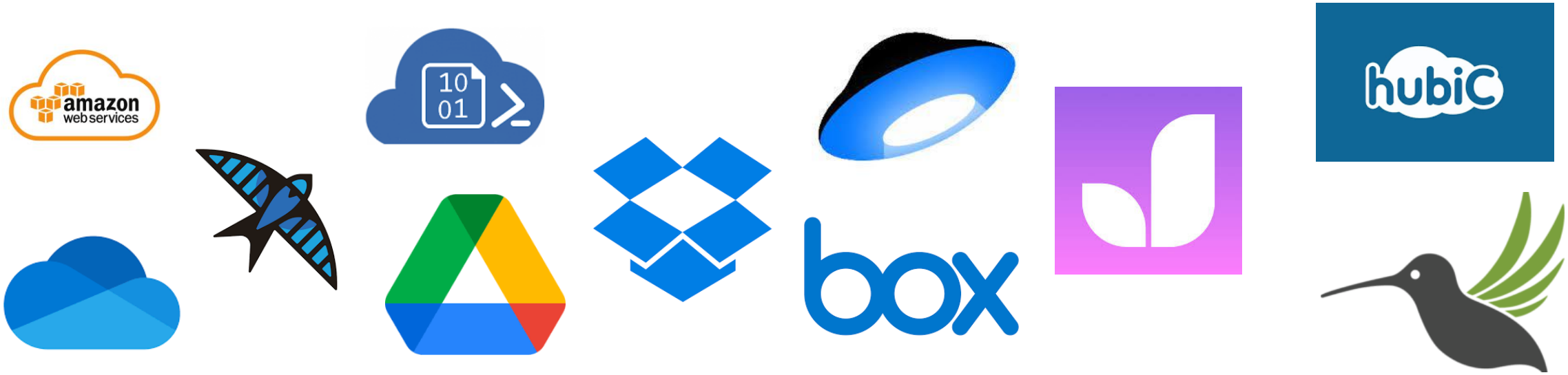
- Miroslav wants to share with Hugo
  - Invitation workflow to get name of remote sync-and-share system and user id



Ad-hoc data transfers



- 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

Managed data transfers

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

RUCIO
SCIENTIFIC DATA MANAGEMENT

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

(Slide stolen from E. Karavakis)

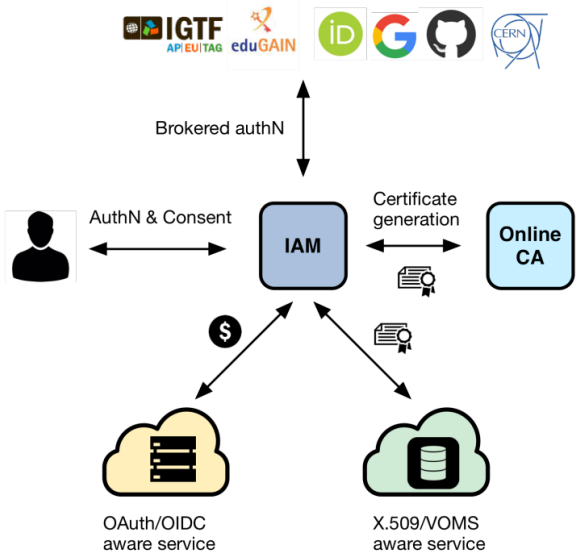
Federated Identity Management

- Indigo IAM
  - Used by WLCG and ESCAPE
  - Rucio, FTS and EFSS' use the same IAM
  - OIDC & OAUTH

INDIGO Identity and Access Management Service

An authentication and authorization service that

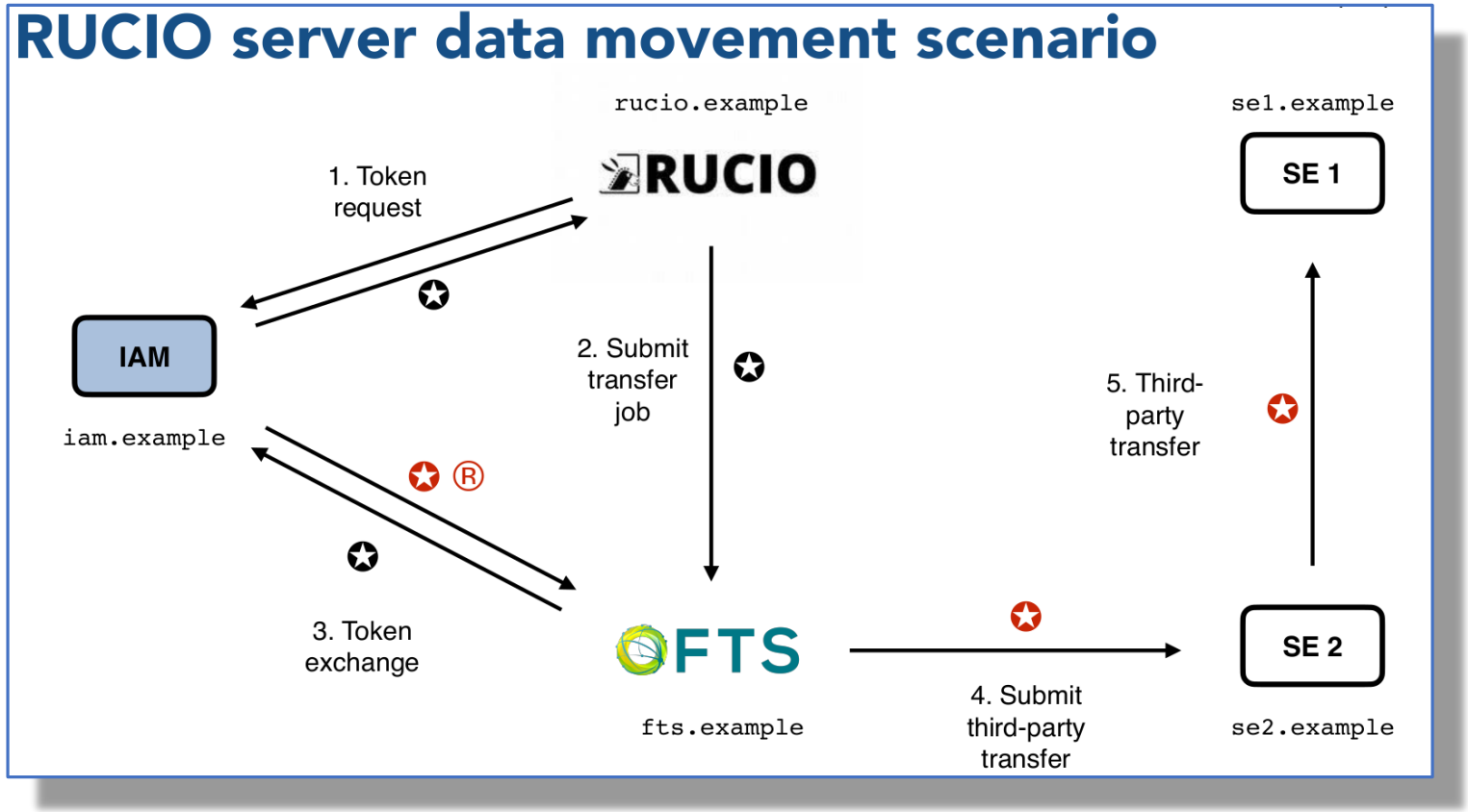
- supports **multiple authentication mechanisms**
- provides users with a **persistent, organization scoped** identifier
- exposes **identity information**, **attributes** and **capabilities** to services via **JWT** tokens and standard **OAuth & OpenID Connect** protocols
- can integrate existing **VOMS**-aware services
- supports **Web** and **non-Web access**, **delegation** and **token renewal**



(Slide stolen from A. Ceccanti)

Managed data transfers

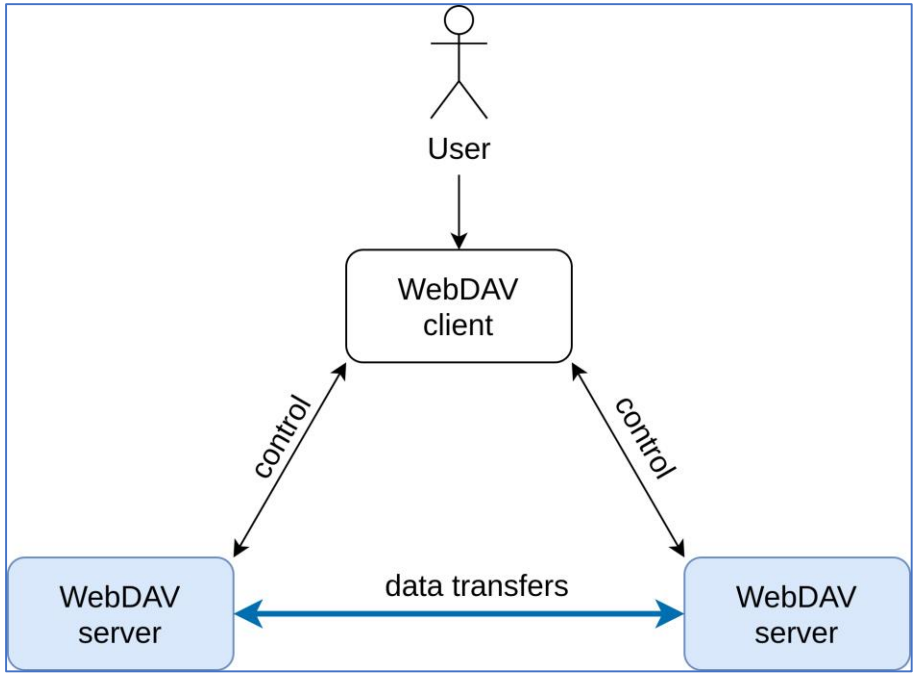
- EFSS as Rucio or FTS storage endpoint
- From EFSS issue
  - Uploads
  - Downloads
  - Metadata queries
  - …….



(Slide stolen from A. Ceccanti)

Third-party WebDAV copies

- Google summer of code project (Rahul Cauhan)
  - https://hepsoftwarefoundation.org/gsoc/2021/proposal_rucio_cernbox.html
  - WebDAV third-party transfers for EFSS



Dataset transfer between research groups

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

On demand large dataset transfer

Science Mesh

RCLONE

FTS

RUCIO

Data shipped to Kraków
for creating science quality images

LOFAR Surveys Key Science Project
Collaboration between researchers

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

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**.