



CS³
MESH⁴
EOSC

Connecting European Data



Data Transfers in the Science Mesh

Ron Trompert

SciDataCon 2021



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

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 3rd party data transfer mechanisms

Rclone

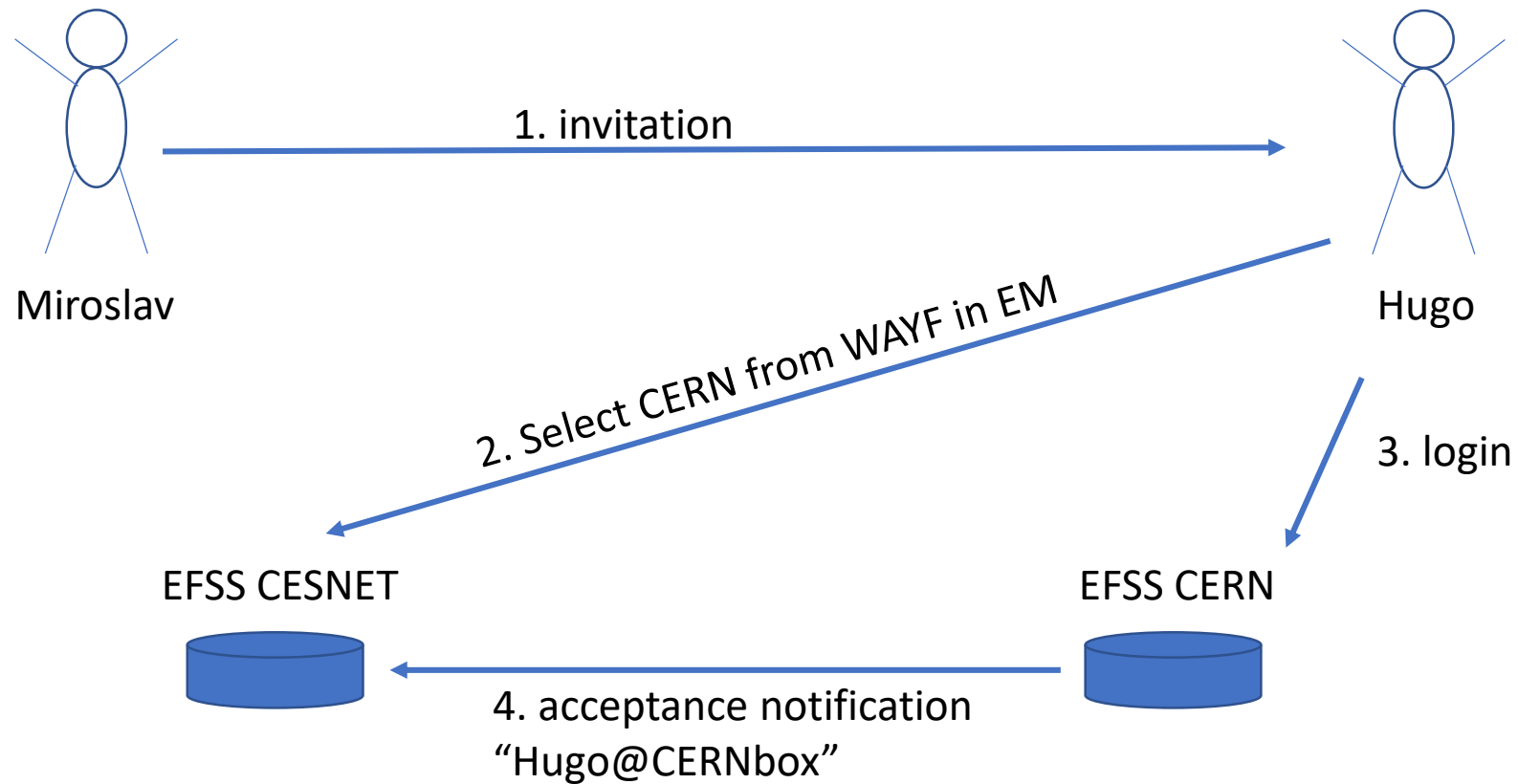
- # For small/medium sized data transfer needs
- # Individuals or individual research groups

Rucio/FTS

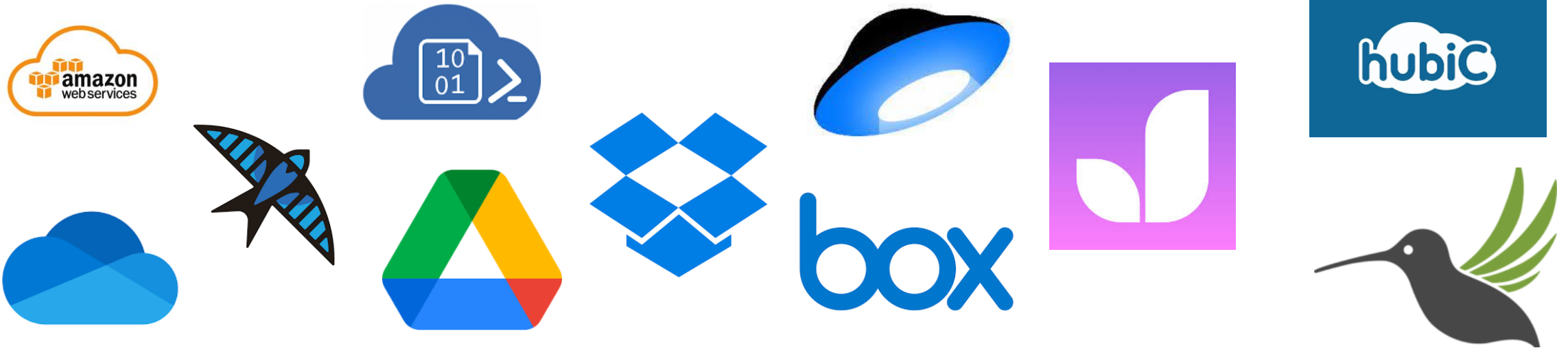
- # Connecting Big Science with sync-and-share (EFSS)
- # Communities



Establishing trust

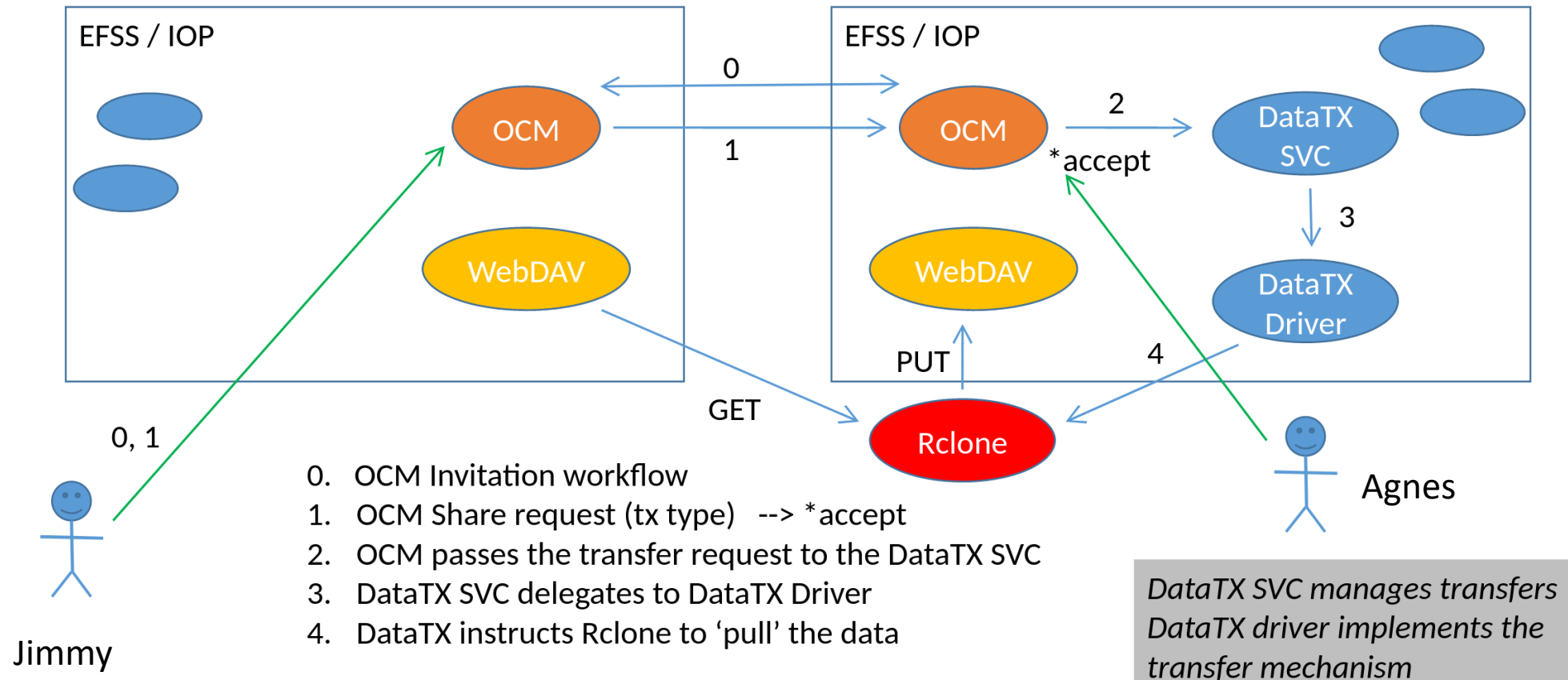


- # Commandline programme to manage files on cloud storages
- # It can interface with over 40 cloud storage products



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

Data transfers between two EFSS' based on Rclone

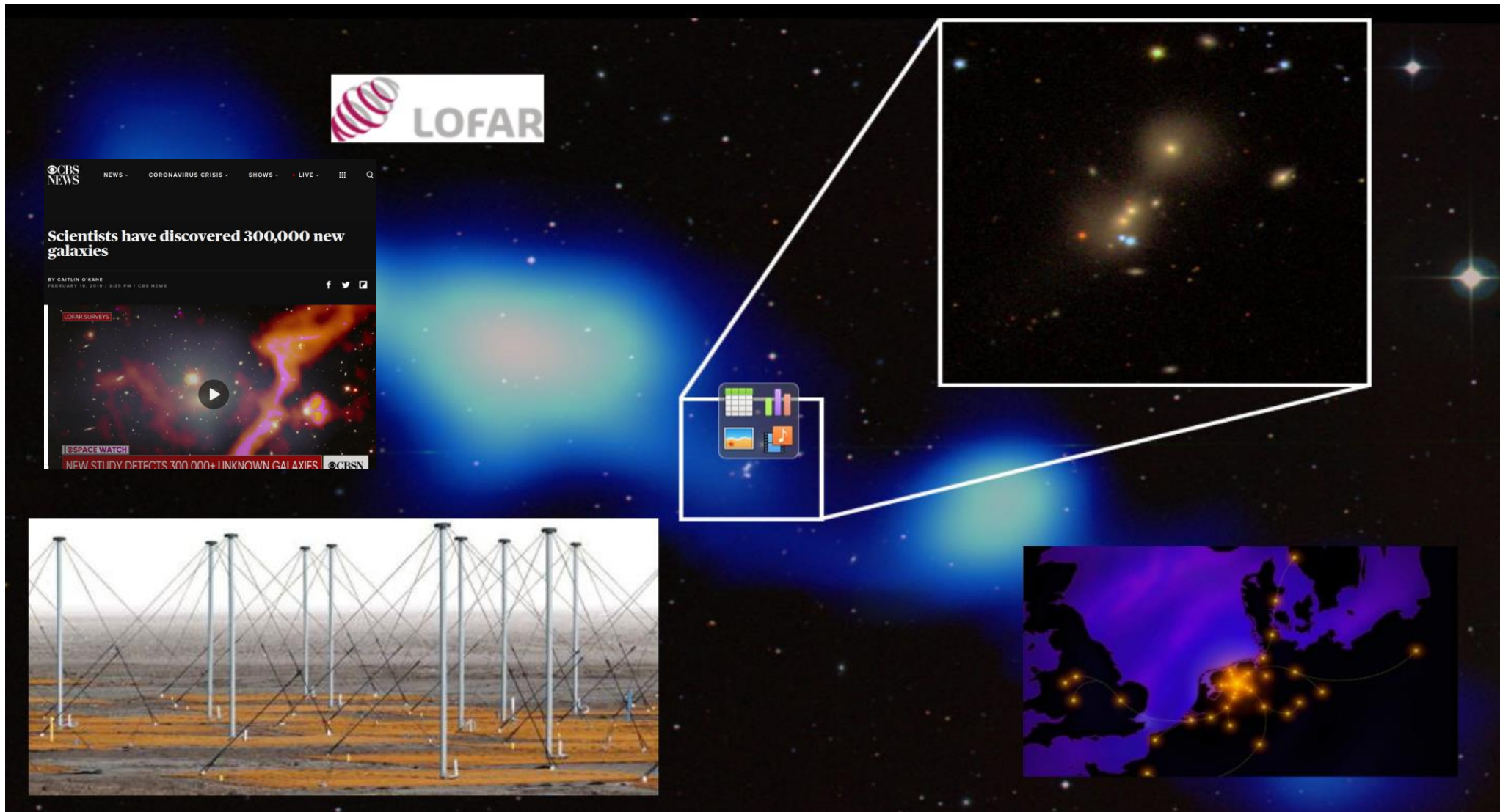


 **CS³MESH⁴EOSC**
Connecting European Data

CS3MESH4EOSC supports
on demand
data transfer between
Sync&Share systems
using the Reva IOP
interoperability layer.



<https://youtu.be/m1PlzTrazoA>



The image is a collage illustrating the LOFAR use case. It features several key elements:

- Top Left:** A screenshot of a CBS News article titled "Scientists have discovered 300,000 new galaxies" by Caitlin O'Keane, dated February 19, 2019. The article includes a video player and social media sharing icons.
- Top Center:** The LOFAR logo, which consists of a stylized antenna structure and the text "LOFAR".
- Top Right:** A zoomed-in view of a galaxy cluster, showing numerous bright, colorful stars and galaxies against a dark background.
- Center:** An icon representing a data dashboard or analytics tool, featuring a grid, bar chart, and other data visualization elements.
- Bottom Left:** A photograph of the LOFAR antenna array, showing a series of tall, silver metal poles connected by a network of cables, situated in a flat, open landscape.
- Bottom Right:** A map of Europe with several glowing orange dots representing LOFAR radio telescope stations, connected by thin lines.



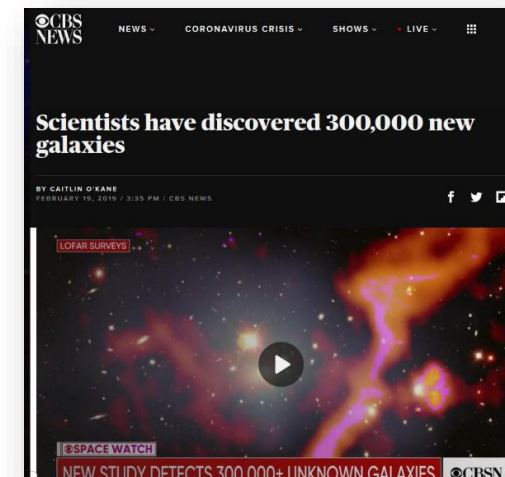
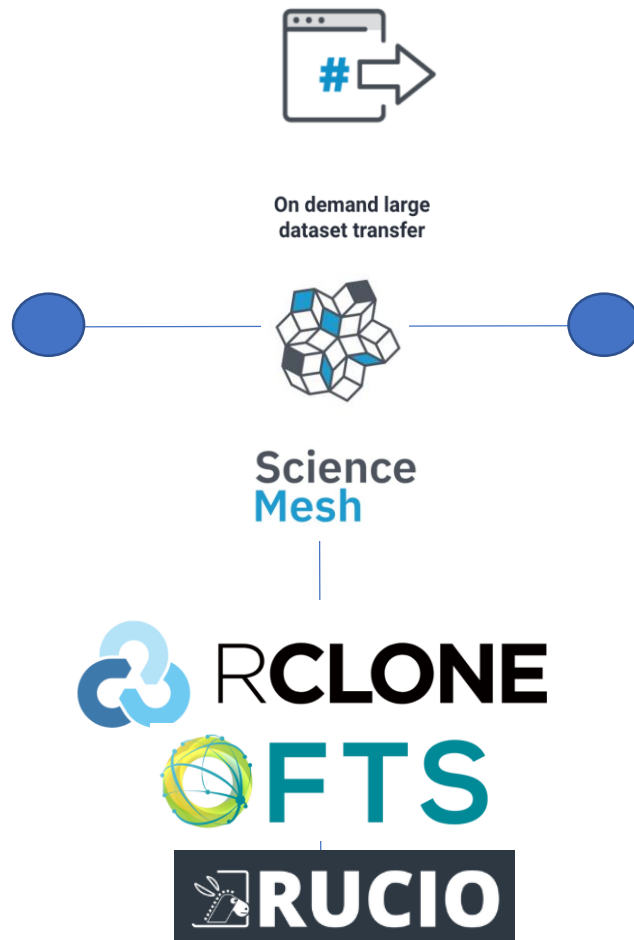
*Data stored at SURF and FZJ.
Initially processing (64x reduction).*



LOFAR Surveys Key Science Project
Collaboration between researchers

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

Dataset transfer between research groups



*Data shipped to Kraków
for creating science quality images*



- # 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)



11/04/2019

FTS improvements for
LHC Run-3 and beyond

3



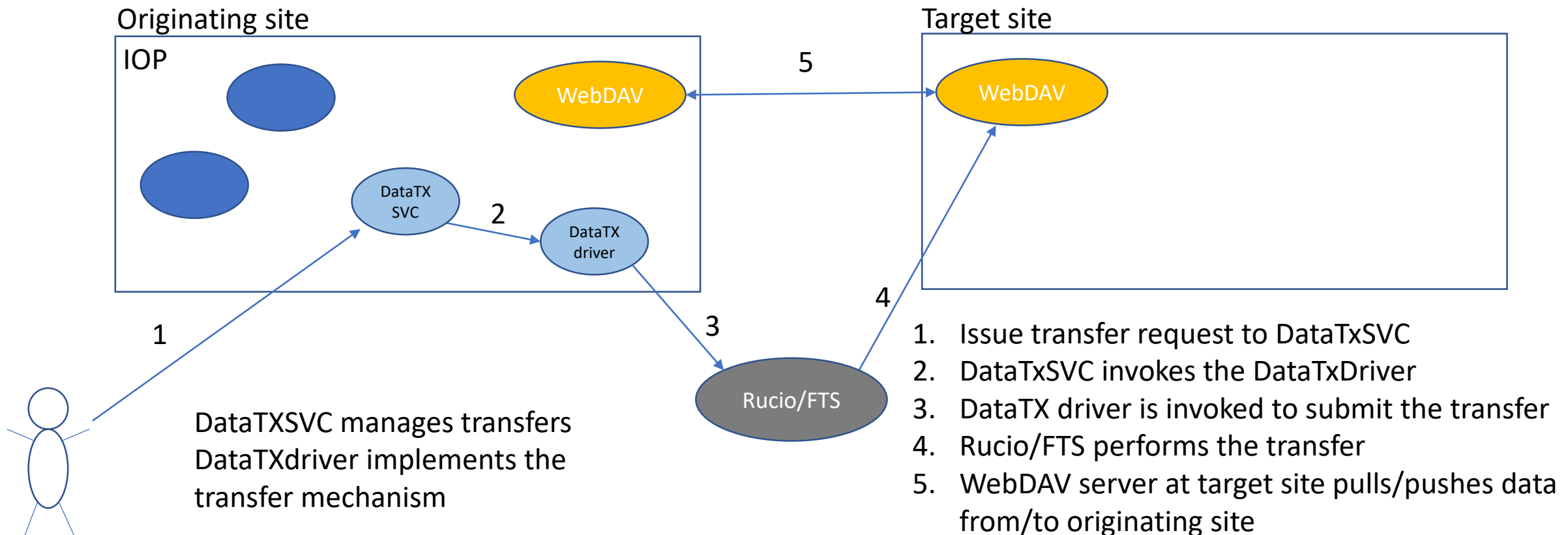
- # Data management
- # Developed by ATLAS
- # What does it do?
 - # Managing data collections
 - # Metadata management
 - # Replicating data to registered storage elements
 - #
- # Uses the FTS for data transfers
- # Not only used by ATLAS, but also by CMS, SKA and many others





- # Work in progress
- # Google summer of code project (Rahul Cauhan)
 - # https://hepsoftwarefoundation.org/gsoc/2021/proposal_rucio_cernbox.html
 - # WebDAV third-party transfers for EFSS
- # EFSS site as a:
 - # FTS WebDAV endpoint
 - # Rucio RSE
- # Federated Identity Management
 - # IAM (Indigo Datacloud)
 - # EFSS', FTS and Rucio use the same IAM

Data transfers between an EFSS site and a regular or EFSS site using Rucio/FTS



1. Issue transfer request to DataTxSVC
2. DataTxSVC invokes the DataTxDriver
3. DataTX driver is invoked to submit the transfer
4. Rucio/FTS performs the transfer
5. WebDAV server at target site pulls/pushes data from/to originating site



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**.