

Bringing HPC clusters into the Science Mesh

Raimund Vogl, Holger Angenent, Daniel Müller

WWU IT, University of Münster, Germany

rvogl@uni-muenster.de



Traditional Data Transferring Mechanism – the problem

- Scientists use HPC (High Performance Computing) resources at multiple institutions/sites
- Research collaborations exchange large data sets across institutions
- Transferring data to and from HPC clusters (at different sites) is normally done via command line on port 22 (ssh)
- This also allows interactive access to the cluster
- For security reasons, access to port 22 is nowadays restricted (VPN required), so different clusters cannot see each other
- Data transfer becomes more difficult (storing data temporarily on third computers might not work)

What is the Science Mesh?

- Born out of 3.5-year EOSC (European Open Science Cloud) EU project lead by CERN
- Based on Open Standards and Open Source Software
- Federated research space for Europe
- Interoperability Platform to develop and connect new applications
- Project name: CS3MESH4EOSC <https://cs3mesh4eosc.eu/>
- Science Mesh is the product name: <https://sciencemesh.io>
- Make data exchange for research easy (again) to support FAIR (Findable, Accessible, Interoperable, Reusable) data usage



What is the Science Mesh?

- ScienceMesh = OCM (Open Cloud Mesh) + discovery mechanism
- **Additional Services for (data) science use:**
- JupyterLab integration
- Collaborative editors (like Collabora, CodiMD...)
- Open data systems (<https://www.research-data-services.org/>)
- Large data transfers (based on Rclone, Rucio and FTS)



Data Science
Environments



Open Data Systems



Collaborative
Documents



On demand large
dataset transfer



Science
Mesh

Data transfers

- Point to point (Rclone)
- PcC successful
- User Interface: WIP
- Between VOs
- FTS <-> Reva: working
- Rucio <-> Reva: WIP



How to join the Science Mesh

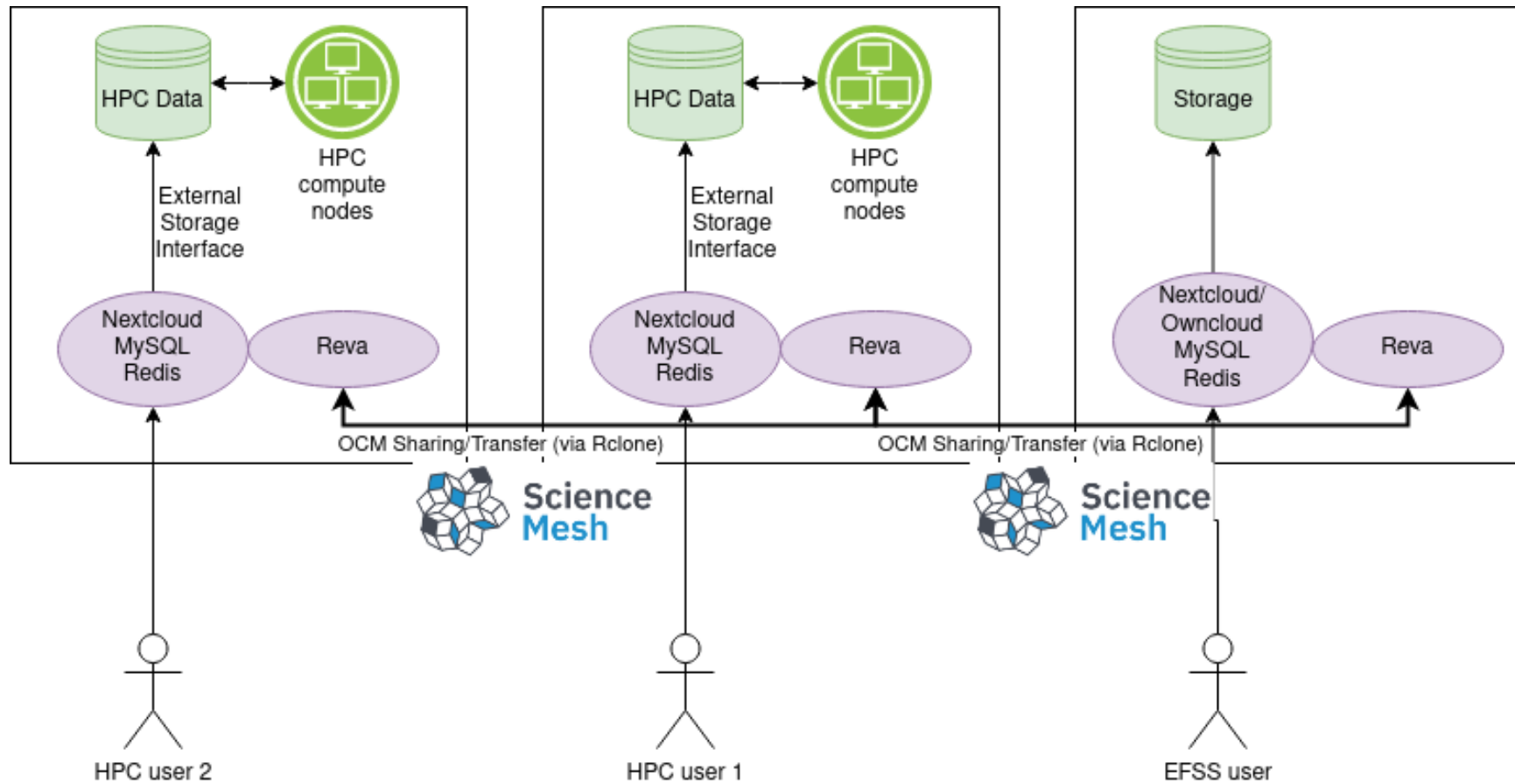
- Read and follow <https://developer.sciencemesh.io/docs/how-to-join-sciencemesh/>
- **Science Mesh node consists of:**
 - Owncloud/Nextcloud +
 - front end app +
 - Reva interoperability daemon



How can the Science Mesh help for HPC data transfers?

- If only your HPC filesystem would be an EFSS...
- Install Nextcloud on Headnode
- Export subdirectory of filesystem via samba
- Mount samba share via external_storage mechanism in Nextcloud
- Make this Nextcloud part of the Science Mesh

Final data transfer setup



How can the Science Mesh help for HPC data transfers?

Desired final state:

- Users can access the cluster's filesystem (or a part of it for performance reason) via a web interface
- There they can push upload and download files directly
- Data transfers to other Science Mesh systems can be orchestrated

Thank you for your attention

Contact: rvogl@uni-muenster.de