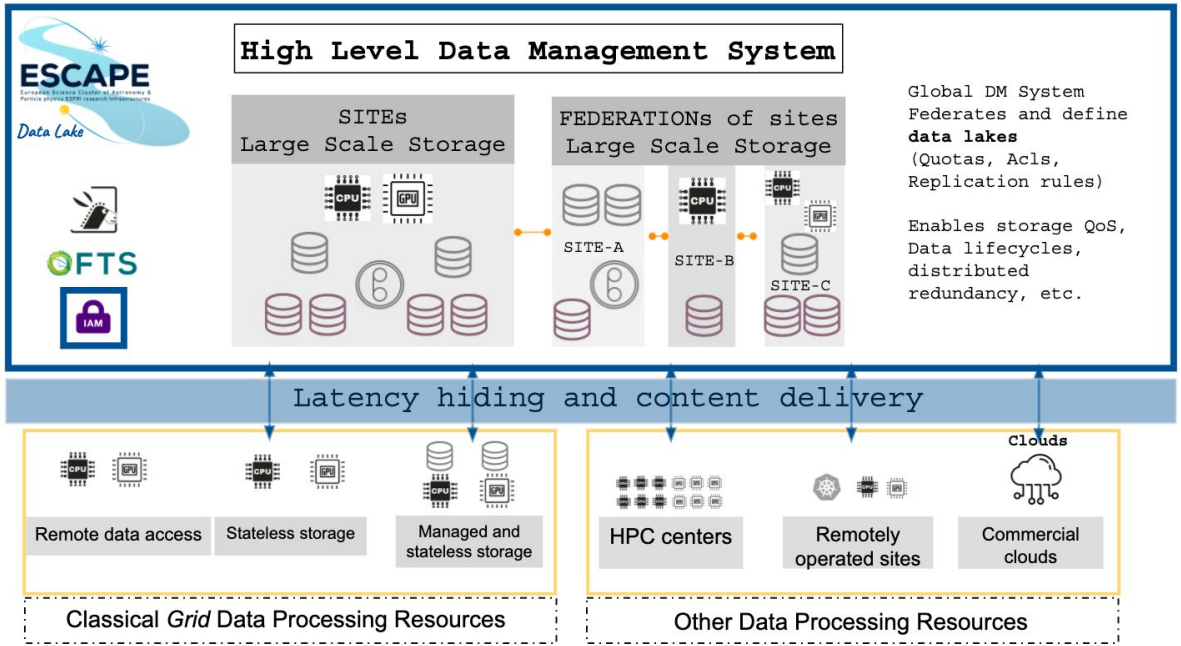
ESCAPE and Science Mesh

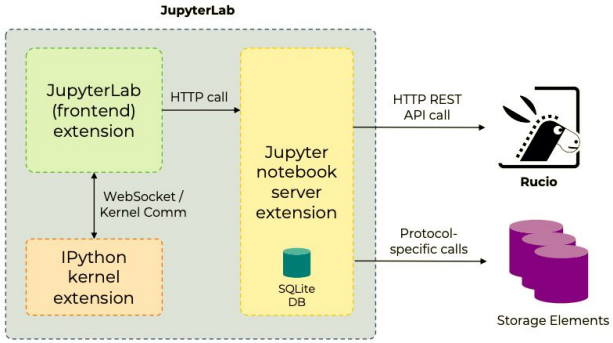
Xavier Espinal (CERN, ESCAPE WP2 lead)

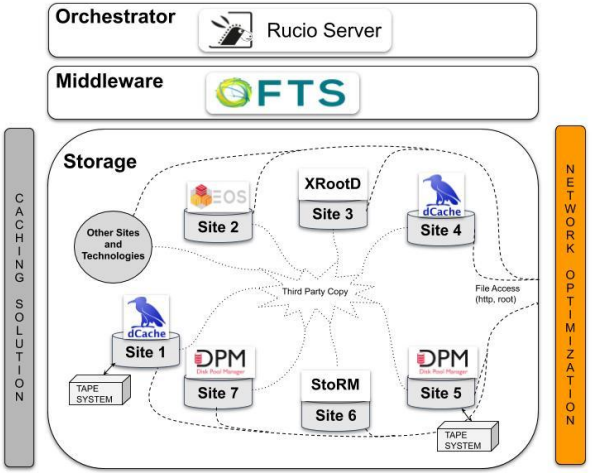
Science Mesh Workshop - Panel discussion 26 Jan 2022

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

ESCAPE and Science Mesh: The ESCAPE Scientific-Data Lake







- **Bringing big science data to the researcher fingertips**: data lake integration with notebooks and analysis platforms
- Data lake model and tools **evolving and being adopted beyond particle physics** covering different use cases and needs. Possibilities to **further extend capabilities**, ie. via CS3MESH4EOSC/ScienceMesh

Funded by the European Union's
Horizon 2020 - Grant N° 824064

ESCAPE and Science Mesh: a possible collaboration scope

- **Goal**: Provide a **Data Management and Access *facilitating*-service**, in the EOSC portal and/or an integral solution deployable for projects, experiments or collaborations.
- **Why?** Bridge diverse scientific data (size does not matter) with researchers, outreach activities and open science initiatives.
  - Data Lake services for the heavy lifting done by experiment experts (ie. data injection, policies and rules)
  - ScienceMesh putting data at the service of the population, ie. via notebooks and integrated with sync&share tools to boost scientific collaborations, and home directories.
- **What?** Provide hints in the form of PoC to put forward these ideas and offer the possibility to connect ScienceMesh with the Data Lake infrastructure(s)
- **How**? Enabling Science Mesh activities in the ESCAPE Data Lake.
  - Integrated access to the Data Management system (Rucio) and the Data Transfer/Movement service (FTS)
  - Certify transversal token-based AAI (IAM based): ensure IAM-issued tokens are trusted by Science Mesh nodes and coherent across the full “data chain”
- **A possible catalizer**? LOFAR very active in ESCAPE, production workloads certified from data injection to analysis. Also with applied use-case in Science Mesh. Sounds like reasonable PoC to start with?

Funded by the European Union’s Horizon 2020 - Grant N° 824064

FAIR data ecosystem and repository certification

Science Mesh Workshop 'Global Platform for Scientific Collaboration' -
Session '*Scientific disciplines embracing no border Research Environment thanks to Science Mesh*'

Mari Kleemola, SSHOC (CESSDA and FSD/Tampere University)

January 26, 2022

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Project: SSHOC

social sciences & humanities open cloud

Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:

Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 45

(20 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Duration: 40 months

(January 2019 – 30 April 2022)

Project budget:

€ 14,455,594.08

Project website:

www.SSHopencloud.eu

Objectives:

- creating the social sciences and humanities **(SSH)** part of European Open Science Cloud **(EOSC)**
- maximising **re-use** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

Challenges

- Data is an asset that needs to be managed in order to maximise its value
- Varying level of maturity of services and repositories
- Sharing of responsibilities unclear
- Lacking governance, business models, metadata, documentation

How SSHOC Meets the Needs of Researchers

SSHOC Certification Support: Enabling open research with enhanced access and reuse of our rich European cultural and scientific heritage.

A Conversation with **Tomasz Parkola**

Meet our SSHOC Champion

I'm Tomasz Parkoła. I work at Poznań Supercomputing and Networking Center where I manage the ICT R&D department focused on digital repositories and knowledge platforms. Among various activities, we provide e-infrastructure and software technologies for nationwide Digital Repository of Scientific Institutes (DRSI). One of the long-term goals of DRSI is to provide trusted datasets and digital resources for end-users. To ensure we have everything in place to enable open research, we're working towards a CoreTrustSeal certification. This is where the SSHOC Certification Support comes into play as a much-appreciated service on our side.

SSH Research Context

In PSNC, I work on innovative software technologies for digital repositories, data aggregation and digital humanities. Together with cultural heritage professionals and digital humanities researchers we define new ways of accessing and reusing our rich European cultural and scientific heritage. Thanks to these efforts, various research datasets, digital books and paintings as well as georeferenced resources are easily accessible online.

The Research Challenge: Accessing Trustworthy Data Sources

In my opinion trustworthiness is one of the most important aspects of every digital repository. It is especially helpful in the context of research workflows, data citation, linked open data or even advanced data analytics. Without reliable, safe and secure sources of data it is impossible to conduct open research activities and effective cross-national and cross-domain collaboration.

SSHOC: Making the Difference

We found the SSHOC Certification Support service very helpful when preparing the CoreTrustSeal application for Digital Repository of Scientific Institutes. The support service offered us a workshop, individual online discussions, comments, hints and suggestions on how to improve and make our CTS application better. Without that, the preparation process would have been much longer and required a lot more effort.

Tips and Tricks for a CoreTrustSeal

If you're thinking about preparing a CoreTrustSeal, then organise a team composed of data curators, data users and IT specialists. They all will need to fill out the application form. Getting advice from preservation professionals, such as those involved in SSHOC Certification Support, is a key to moving through the application process smoothly and seamlessly. Their experience and valuable input will guide you through the CTS requirements, help better understand them and improve your internal data-related processes.

Try the SSH Open Marketplace Beta version!

Have other research needs you'd like to discuss?

LET US KNOW HERE

Community

www.sshopencloud.eu
info@sshopencloud.com
@SSHOpenCloud
in/SSHOpenCloud

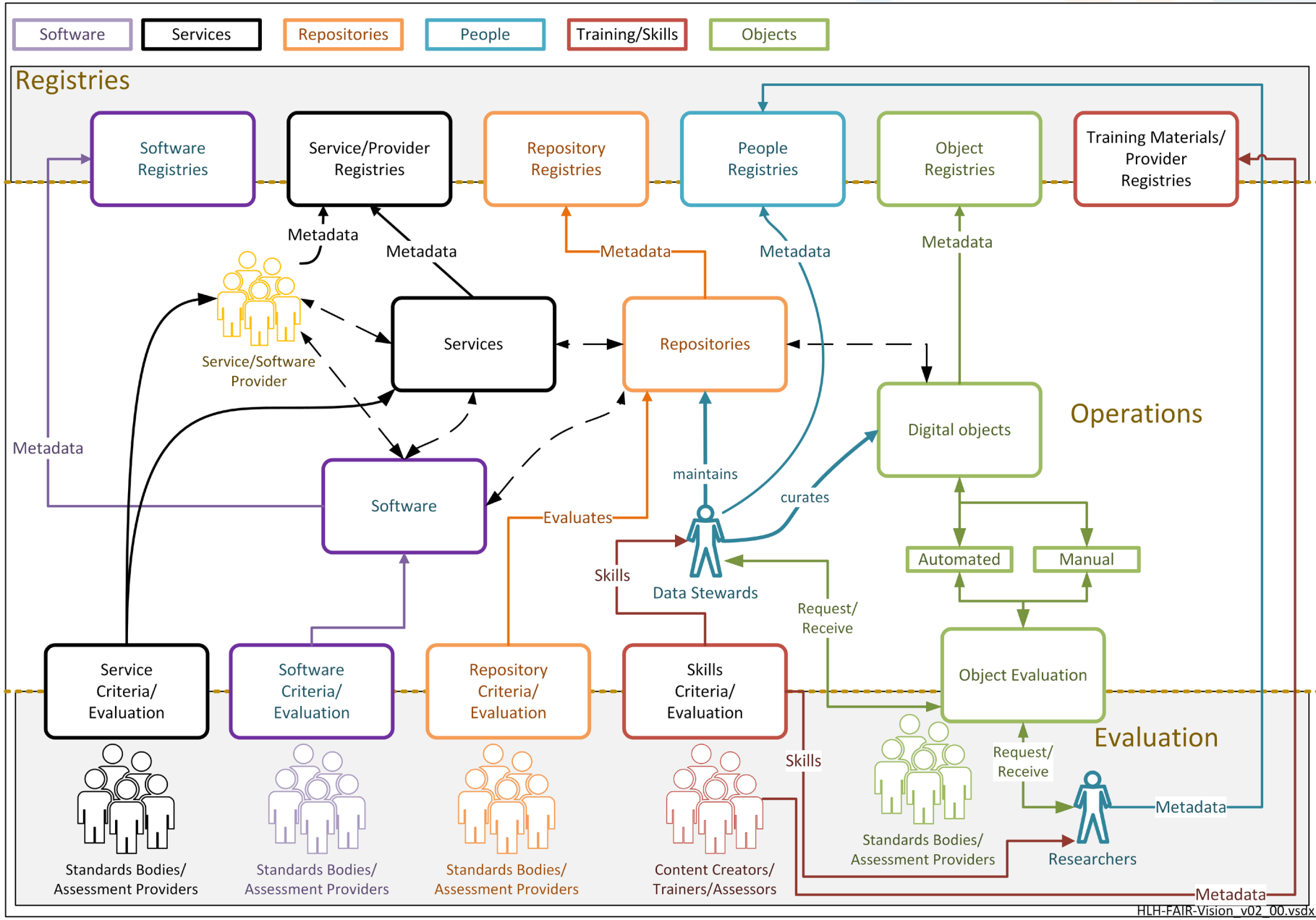
SSHOC "Social Sciences and Humanities Open Cloud", has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, Grant Agreement No.823782

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Potential synergies

- FAIR enabling services and repositories are vital part of the ecosystem
- Trustworthy Digital Repositories (TDR) are key to increasing confidence
- Clarifying responsibilities to avoid overlap and to build on synergies
- Evaluation and certification mechanisms need to be developed further

Source: Hervé L'Hours, & Ilona von Stein. (2020). FAIR Ecosystem Components: Vision (02.00). Zenodo. https://doi.org/10.5281/zenodo.3734273



This project is funded from the EU Horizon 2020 Research

Thank you for your attention!

Join our community

https://www.sshopencloud.eu

@SSHOpenCloud

info@sshopencloud.eu

/in/sshopencloud

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

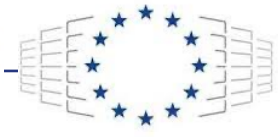
Challenges

- EUDAT's vision:

Data is shared and preserved across borders and disciplines

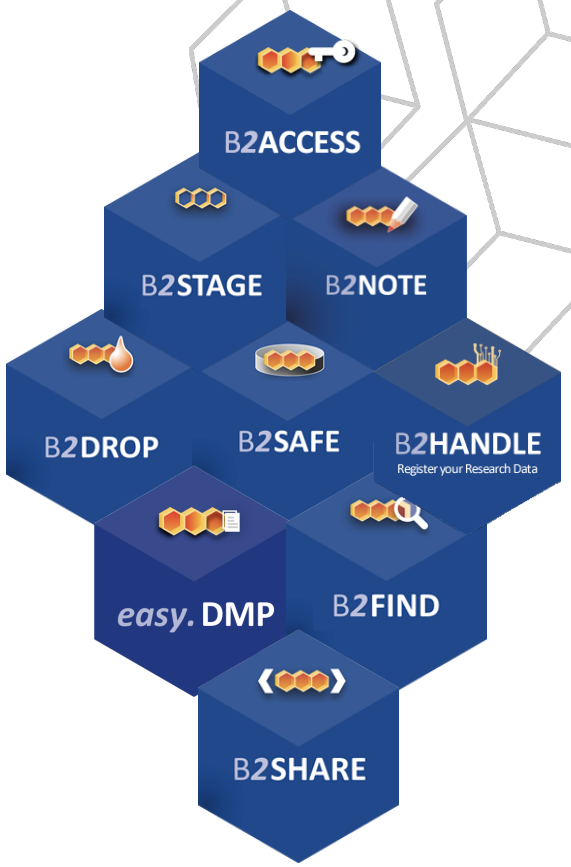
Main Challenge

- Technical interoperability
- Access to resources, data and services
- Policies across domains and disciplines



EUDAT Collaborative Data Infrastructure

Potential Synergies



- B2DROP is EUDAT sync and share service
- ScienceMesh is more as just a sync and share service
- Develop common approaches on interoperability
  - Connecting to computing and analysis
  - Publishing workflows
  - Federated AAI
  - …
- Make services and resources available across infrastructures

- Collaborate on common use cases from communities

EUROPEAN OPEN SCIENCE CLOUD

ENVRI-FAIR & CS3MESH4EOSC

Potential synergies

Zhiming Zhao[1,2], Andreas Petzold[3]

[1]University of Amsterdam,

[2]LifeWatch ERIC Virtual Lab and Innovation Center (VLIC)

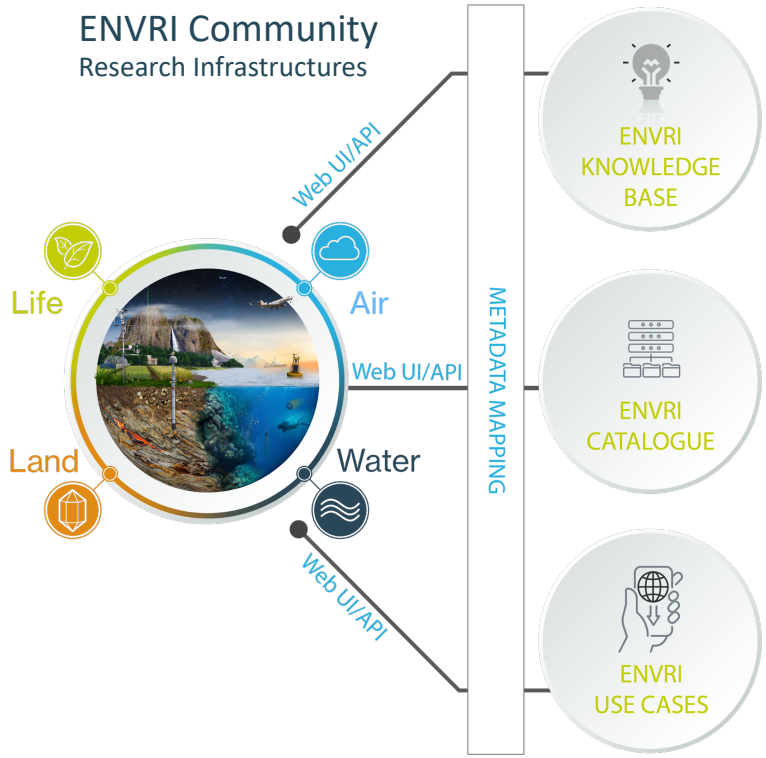
[3]IAGOS, Forschungszentrum Juelich, Germany

ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

ENVRI-FAIR challenges and mission

- Develops **FAIR-based resources** for easy and seamless access to ENVRI data and services
- Implements **common standards and policies** for data life cycle, cataloguing, curation, provenance and service provision
- Realizes **the service platform ENVRI-Hub** for
  - discovery of services and data
  - documented standardized interface and machine actionability
  - re-usability and user support via Jupyter notebooks

ENVRI-Hub and potential synergies



Web UI based access to the ENVRI ecosystem
- quick discovery of data, services and assets
- sharing of engineering practices, technologies and knowledge

Machine actionable interface to the ENVRI ecosystem
- cataloguing all RIs in the ENV domain
- accessing RIs datasets via metadata search
- interface to EOSC and other users (e.g. Copernicus)

Jupyter notebooks on scientific use cases
- Use cases are made accessible as executable demonstrators and demonstrator VREs

Possible synergies :

1. Services for **data access and sharing** in distributed environments
2. On demand **interoperable infrastructure** (e.g., from EOSC or public cloud) **services automation**, e.g., for DevOps, testing data services, and running distributed computing tasks
3. **Virtual Research Environments** (including Jupyter Lab, Hub environments) for data intensive workflow
4. …