



Second EOSCpilot Stakeholders

Forum

Vienna

21 – 22 November 2018

**Enabling interoperability**

**-- experience from the EOSCpilot work**

Doina Cristina Duma (INFN - CNAF)

*On behalf of WP6*

**EOSC** pilot

The European Open Science  
Cloud for Research Pilot Project

[www.eoscpilot.eu](http://www.eoscpilot.eu)




# EOSCpilot: High Level Aims & Impact


The **EOSCpilot** project supports the first phase of development of EOSC:


-  **Engage with a broad range of stakeholders**, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research
-  **Develop a number of demonstrators** functioning as high-profile pilots that **integrate services and infrastructures to show interoperability and its benefits** in a number of scientific domains
-  **Establish the governance framework** for the EOSC and contribute to the development of European open science policy and best practice

## Impact

-  **Reduce fragmentation** between data infrastructures by working across scientific and economic domains, countries and governance models; and
-  **Improve interoperability** between **data infrastructures** by demonstrating how data and resources can be shared even when they are large and complex and in varied formats


Three types of challenges:

 **Scientific Challenges:** deploying the EOSC to deliver Open Science

 Needs and provision


 *What do research communities need from an “Open Data Science Environment”?*

 **Technical Challenges:** developing technical solutions that meet the scientific needs **WP 5-6**


 Services and integration

 *How can EOSC deliver integrated services that are relevant to community needs?*

 **Cultural Challenges:** adopting new, more open ways of working

 Skills and engagement

 *What changes are needed in capability and practices?*

 Governance and policy

 *How should provision be overseen to maximize benefit?*





Interoperability:  
a crucial aspect of the EOSC





**EOSC** pilot  
The European Open Science  
Cloud for Research Pilot Project  
[www.eoscpilot.eu](http://www.eoscpilot.eu)



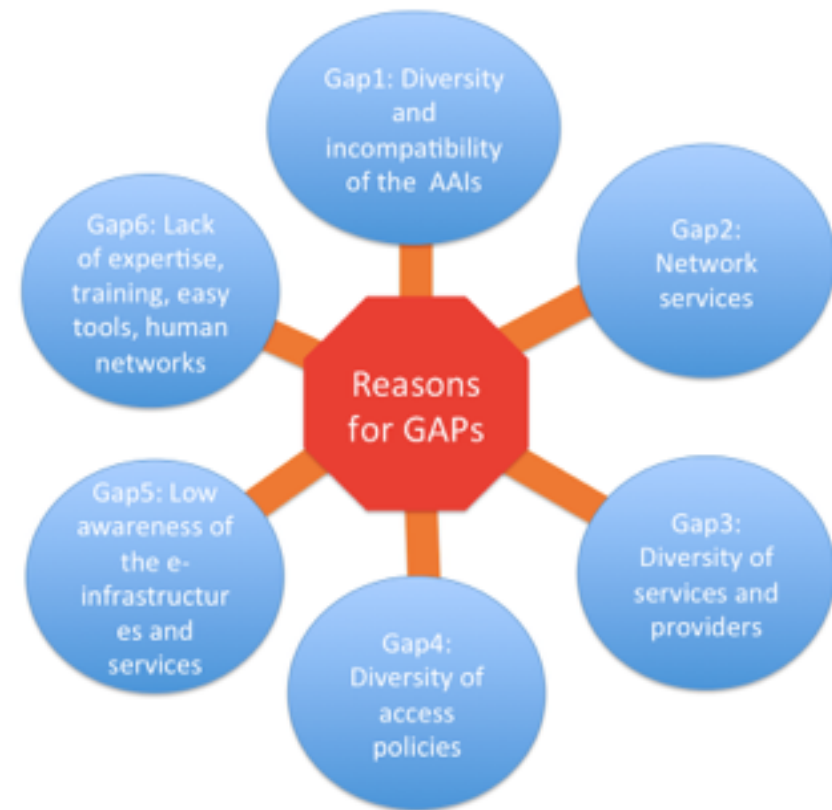
## Infrastructure Interoperability:





-  The **provider view centric**
-  The complementary usage of Cloud, Grid, HTC and HPC infrastructures, including large data-stores, through high speed networks and performant data transfer protocols and tools.
-  The high level objective is to **facilitate the most adequate infrastructures** for the treatment of extensive amounts of data
-  The federated infrastructure pilots have to be set up ... will enable **the analyses of the existing interoperation mechanisms** for software components, services, workflows, users and resource access within existing RI systems.

## Research and Data Interoperability:

-  Provides the **research infrastructure and domain expert view** in the work programme with focus on data interoperability.
-  The definition of a Data Interoperability framework in EOSC is based on the FAIR principles - data and services need to be Findable, Accessible, Interoperable and Reusable.

***Through - instantiation of multi-infrastructure, multi-community pilots.  
Services & SDs – deployed and validated => maturity, scalability, and usability  
for a future EOSC.***



-  **Perform a gap analysis**
-  **Establish synergies and collaborations**
-  **Define procedures and testbeds**
-  **Define the architecture for interoperating various types of infrastructures in the EOSC**

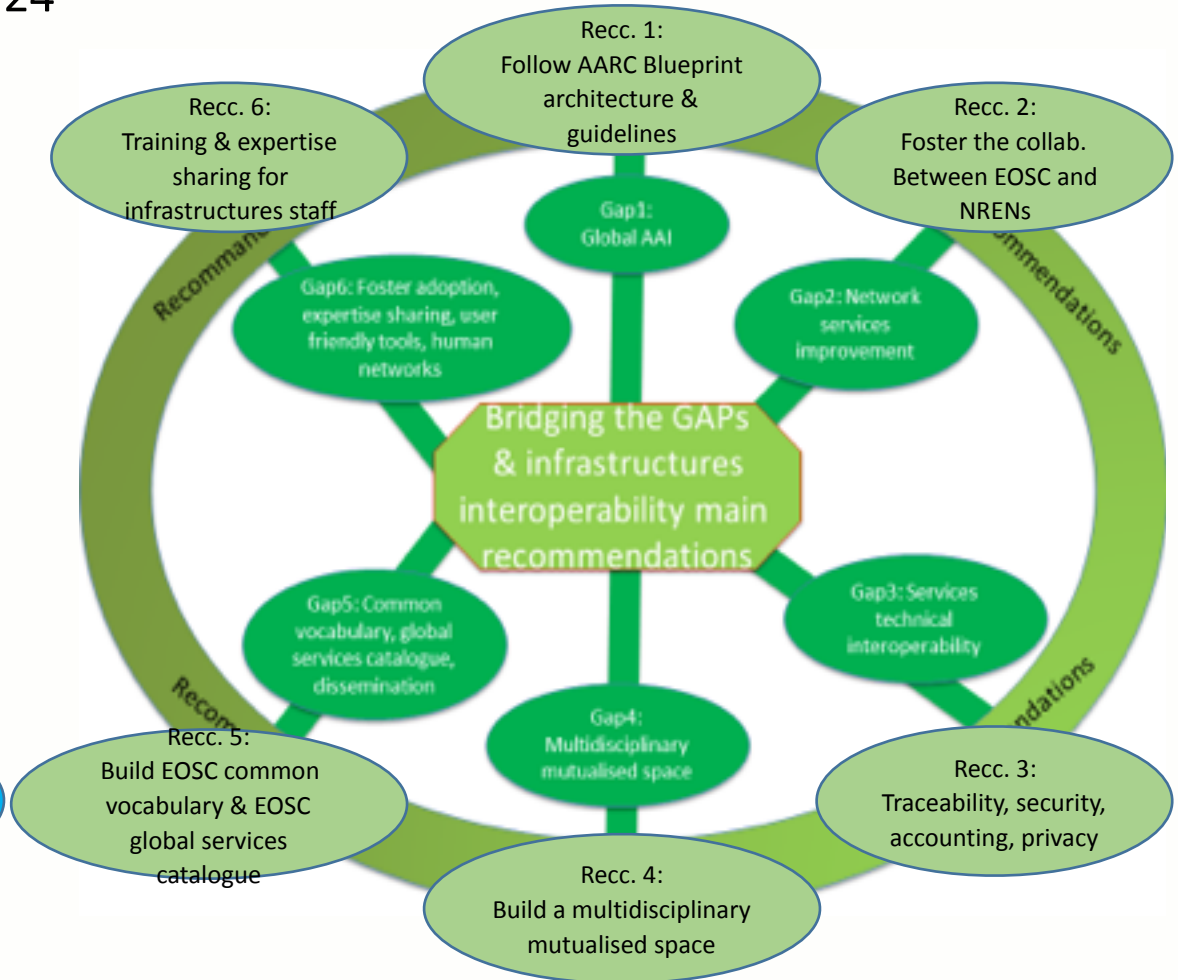


## D6.1: e-infrastructure gap analysis & D6.2: EOSC architecture design and validation procedure

- describes the framework to be set to allow the interoperability between the e-infrastructures and Research infrastructures involved in EOSC






D6.8, due on M24



Risks and recommendations for each gap and bridge identified

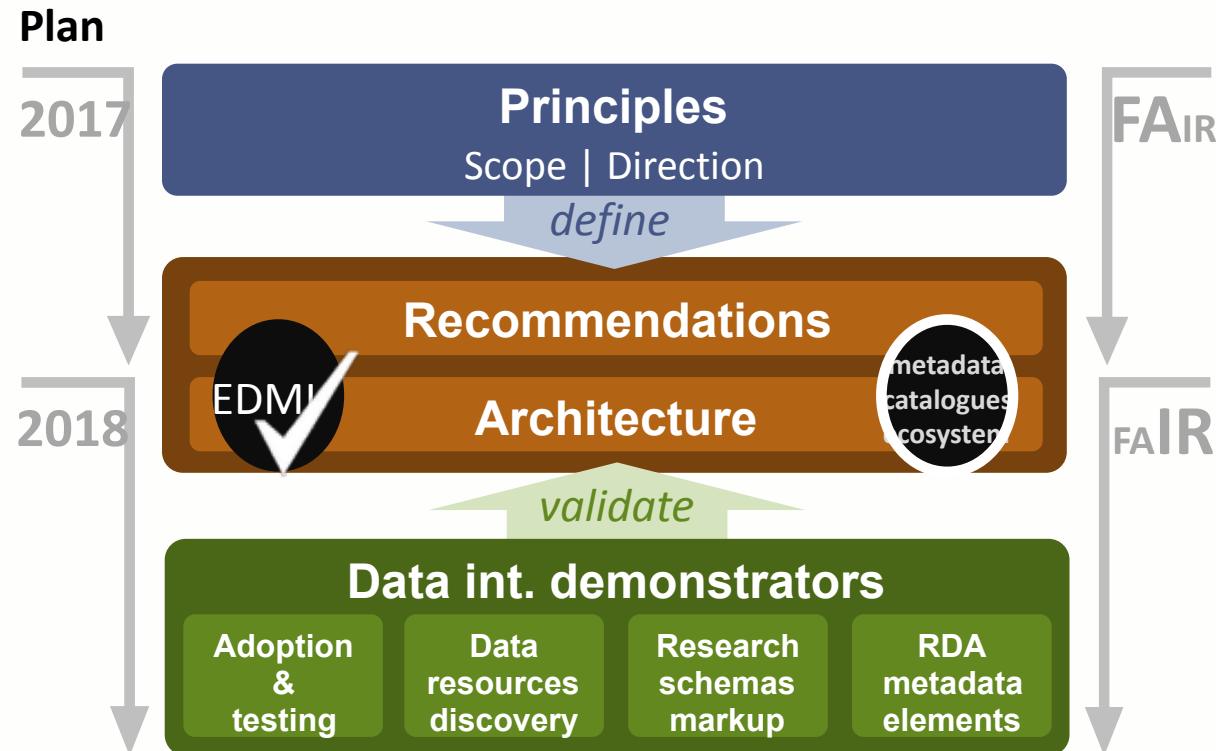


Aims to **establish principles and develop mechanisms** that enable the EOSC to provide research and data interoperability across the diversity of existing (and potential future) research communities, research infrastructures and other research organizations.

-  **analyses** the existing **interoperation mechanisms** for data, software components, workflows, users and resource access with particular attention to the use of standards and their syntactic and semantic representations.
-  **provides the knowledge management framework** - the content descriptions - consumed by the services established in WP5 and technical interoperability defined in task 6.1 and 6.3.
-  **gathers requirements** from the participating RIs and science partners

Organized following the FAIR principles, and address the Findability, Accessibility, Interoperability and Reusability of research assets.

**Goal:** To demonstrate how to ensure availability of scientific data and data-analysis services to users and services through a cloud infrastructure and design a stakeholder driven governance framework












“Develop a number of pilots that integrate services and infrastructures to demonstrate interoperability in a number of scientific domains”







- set up **demonstrators** to show **interoperability** among infrastructures and to **foster the adoption** of the solutions according to the FAIR principles,
- **assess the maturity level** of solutions for what regards **TRL**, openness, scalability, user community adoption and sustainability.
- **Validation** regarding:
  - **AAI requirements** - distributed authorization policies, support for multiple authentication methods using standards
  - **Resource brokering solutions** - high-level resource discoverability and addressability
  - **Accessibility** - EOSC local, Grid, HPC and Cloud resources <- by multiple communities via different interfaces
  - **Data accessibility** – through personal resources, scientific portals, CLI
  - **Interoperability** - underlying distributed storage systems with the EOSC platform services
  - **Transparent networking solutions** - connecting multiple infrastructures and communities, spanning European countries
  - **Data interoperability** - data placement, findability and accessibility in pilot solutions reusing common components




-  **Support** the setup of the Science Demonstrator pilots, following their interoperability requirements and matching them again with available services and solutions
-  **Setup** of different pilot addressing different interoperability aspects:
  -  Transparent Networking – PiCo2 (Pilot for COnnexion between COmputing centers)
  -  Grid & Cloud interoperability – pilot demonstrator for one of the HEP experiments
  -  AAI – through the setup of a scoped interoperability pilot as part of the WLCG Authorization WG & AARC & EOSCpilot collaboration
  -  Resource Brokering & orchestration – leveraging INDIGO-DataCloud solutions
  -  Data accessibility & interoperability of underlying storage systems – distributed Onedata deployment


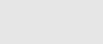


## **PiCO2 (Pilot for COnnecting COmputing centers)**

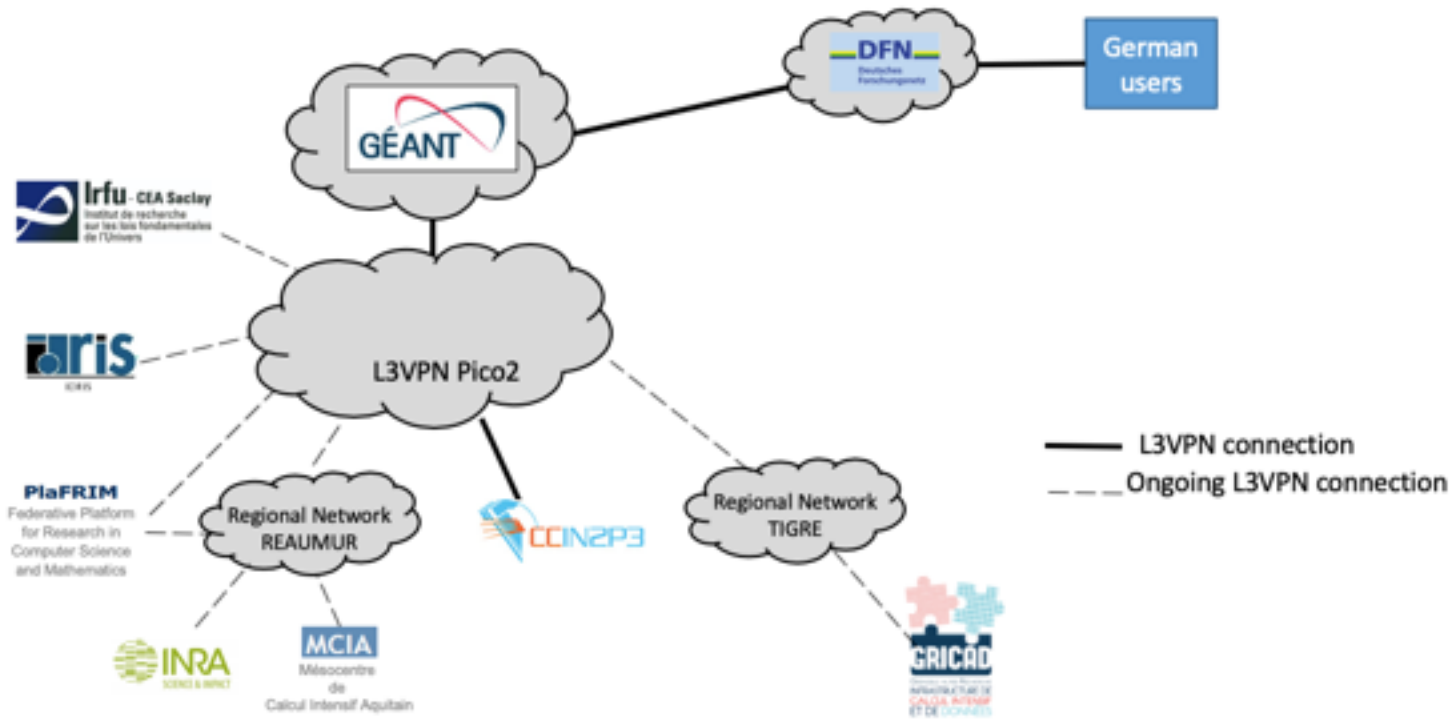
-  one of the first interoperability pilots between generic, community agnostic, infrastructures, especially Tier-1 (National HPC/HTC centres), and Tier-2 (HPC/HTC regional centres).
-  Its main objective is the automation of frequent, community agnostic, data flow (many large files) and code exchange between HPC (National, Regional) and HTC (national, grid) infrastructures
-  **technical groups** have been set up :
  -  a) one for building a network of peer to peer federations between iRODS zones (data storage service), between Tier1 & Tier 2, between Tier2, and between Tier 2 and the grid
  -  b) one for connecting the infrastructures within a L3VPN network and monitoring the performance of the network between sites
  -  c) one for facilitating the mobility and use of codes between different machines, using containers, packages for configuration management, and notebooks



 Current participants:




-  GRICAD, Tier 2 HPC/HTC/Grid/cloud, Grenoble
-  MCIA, Tier 2 HPC/Grille/Cloud, Bordeaux
-  Plafrim, cluster INRIA Bordeaux, Tier 2 HPC
-  IDRIS, CNRS National Computing Center, Tier 1 HPC
-  CC-IN2P3, IN2P3/CNRS Computing Center, Tier 1 HTC/Grid/Cloud
-  RENATER, REN, France
-  France-Grille, EGI NGI, France ; Tier 1 Grid/Cloud
-  GRIF / CEA, Tier 2, Saclay
-  DESY, Tier-2, Hamburg

**Pico2 Layer 3 Virtual Private Network**  
... towards a European wide interoperability









## **Dynamic On Demand Analysis Service**

-  a PaaS tool built combining several solutions and products developed by the **INDIGO-DataCloud**
-  extensively tested on a dedicated interoperability testbed under the umbrella of the **EOSCpilot** project
-  originally designed for the CMS Experiment at LHC => **Thematic Service that will provide multi-disciplinary solutions** in the **EOSC-hub** project



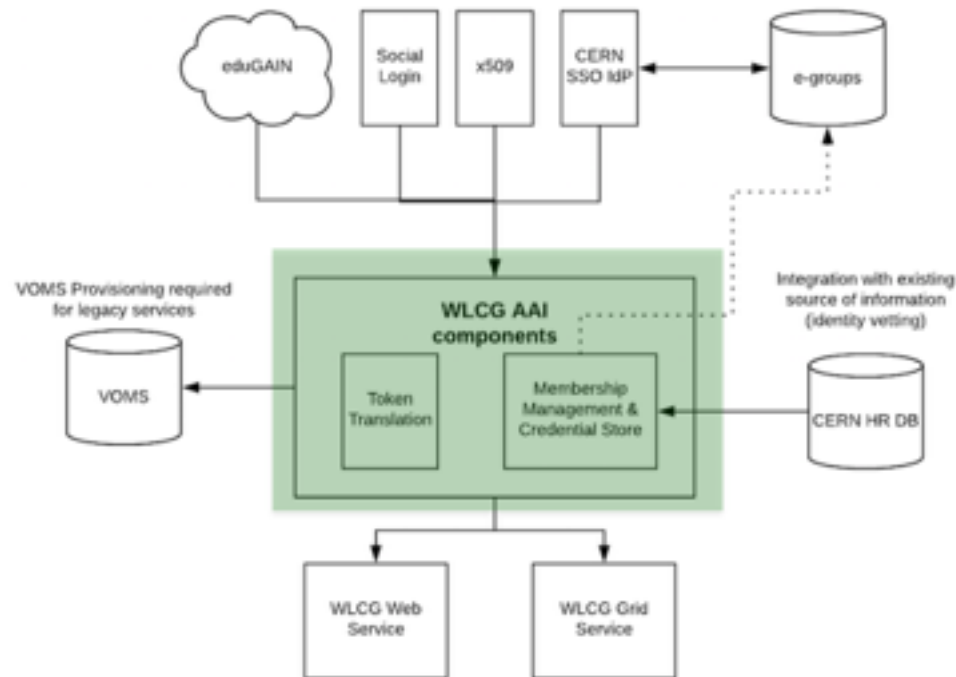
## Weaknesses identified and recommendations:

-  Federated AAI -> **crucial to have it as a EOSC provided service.**
-  **Transparent Data Access:** - scalable solution (XrootD) - A **more generic solution** would be a big plus.
-  **Resource monitoring – common strategy** for cloud resources monitoring
-  **PaaS Orchestration** - improvement both in the interface and in the management of IaaS ranking.

Collaboration on the field of **authorization and authentication, policies and recommendations** regarding solutions design -> <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGAuthorizationWG>

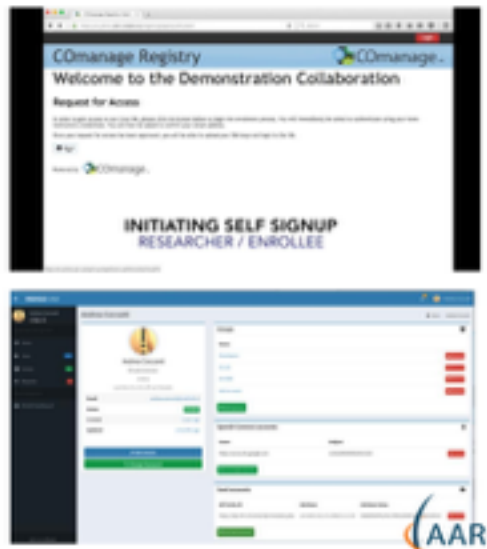
Main objectives:

- Design and testing of a **WLCG Membership Management and Token Translation service**, facilitated by pilot projects with the support of AARC
- Definition of a **token-based authentication and authorization profile for WLCG**



### AAI Pilot Projects

- Two solutions appear to meet the majority of requirements
  - EGI Check-in & COmanage
  - **INDIGO IAM**
- Additional integration required for
  - VOMS provisioning & lookup
  - CERN HR DB integration
  - AUP re-signing

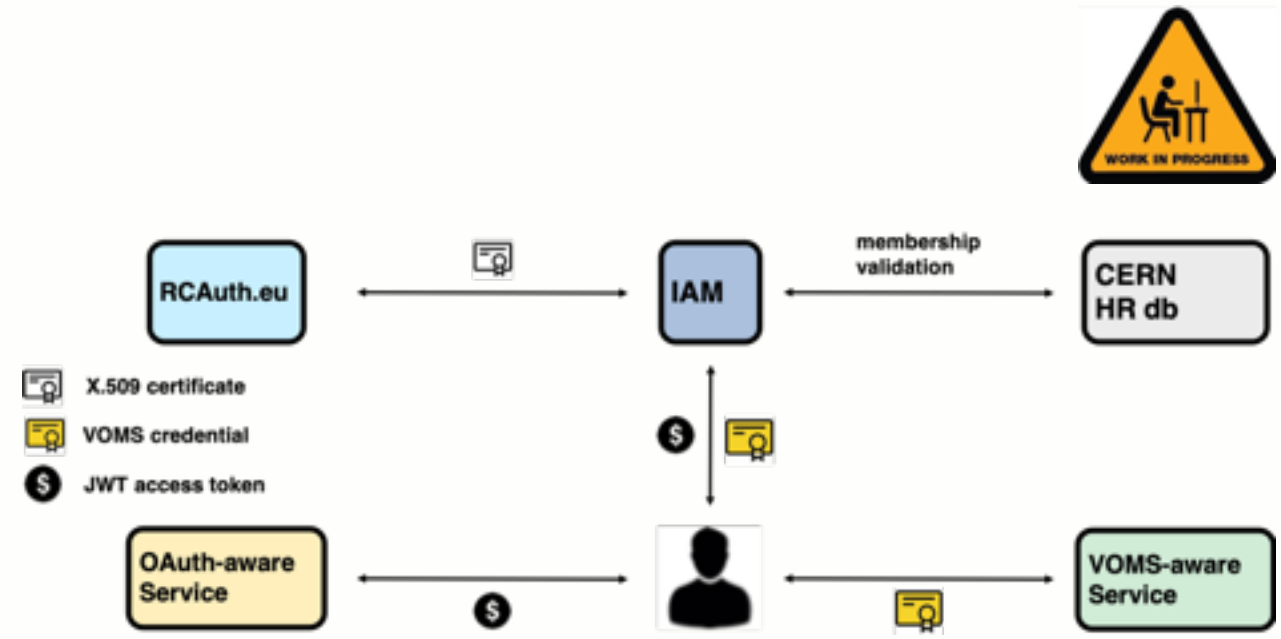
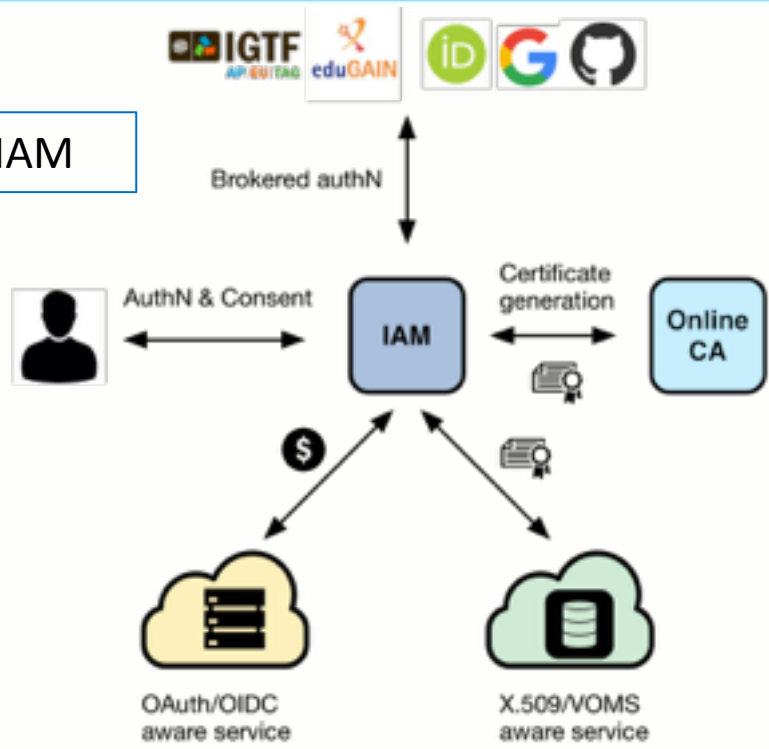


WLCG  
WLCG AuthZ WG

[andrea.ceccanti@cnaf.infn.it](mailto:andrea.ceccanti@cnaf.infn.it)



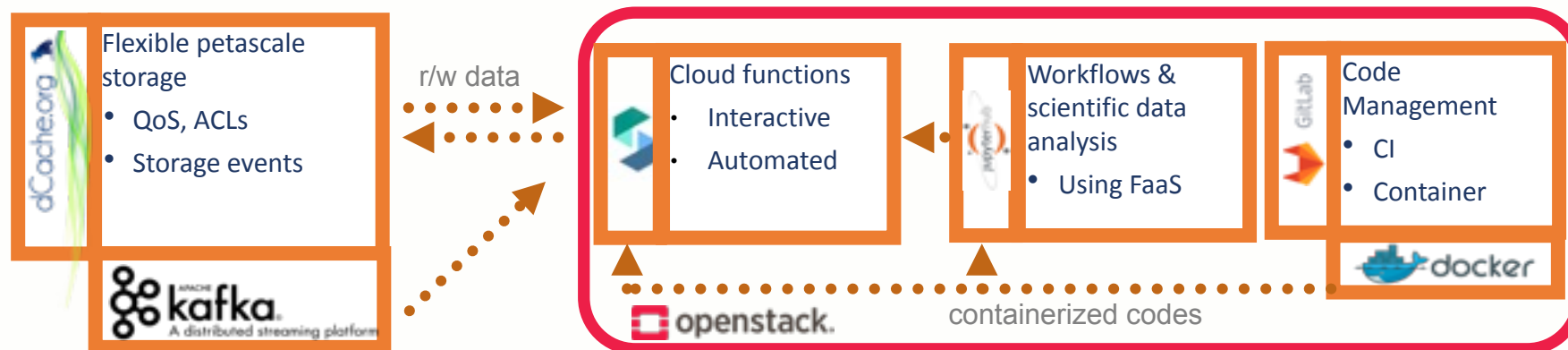
INDIGO IAM



- IAM instance deployed @ INFN-CNAF since January 2018 to showcase main features and integration capabilities
  - <https://wlcg-authz-wg.cloud.cnaf.infn.it/login>
- This deployment is being migrated to CERN infrastructure for further validation & feedback on
  - [RCAuth.eu](http://RCAuth.eu) and CERN HR database integration
  - Registration & administration management functionality



- Interoperability between cloud storage and virtualized compute clusters in the EOSC
  - dCache storage events (Kafka and Server Sent Events (SSE))
  - RESTful API based interaction
  - Single Sign On (SSO) e.g. OpenID Connect (OIDC)
  - Secure delegation of read/write access rights to stateless cloud functions (using Macarons)
- Cloud functions as a service (FaaS) using docker runtimes
  - Preserving environments, versions, configurations for function catalogues and user codes
  - Enhanced portability on federated cloud environments in the EOSC
  - Same architecture for automated pipelines also used for interactive workflows (Jupyter Notebooks)
  - CI/CD for docker containers (scaling load with GitLab Runners)
  - Seamless scaling for headless applications; graphical interaction needs co-development to facilitate integration of scientific applications on cloud based infrastructures and platforms



## Objectives

### Infrastructure interoperability:

facilitate the most adequate infrastructures for the treatment of extensive amounts of data. Demonstrate with multi-infrastructure, multi-community pilots (science demonstrators from WP3 & WP4)

### Research and Data Interoperability:

Data & services to be findable, accessible, interoperable and reusable (FAIR)

### Testbeds for interoperability:

Put to work Science Demonstrator and learn about interoperability issues and solutions

### Key Output:

*The design of a future EOSC based on **federated interoperable services** meeting the **needs** of the thematic research domains and wider user base.*

### WP Outputs so far:

- Gap analysis
- EOSC architecture
- Reports on data interoperability and testbeds
- Initial and Updated Requirements for testbeds

### More to come:

- **Final** EOSC architecture
- **Final** report on data interoperability
- **Final** Interoperability Testbeds report



# Interoperability: Key element of the EOSC

WP lead: Volker Beckmann (CNRS / IN2P3),  
T6.1 lead: Geneviève Romier, Eric Fede (CNRS / IN2P3)  
T6.2 lead: Rafael C. Jimenez Domenech (ELIXIR)  
T6.3 lead: Doina Cristina Duma (INFN)

