

# EOSCpilot: Interoperability Interim results

C. Duma<sup>1</sup>, A. Costantini<sup>1</sup>, D. Michelotto<sup>1</sup>, A. Ceccanti<sup>1</sup>, E. Fattibene<sup>1</sup> and D. Salomoni<sup>1</sup>

<sup>1</sup>INFN Division CNAF, Bologna, Italy

E-mail: ds@cnafe.infn.it

## Abstract.

The EOSCpilot project is the first project in the entire EOSC programme, tasked with exploring some of the scientific, technical and cultural challenges that need to be addressed in the deployment of the EOSC. The EOSCpilot project has been funded to support the first phase in the development of the European Open Science Cloud (EOSC). In this paper we present a summary of the second year activities results in the field of interoperability containing the first results of the validation of services and demonstrators in the interoperability testbeds and the revised interoperability requirements derived from these activities.

## 1. Introduction

The European Open Science Cloud (EOSC) programme aims to deliver an Open Data Science Environment that federates existing scientific data infrastructures to offer European science and technology researchers and practitioners seamless access to services for storage, management, analysis and re-use of research data presently restricted by geographic borders and scientific disciplines.

In the framework of the EOSCpilot, WP6, "Interoperability", aims to develop and demonstrate the interoperability requirements between e-Infrastructures, domain research infrastructures and other service providers needed in the European Open Science Cloud. It provides solutions, based on analysis of existing and planned assets and techniques, to the challenge of interoperability. Two aspects of interoperability are taken into consideration: **Data interoperability**, ensuring that data can be discovered accessed and used according to the FAIR principles, and **Service interoperability**, ensuring that services operated within different infrastructures can interchange and interwork.

In the framework of the EOSCpilot project INFN, and in particular CNAF is the coordinator of the activities of the task T6.3 - Interoperability pilots (service implementation, integration, validation, provisioning for Science Demonstrators).

One of the project's main Objectives related to WP6 is to:

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

mapped into some specific Objectives addressed by the T6.3 task:

- Validating the compliance of services provided by WP5, "Services", with specifications and requirements defined by the Science Demonstrators in WP4, "Science Demonstrators"

- Defining and setting up distributed Interoperability Pilots, involving multiple infrastructures, providers and scientific communities, with the purpose of validating the WP5 service portfolio.

## 2. Activities and Achievements

During 2018 the main activities coordinated by INFN-CNAF were:

- 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 COnection between COmputing centers)
  - Grid and 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 and EOSCpilot collaboration
  - Resource Brokering & orchestration leveraging INDIGO-DataCloud solutions
  - Data accessibility & interoperability of underlying storage systems distributed Onedata deployment
- Continuous interaction and communication with Science Demonstrators shepherds in order to collect eventual new requirements result of the activities done in the implementation of the SDs specific use cases.

### 2.1. Interoperability pilots: Transparent Networking

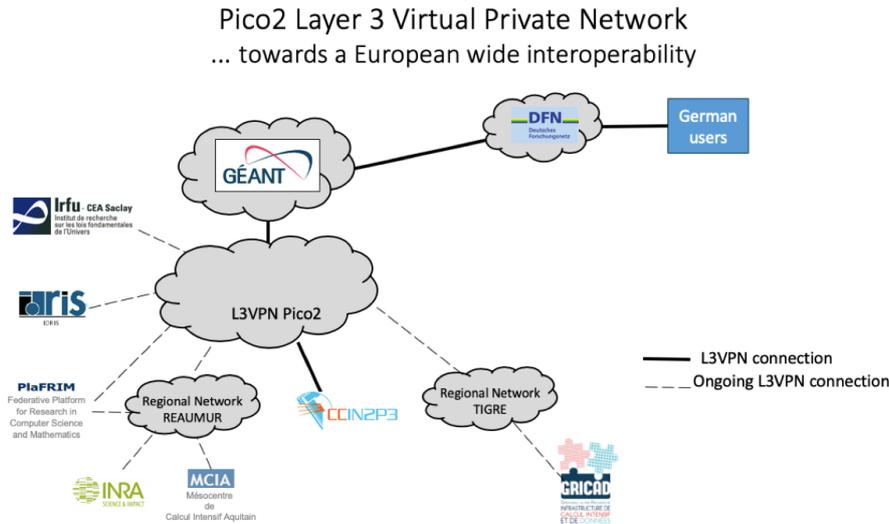
The **PiCO2 (Pilot for COnecting COmputing centers)** is one of the first interoperability pilots between generic, community agnostic, infrastructures, especially Tier-1 (National HPC/HTC centres), and Tier-2 (HPC/HTC regional centres). Its main objective is the automation of frequent, community agnostic, data flow (many large files) and code exchange between HPC (National, Regional) and HTC (national, grid) infrastructures During 2018 technical groups have been set up:

- 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
- one for connecting the infrastructures within a L3VPN network and monitoring the performance of the network between sites
- one for facilitating the mobility and use of codes between different machines, using containers, packages for configuration management, and notebooks

In Figure 1 we see the current status of the project with the sites involved.

### 2.2. Interoperability pilots: Grid-Cloud interoperability demonstrator for HEP community

Dynamic On Demand Analysis Service (DODAS) is a Platform as a Service tool built combining several solutions and products developed by the INDIGO-DataCloud H2020 project. It has been extensively tested on a dedicated interoperability testbed under the umbrella of the EOSCpilot project, during the first year of the project. Although originally designed for the Compact Muon Solenoid (CMS) Experiment at LHC, DODAS has been quickly adopted by the Alpha Magnetic Spectrometer (AMS) astroparticle physics experiment mounted on the ISS as a solution to exploit opportunistic computing, nowadays an extremely important topic for research domains where computing needs constantly increase. Given its flexibility and efficiency, DODAS was selected as one of the Thematic Services that will provide multi-disciplinary solutions in the EOSC-hub project. An integration and management system of the European Open Science Cloud starting in January 2018. During the integration pilot the usage of any cloud (both public and private)



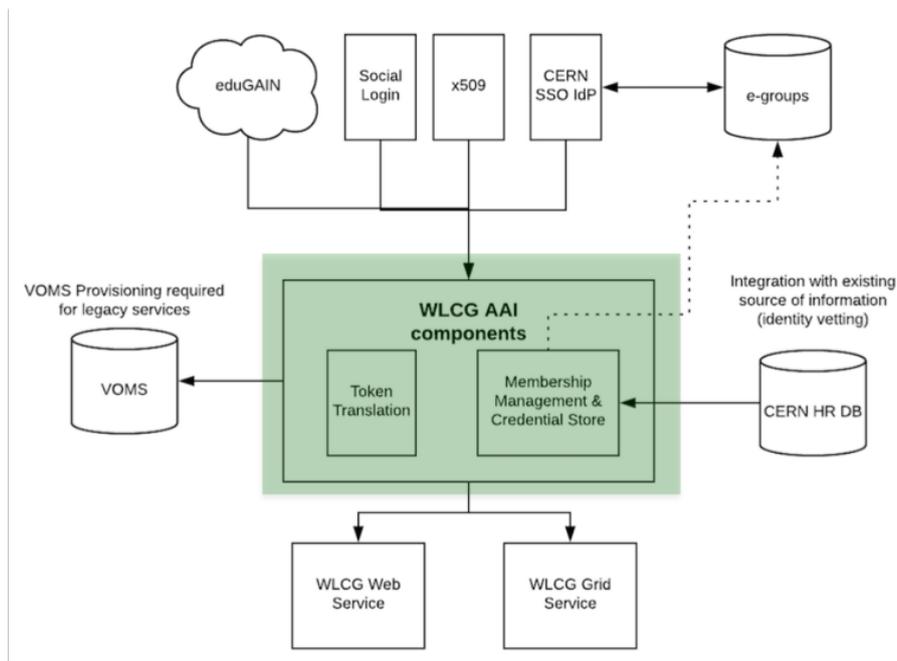
**Figure 1.** PiCO2 Layer 3 VPN

to seamlessly integrate existing Grid computing model of CMS was demonstrated. Overall, integration has been successful and much experience has been gained resulting in improved understanding of weaknesses and aspects to improve and to optimise. Weaknesses, and aspects to be improved include:

- Federation: federated access to underlying IaaS is a key. So far we have experienced several issues. Frequently we had issues with the IaaS provider already using OpenID Connect Authorisation Server and thus unable to federate additional services. We adopted ESACO solution to solve such a problem. It would be crucial to have it as a EOSC provided service.
  - for non-proprietary IaaS would be extremely important in the EOSC landscape. A scenario where, as example, a commercial cloud is used, would benefit of such functionality for counting the overall HEPSpec .
- Transparent Data Access: so far the only scalable solution we can use is XrootD . However, this might not fit all possible use cases. A more generic solution would be a big plus.
- Resource monitoring: we didn't find a common solution for monitoring cloud resources. Although we implemented our own we are convinced that a common strategy would be extremely valuable.
- PaaS Orchestration: Although the current INDIGO PaaS Orchestrator has been fully integrated and show enormous advantages while dealing with multiple IaaS, there is room for improvement both in the interface and in the management of IaaS ranking.

### 2.3. Interoperability pilots: AAI

The EOSCpilot and AARC (add reference) projects started a collaboration activity in the field of authorisation and authentication, policies and recommendations regarding their design, that took shape, in the scope of the WP6 activities, under the form of an AAI interoperability demonstrator setup as part of the AARC pilots Task 1: **Pilots with research communities based on use cases provided - the WLCG use case**, regarding the *Implementation of IdP/SP Proxy, mainly to provide Token Translation Services to allow end users to login without the need of manually managing X.509 certificates*. A team of people was formed, under the WLCG coordination, to deal with the various activities, the **WLCG Authorisation WorkingGroup (WG)**, motivated by:



**Figure 2.** WLCG AAI system

- Evolving Identity Landscape
  - User-owned x509 certificates - Federated Identities
  - Federated Identities linkage with existing VOMS authorisations not supported
  - Maintaining assurance and identity vetting for federated users not supported
- Central User Blocking
  - Retirement of glxexec removes blocking capability (& traceability)
  - VO-level blocking not a realistic sanction
- Data Protection
  - Tightening of data protection (GDPR) requires fine-grained user level access control

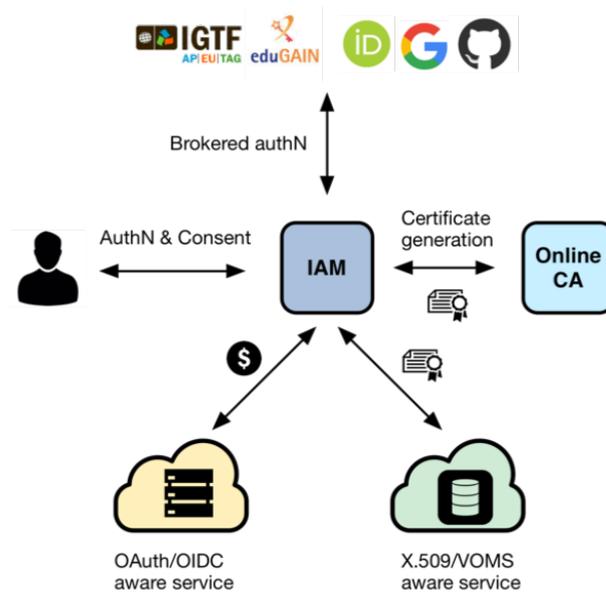
The federated identities and the adoption of new authorisation standards by industry is a strong signal for WLCG to adapt its authorisation infrastructure, of which we can see the schema in Figure 2.

After an initial requirements gathering , and analysis of how existing solutions functionalities match the requirements , two main activities started:

- (i) Design and testing of a WLCG Membership Management and Token Translation service, facilitated by pilot projects with the support of AARC (AAI Pilot Projects)
- (ii) Definition of a token based authorization schema for downstream WLCG services and token issuers (JWT)

The activities done during 2018 regarded the:

- IAM instance deployed @ INFN-CNAF since January 2018 to showcase main features and integration capabilities of the INDIGO-IAM service (figure 3)
  - <https://wlcg-authz-wg.cloud.cnaf.infn.it/login>
- The migration of this deployment to CERN infrastructure for further validation & feedback on



**Figure 3.** INDIGO-IAM service architecture

- RAuth.eu and CERN HR database integration
- Registration & administration management functionality

#### 2.4. Data Interoperability Demonstrators

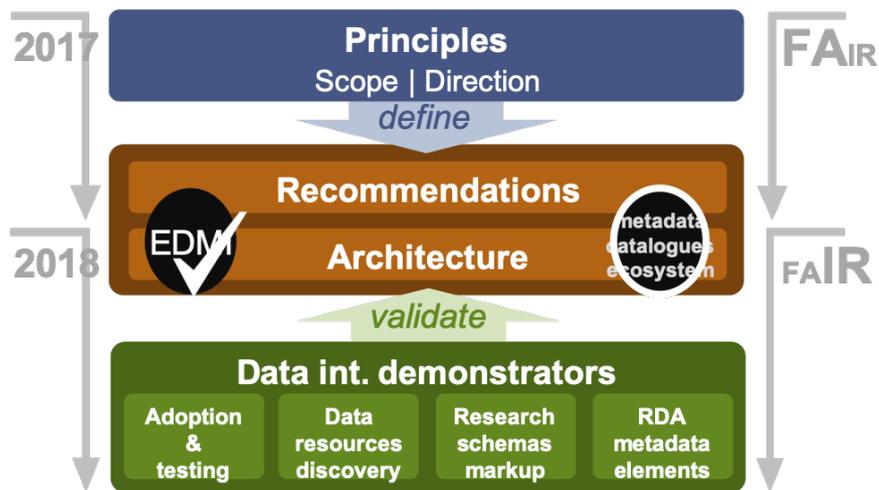
One of the two tracks into which interoperability is mapped in the EOSCpilot WP6 is the Research and Data Interoperability track, task T6.2 that provides the research infrastructure and domain expert view in the work programme with focus on data interoperability.

T6.2 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 organisations. It

- 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
- is organised following the FAIR principles, and address the Findability, Accessibility, Interoperability and Reusability of research assets.

After the delivery of the first draft of the strategy and recommendations done in 2017 and beginning of 2018, four data interoperability demonstrators have been proposed to test components of the strategy:

- Evaluation of the EDM1 (EOSC Datasets Minimum Information) metadata guidelines to find and access datasets
- Discovery of compliant data resources and metadata catalogues
- Research schemas for exposing dataset metadata
- Description and guidelines per metadata property



**Figure 4.** Data Interoperability activities plan

Our role during 2018 was to:

- Facilitate & Support through
  - Integration of the feedback from demonstrators into the EOSCpilot data interoperability strategy
  - Organisation of phone calls, F2F meetings and other events to help delivering the proposed tasks
  - Track the outcomes produced by the data interoperability demonstrators
  - Promote the activities and results of the demonstrators and work on ways to recognise the contribution of the demonstrator participants.

### 3. Other activities

During 2018 we contributed also to the writing of two of the project deliverables summarizing the activities done:

- **D6.5: Interim Interoperability Testbed report** - highlighting the status of all Science Demonstrators testbeds and activities, most of them in line with what planned initially, some of them requiring extensions in order to finalise the work.
- **D6.7: Revised Requirements of the Interoperability Testbeds** - providing an updated picture of the different actors involved in the EOSCpilot project, that, through their activities, aim in shaping the EOSC environment, improving the services and e-infrastructures it consists of, and also provide requirements and recommendations, based on the experiences they gained during the project, to help in the prioritisation of the new features of the existing services and of the development of new services that are aligned with the needs and expectations of researchers.

### 4. Conclusions

After the collection of the initial requirements on the interoperability testbeds, reported in deliverable D6.4, at the end of 2017, the activities continued on supporting the projects Science Demonstrators and an interim report on the status of the testbeds was provided in the D6.5 deliverable. In the second part of the project after the first round of selected Science Demonstrators were almost at the end of their activities, while the second round was at the beginning, an updated list of requirements regarding interoperability aspects was provided in the

deliverable D6.7. The activities of the Interoperability Pilots task will be concluded during the first part of 2019 by providing the validation of the e-infrastructures and services deployed. For this final assessment we will take into considerations the tools/services developed as part of other EC projects to implement the interoperability aspects, e.g. the Interoperability (IOP) Quick Assessment Toolkit (add reference), developed in the context of Action 2.1 of the Interoperability Solutions for European Public Administrations (ISA) Programme (add reference)

### **Acknowledgments**

EOSCpilot has been funded by the European Commission H2020 research and innovation program under grant agreement RIA 739563.