

**EGI-Engage**

Report on evaluation of EPOS EGI pilot studies

D6.23

|  |  |
| --- | --- |
| **Date** | 16 August 2017 |
| **Activity** | WP6 |
| **Lead Partner** | INFN/INGV |
| **Document Status** | FINAL |
| **Document Link** | <https://documents.egi.eu/document/3040> |

Abstract

The EPOS Competence Centre (CC) of the EGI-Engage project drives collaboration between EGI and the European Plate Observing System (EPOS) service developers and providers in order to collect, analyse and compare Earth Science community needs with EGI technical offerings. The work of the CC in the project was driven by three specific use cases that were selected by the CC members in the first project year (2015). This document describes the three pilots that have been developed during the second and third project years (18 months) based on these use cases: (1) a prototype service for the Authorisation-Authentication of EPOS users; (2) a cloud-based computational portal to conduct MISFIT earthquake simulations; and (3) a data access mechanism to enable the computation of Sentinel-1 data on a scalable platform. This document draws conclusions and lessons learned from the pilots and outlines possible future work within the EPOS-EGI collaboration.

**COPYRIGHT NOTICE**



This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). The EGI-Engage project is co-funded by the European Union Horizon 2020 programme under grant number 654142.

**DELIVERY SLIP**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Name*** | ***Partner/Activity*** | ***Date*** |
| **From:** | Daniele Bailo | INGV / SA2 | 18/Aug/2017 |
| **Moderated by:** | Gergely Sipos | EGI Foundation/ SA2 |  |
| **Reviewed by** | Luca Trani  Keith Jeffery | KNMI  NERC/BGS | 28/Aug/2017 |
| **Approved by:** | AMB and PMB |  | 13/Sept/2017 |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
| **v.1** | 18/08/2017 | First full draft for external review with input from   * Mariusz Sterzel (CYFRONET) * Alessandro Spinuso (KNMI) * Andre Gemund (SCAI Farunhofer) * Genevieve Romier (CNRS) * Diego Scardaci (EGI.eu) * Gergely Sipos (EGI.eu) | Daniele Bailo / INGV |
| **FINAL** | 08/09/2017 | FINAL version, improved based on reviewers’ feedback | Mariusz Sterzel/ACK Cyfronet AGH |

**TERMINOLOGY**

A complete project glossary and acronyms are provided at the following pages:

* <https://wiki.egi.eu/wiki/Glossary>
* <https://wiki.egi.eu/wiki/Acronyms>

**Contents**

[1 Introduction 6](#_Toc493079982)

[1.1 Background and motivations 6](#_Toc493079983)

[1.2 EPOS infrastructure 6](#_Toc493079984)

[2 Prototype of the EPOS Authorisation Authentication Infrastructure (AAI) 8](#_Toc493079985)

[2.1 Overview 8](#_Toc493079986)

[2.2 Architecture 9](#_Toc493079987)

[2.2.1 High-Level Service architecture 10](#_Toc493079992)

[2.3 Demonstration 11](#_Toc493079993)

[2.3.1 Scenario 11](#_Toc493079994)

[2.3.2 Feedback 14](#_Toc493079995)

[2.4 Future plans 14](#_Toc493079996)

[3 Computational Earth Science, (Earthquake simulation and MISFIT analysis) 16](#_Toc493079997)

[3.1 Overview 16](#_Toc493079998)

[3.2 Architecture 17](#_Toc493079999)

[3.2.1 High-Level Service architecture 17](#_Toc493080003)

[3.2.2 Integration and dependencies 18](#_Toc493080004)

[3.3 Demonstration 21](#_Toc493080005)

[3.3.1 Scenario and current status 21](#_Toc493080006)

[3.3.2 Feedback 21](#_Toc493080007)

[3.4 Future plans 22](#_Toc493080008)

[4 Satellite Data (P-SBAS InSAR Sentinel-1 TOPS) 23](#_Toc493080009)

[4.1 Overview 23](#_Toc493080010)

[4.2 Architecture 24](#_Toc493080011)

[4.2.1 High-Level Service architecture 24](#_Toc493080015)

[4.3 Demonstration 25](#_Toc493080016)

[4.3.1 Scenario 25](#_Toc493080017)

[4.3.2 Feedback 26](#_Toc493080018)

[4.4 Future plans 26](#_Toc493080019)

[5 Conclusions, lessons learnt, future work 27](#_Toc493080020)

**Executive summary**

The EPOS Competence Centre (CC) of the EGI-Engage project drives collaboration between EGI and the European Plate Observing System (EPOS) service developers and providers in order to collect, analyse and compare Earth Science community needs with EGI technical offerings. The work of the CC in the project was driven by three use cases selected in the first project year (2015). This document describes: a) the three pilots that have been developed during the second and third project years (18 months) based on the selected use cases; and the demonstrations of such pilots that were presented to the EPOS community.

The pilots highlight how EPOS could benefit from EGI services and provide the first insights about how services from the EGI infrastructure could be connected to the EPOS Research Infrastructure as Distributed Integrated Core Services (ICS-D)[[1]](#footnote-1). Roadmaps for the future enhancements of the pilots are also included in the document, together with plans for exploiting and disseminating the work done so far to the wider EPOS community.

The implemented pilot demonstrators are the following:

* AAI: This pilot demonstrated interoperability between the EPOS Authentication Authorisation Infrastructure (AAI) and the EGI CheckIn services, showing that EGI infrastructure services can be capable of acting as ICS-D in the wider EPOS architecture[[2]](#footnote-2). The EPOS AAI prototype was developed based on the UNITY IDM technology interfaced with EGI CheckIn.
* Earthquake simulation (MISFIT): this pilot shows how an existing application, i.e. MISFIT from the field of Computational Seismology, can be improved by integration with the EGI Federated Cloud. Furthermore, it demonstrates how EGI and EUDAT services can jointly serve a research community with compute and storage services. (using compute cloud from EGI, and data preservation services B2SAFE & B2STAGE from EUDAT).
* Satellite Data: the aim of this demonstrator was the setup of an environment with the use of EGI services that can facilitate the development of new services in the satellite data processing domain. Within EPOS the satellite data TCS will deal with the processing of the Earth Observation datasets collected by various satellites, including the Sentinels of the Copernicus programme. The pilot deployed an EPOS service[[3]](#footnote-3) on top of the Geohazard Thematic Exploitation Platform by Terradue from the satellite data TCS, and interconnected it to the EGI Federated Cloud to exploit its computing and storage resources (work done in collaboration with the EGI-Engage task SA1.3[[4]](#footnote-4)).

The demonstrators used the CheckIn security proxy service (1st demonstrator), and two federated cloud sites from EGI: SCAI in Germany (2nd demonstrator) and BEgrid-BELNET in Belgium (3rd demonstrator).

By the end of EGI-Engage one of the pilots (the AAI pilot) reached demonstration stage, and its first version was demonstrated to the EPOS community in 2016. Based on the feedback received during that demo a further developed version is to be demonstrated again later in 2017. The other two pilots still require further work to reach demonstration stage. The Satellite Data pilot expects to reach this during 2017, the Earthquake simulation pilot in 2018. The latter will be based on follow-up ‘scientific demonstrator’ activity funded for VERCE in the EOSCpilot H2020 project.

# Introduction

## Background and motivations

EPOS[[5]](#footnote-5), the European Plate Observing System, is a long-term plan to facilitate integrated use of data, data products, and facilities from distributed research infrastructures for solid Earth science in Europe. EPOS will bring together Earth scientists, national research infrastructures, ICT (Information & Communication Technology) experts, decision makers, and public to develop new concepts and tools for accurate, durable, and sustainable answers to societal questions concerning geo-hazards and those geodynamic phenomena (including geo-resources) relevant to the environment and human welfare.

In this framework, e-Infrastructure providers are supposed to support EPOS platform by providing access to existing computational resources and specific services coordinated and harmonised at European level. The CC was established in EGI-Engage for testing and piloting purposes, i.e. to verify the effective interoperability level and the maturity of the EPOS and EGI infrastructures.

Members of the CC were selected based on their competence and involvement in the investigations of the pilots, that is to say: 1) harmonised provisioning of AAI solutions, for which Cyfronet was reckoned to be one of the main players;2) provisioning of computational resources in the framework of computational seismology, which triggered the participation of the main stakeholders in the related VERCE project (i.e. KNMI, IPGP - CNRS, SCAI Fraunhofer, GRNET); and 3) other EPOS participants already involved in joint EPOS / EGI developments on satellite data (i.e.CNR).

## EPOS infrastructure

The EPOS functional architecture is composed of three connected technical and organizational elements: NRIs, TCS, ICS with increasingly harmonised IT solutions:

* The National Research Infrastructures (NRIs) represent the underpinning EPOS data providers that will guarantee access to quality-checked data and products. The EPOS architecture ensures that new RIs can be integrated as they become operational in future. The existing solid Earth science NRIs that support the EPOS integration plan are listed in the RIDE database[[6]](#footnote-6). RIs contributing to EPOS will continue to be owned and managed at a national level. These have a significant economic value both in terms of construction and yearly operational costs, which are typically covered by national investments that must continue during EPOS implementation, construction and operation.
* Thematic Core Services (TCS) are community-specific integrations (e.g., seismology, volcanology, geodesy, experimental laboratories). They represent transnational governance frameworks where data and services are provided to answer scientific questions and where each community discusses their specific implementation, best practices and sustainability strategies as well as legal and ethical issues. The TCS were designed taking into account the requirements of the different EPOS communities. The fact that ten TCS contributed to EPOS demonstrates the multidisciplinary breadth of the integration plan and the potential impact of the community building aspect of EPOS. TCS will be interoperable with ICS thanks to appropriate ICT solutions (the compatibility layer).
* Integrated Core Services (ICS) represent the novel e-infrastructure that will allow access to multidisciplinary data, products (including synthetic data from simulations, processing and visualization tools), and services to different stakeholders, including but not limited to the scientific community (i.e., users). The key element of the ICS in EPOS will be a central hub (ICS-C) where users can discover and access data and data products available in the TCS and NRIs as well as access a set of service for integrating and analysing multidisciplinary data. The technical interface between TCS and ICS is the compatibility layer, which guarantees communication and interoperability. The ICS-C will also provide access to *distributed resources* which form the distributed ICS (ICS-D) and include access to supercomputing facilities as well as to visualization, processing and modelling tools that do not need to be centralised. ICS-D may be (a) additional computing/storage/detector array facilities outside the scope of EPOS; (b) nodes providing general software services used across all TCS such as input/validation, data management, analytics, simulation, mining, visualisation; (c) replicas/mirrors of ICS-C in distributed locations for resilience and performance.

In the Competence Centre the three use cases were tested against their potential inclusion as ICS-D.

# Prototype of the EPOS Authorisation Authentication Infrastructure (AAI)

## Overview

Authorisation, Authentication and Accounting Infrastructure (AAAI) is crucial for the proper functioning of the EPOS Infrastructure and their core components ICS-C, ICS-D and Thematic core services. For the sake of simplicity, the part of AAAI including Authorisation and Authentication Infrastructure (AAI) has been chosen to demonstrate that EGI Infrastructure services can act as ICS–D in the broad EPOS architecture. The EPOS AAI prototype was developed based on the UNITY IDM technology interfaced with EGI CheckIn. The EPOS Anthropogenic Hazards Thematic Core service has been used additionally to test ICS-D like functionality on EGI Infrastructure.

|  |  |
| --- | --- |
| **Name** | Prototype of the EPOS Authorisation Authentication Infrastructure (AAI) |
| **URL** | <https://epos-aai.cyfronet.pl/home/home> |
| **Description** | The service aims to provide the Authentication and Authorisation Infrastructure service prototype for EPOS. For this purpose, the AAI backbone of the EGI Applications on Demand Service[[7]](#footnote-7) has been adopted, the AAI prototype is based on Unity IDM technology. For the sake of simplicity, the EGI AAI is used as Identity provider for EPOS AAI prototype[[8]](#footnote-8). |
| **Value proposition of the demonstrator** | The prototype main aim was to demonstrate that EGI Infrastructure services can act as distributed part of the EPOS Integrated Core Services as well as Thematic Core services. In addition, it was demonstrated that EGI Check-In service can act as idP for EPOS infrastructure. |
| **Customer/user of the demonstrator** | Whole EPOS community, since Integrated Core Services, through Thematic Core services up to national (Earth science) research infrastructures. |
| **Scenario** | The scenario focuses on integration of EGI CheckIn with Anthropogenic Hazards Thematic Core Service in order to allow authentication and authorisation and data analysis execution on EGI Infrastructure. |
| **Success criteria** | A user can be authenticated on the prototype with EGI CheckIn or TCS AH credentials, then (once his/her affiliation is confirmed) it is authorised to run a job on EGI Infrastructure via TCS AH. |
| **User Documentation** | Internal, for developers only. Users’ level documentation will be available from ICS-C Users’ Guide. |
| **Technical Documentation** | [http://www.unity-idm.eu](http://www.unity-idm.eu/), see also <https://wiki.egi.eu/wiki/Applications_on_Demand_Service_-_architecture> |
| **Developer team** | ACC Cyfronet AGH, Istituto Nazionale di Geofisica e Vulcanologia |
| **License** | Open Source, Apache2 license |
| **Source code** | Not released yet. |

## Architecture

In order to provide a robust and efficient architecture, that will address the required users’ needs, the description of the use case, was defined and validated with the EPOS community together with general considerations:

1. Taking into account the distributed character of the EPOS infrastructure, including scientific data sets, high level services, computing facilities, etc. users should be provided with Authentication, Authorization Infrastructure (AAI) system capable to interact with all the EPOS resources.
2. Several TCSs, RIs, etc., have already developed AAI solutions. These will have to be connected to the general EPOS AAI in order to avoid incompatibility of existing services.
3. To connect existing AAI approaches in EPOS into one ecosystem, an AAI hub is needed which will assure interoperability between existing technologies. Hub technologies directly passing logins, passwords, etc. have been abandoned by principle, if favour of a modern, attribute based, solution. This will increase the general security among EPOS infrastructure and easy the connection of various existing AAI technologies Once user is authenticated within the infrastructure, all authorisation decisions should be made based on attributes only. The set of the EPOS specific attributes should be defined in the hub and obtained from there by the services.
4. The EPOS AAI is considering supporting EduGAIN’s IdPs and the following protocols LDAP, OpenID, x509 certificates.

The architecture in the Figure: 1 shows the EPOS AAI solution developed so far: it is based on the AAI backbone of the EGI Applications On Demand service.

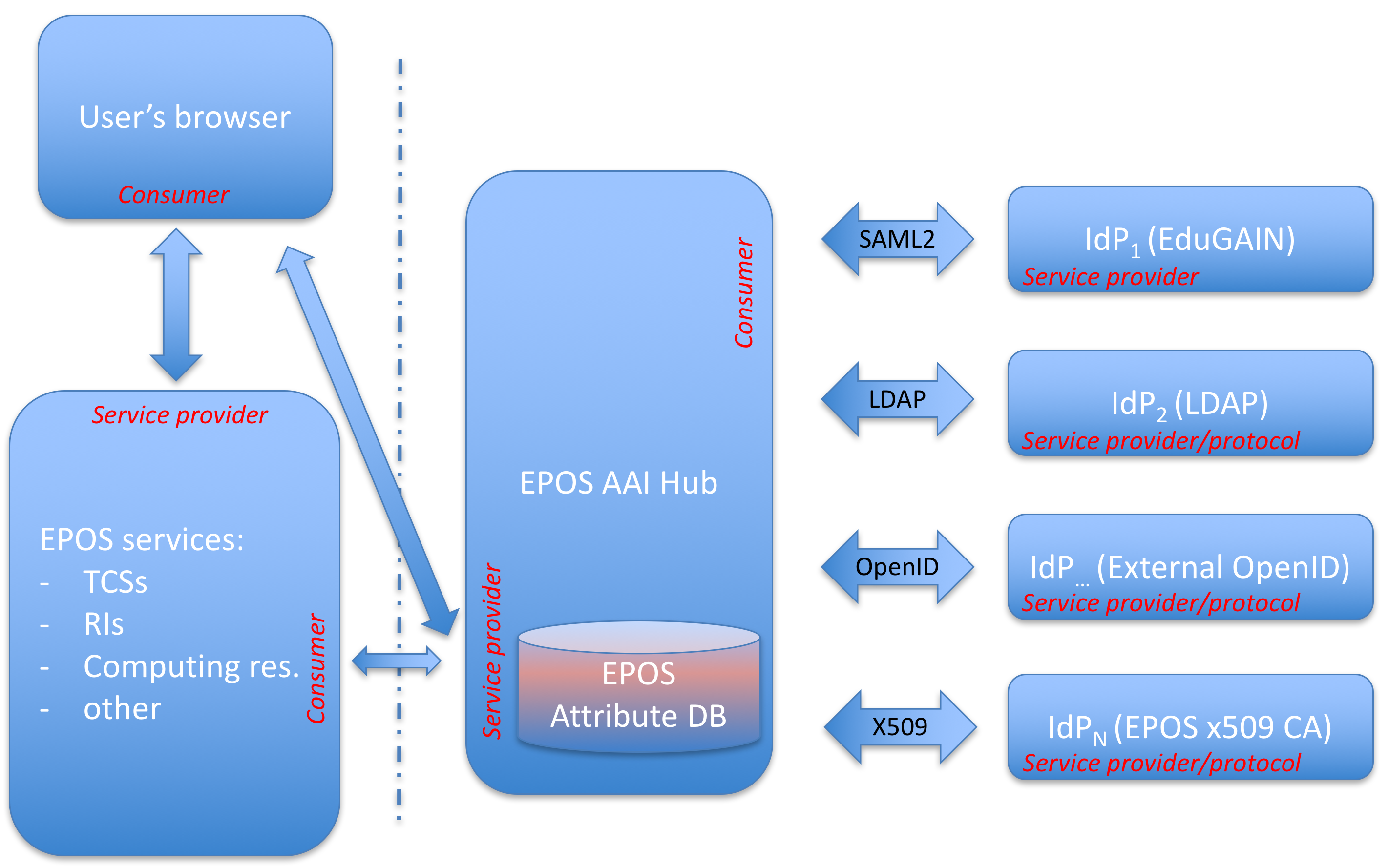


Figure: 1. The proposed AAI scheme for EPOS



### High-Level Service architecture

The prototype has been built according to the scheme sketched on **Error! Reference source not found.**. In the current implementation, two IdPs have been integrated as a proof of concept: EGI CheckIn and TCS AH, both OpenID type Identity providers. Additional IdPs will be integrated within the EPOS ICS-C after initial integration with the ICS-C science gateway. The thick arrows on the scheme represent any further communication (between TCSs, ICS-C and EGI CheckIn) after the authorization has been completed – this includes: job submission with data analysis, processing, advanced visualizations, etc.

The integration with the EGI CheckIn demonstrates how EPOS RI users could exploit EGI services and resources through the ICS in a transparent way (we used the Gaussian VO in EGI as an example). More coupled integration scenarios will be investigated in the future.

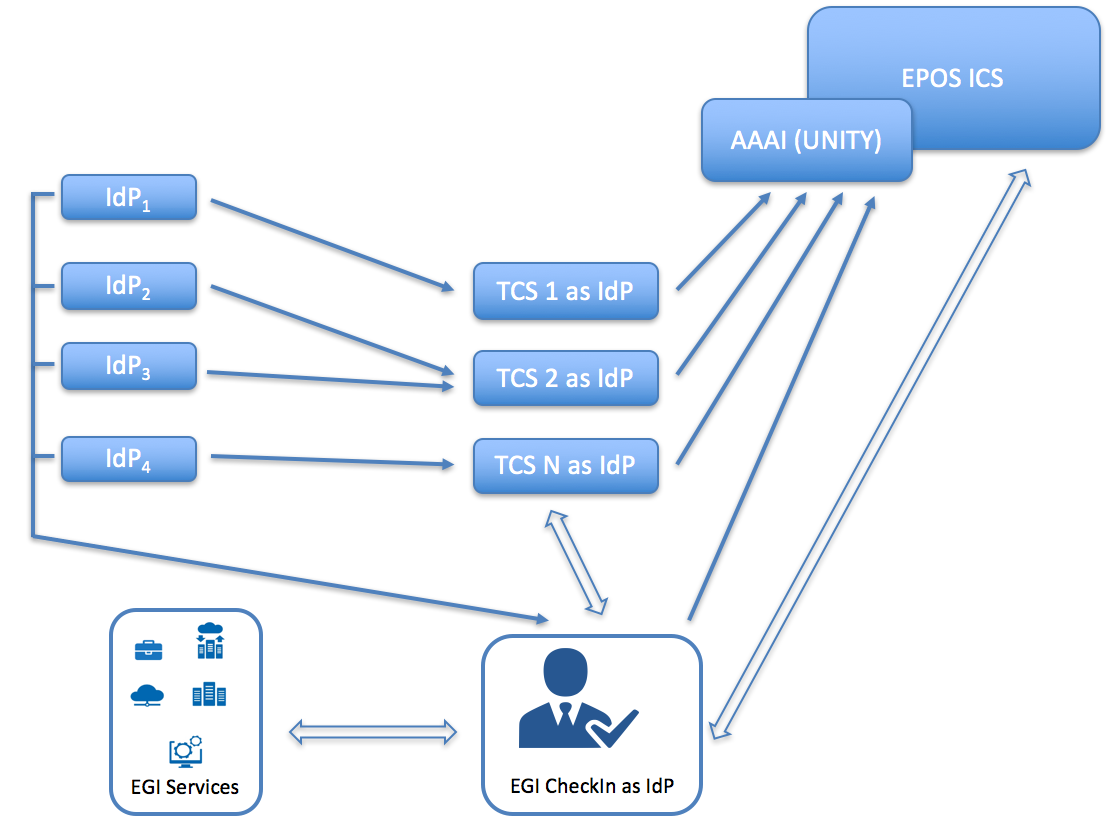


Figure: 2. EPOS AAI prototype scheme

## Demonstration

### Scenario

For the sake of simplicity, a basic scenario has been designed. A user accesses the ICS-C via a web browser. Then, to gain access to the ICS-C resources he/she has to log into the system. To achieve these three possible paths can be utilised:

* Local authentication (not considered for the pilot example)
* Authentication via credentials of one of the EPOS Thematic Core services (Anthropogenic Hazards, TCS AH in this example)
* Authentication via EGI Check-In service.

Once the user has been authenticated and his/her identity is confirmed by one of the IdPs he/she can gain access to ICS-C resources. Such an authentication enables to access also to the TCSs resources (data, data products, software and services, DDSS).

Authorisation to the specific EPOS resources is based entirely on tokens passing/translation (if needed). The same scheme will be utilised in case of EPOS ICS-D.

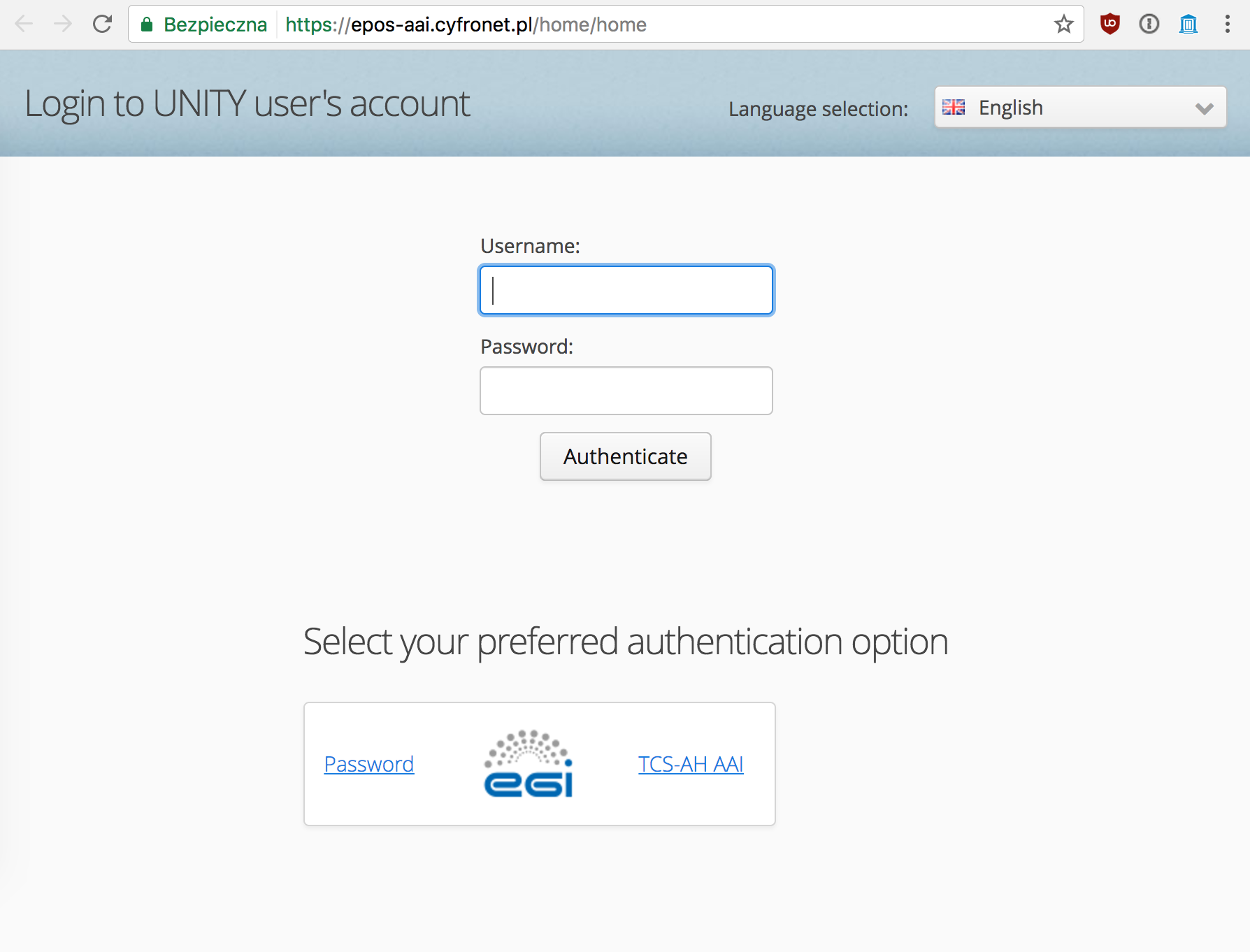


Figure: 3. The pilot's main page

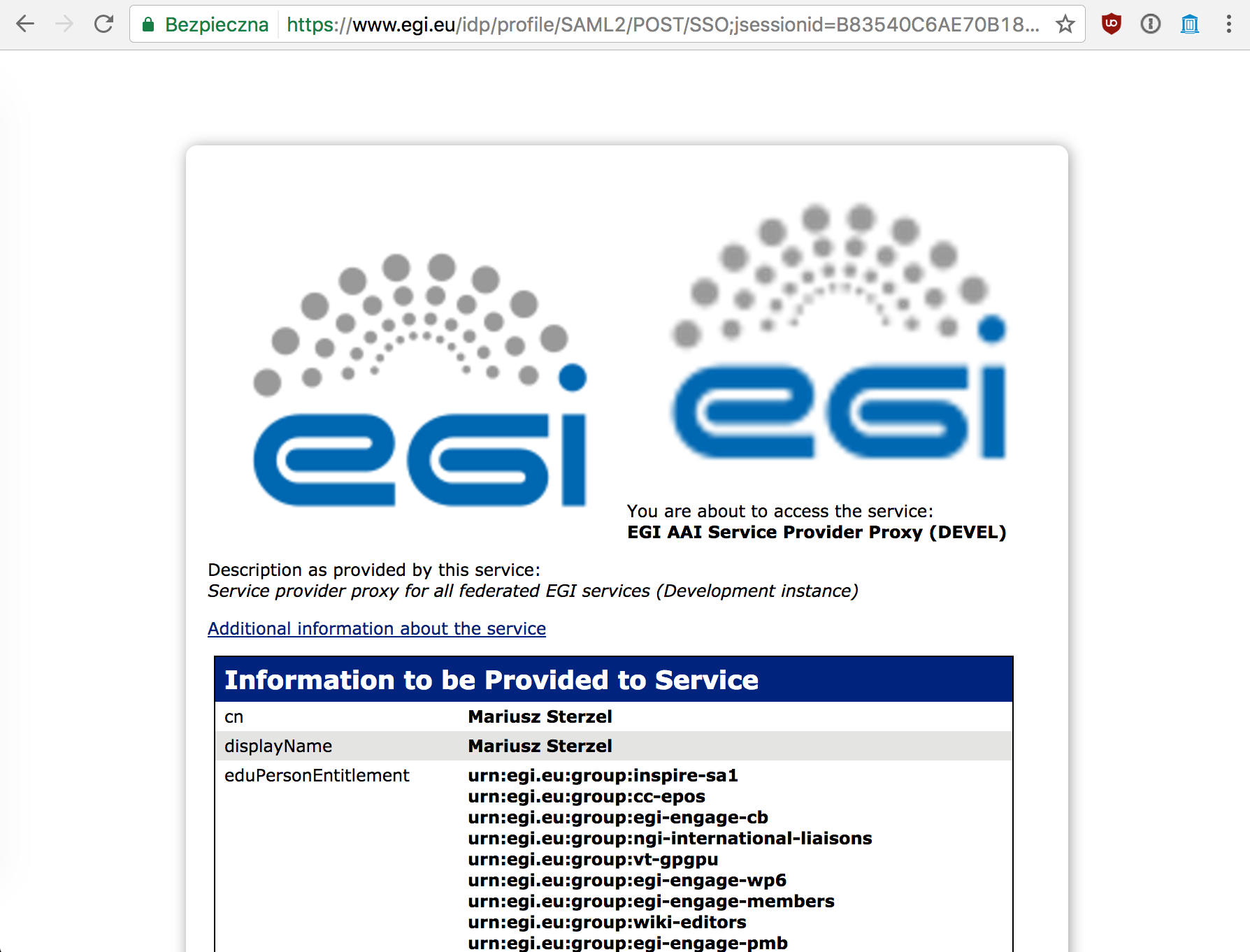


Figure: 4. Authentication walk-through step via EGI Check-In

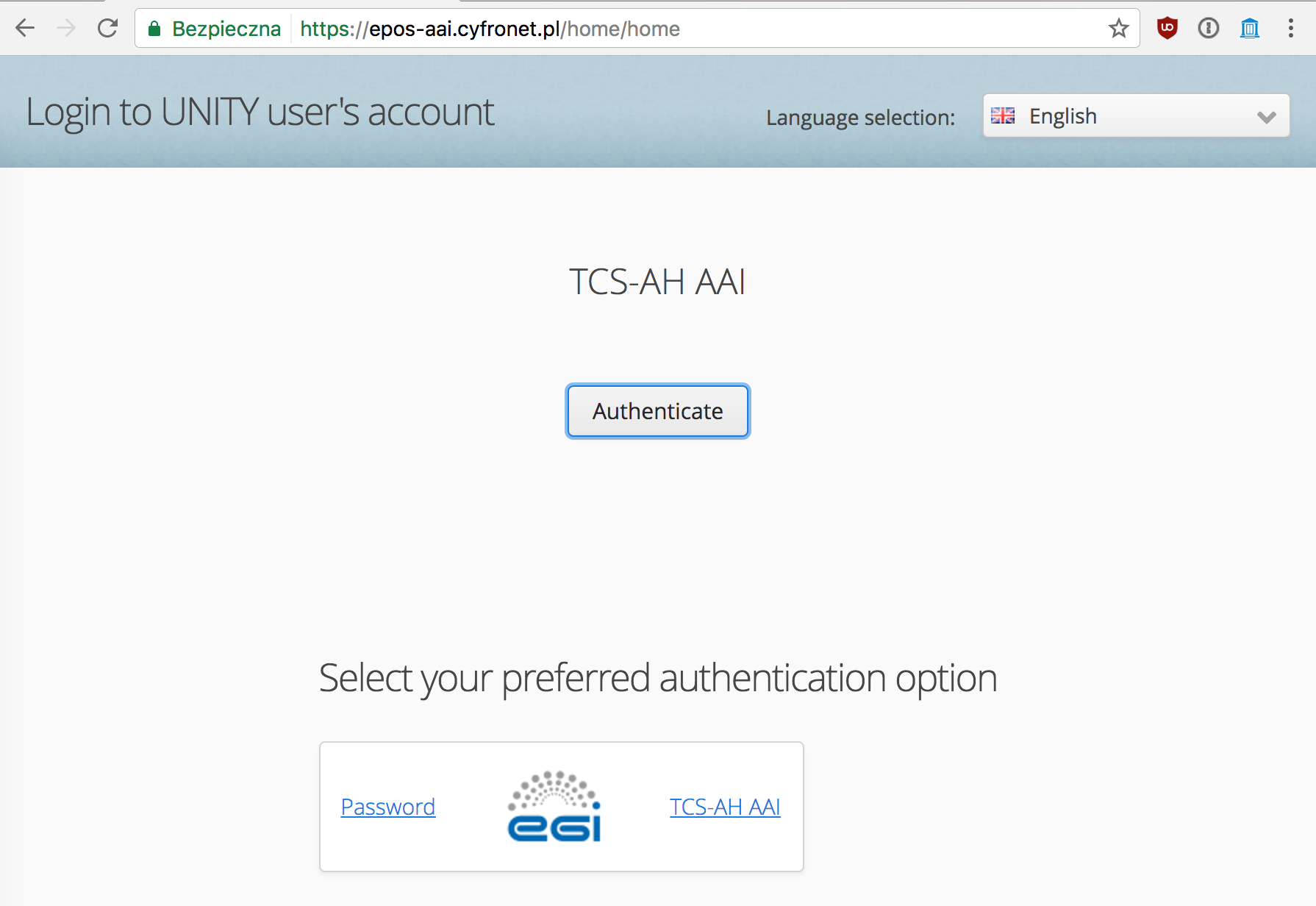


Figure: 5. Authentication walk-through via TCS AH path

### Feedback

The prototype was presented to the EPOS IT teams (EPOS IP project WP6 and WP7 packages). Access to the prototype is open for anyone with valid EGI or TCS AH credentials. After the presentation at the TCS-ICS meeting in Prague in March 2016, the prototype was accepted by EPOS WP6 and WP7 teams as a universal solution for the EPOS AA(A)I. The first version, integrated with the EPOS ICS-C portal, is expected in the fall of 2017.

## Future plans

The major planned step is the integration of the prototype with the EPOS ICS-C science gateway. As mentioned, according to EPOS IT teams’ plans, this step shall be done by fall of 2017. A second stage will include the integration of the remaining Thematic Core Services with the ICS-C platform. The final step will integrate the computing resources provided by EGI with ICS-C and TCSs via ICS-D services.

The EGI Applications On Demand Service is planning to move away from the currently used UNITY-robot proxy based AAI and will use CheckIn and CILogon instead. The new architecture will be based on more broadly used, and externally maintained components, and will also provide ‘AAI as a service’ by EGI for Research Infrastructures. EPOS will consider a similar move to continue using broadly accepted components within the EPOS AAI system.

# Computational Earth Science, (Earthquake simulation and MISFIT analysis)

Many communities within EPOS make use of science gateways to expose high-level scientific tools to experts and researchers. EPOS wants to investigate generic ways of creating these services as part of a more general and integrated portfolio of Computational Earth Science resources. For the EPOS CC we have selected a use-case consisting of an existing scientific application in computational seismology[[9]](#footnote-9). This is implemented with a general purpose Scientific Gateway (WS-PGRADE) and is characterised by two main steps:

**Earthquake Simulation**: Users setup simulations workflows that produce synthetic seismograms for public and custom earth models and earthquakes' parameters, via the execution of HPC simulation codes called solvers, such as Specfem3D Cartesian[[10]](#footnote-10).

**Raw data acquisition and Misfit**: The synthetic data is compared with real observations adopting data-intensive methods. This type of analysis presents three phases: raw data download, pre-processing and misfit calculations through different methods. Each phase is independent and the results can be re-used across phases, testing different methods. The raw data is accessed from federated community services, such as the FDSN[[11]](#footnote-11) network and stored within the user data-space, together with their processed versions and the resulting analytical results.

The EPOS CC will focus on the second part of the Use Case. It aims at the integration of FedCloud resources within science gateway hosting the application, and at the migration to this new infrastructure of the workflows required for the complete Misfit Analysis.

## Overview

|  |  |
| --- | --- |
| **Name** | EPOS Computational Earth Science – Seismology - VERCE Portal |
| **URL** | [http://portal.verce.eu](http://portal.verce.eu/) ; [http://www.verce.eu](http://www.verce.eu/) |
| **Description** | Virtual Earthquake and seismology Research Community e-science environment in Europe |
| **Value proposition of the demonstrator** | The platform generally aims at facilitating the integration of computational resources to be exploited by seismological applications, adopting standard interfaces towards cloud and HPC. Integrating the EGI Federated Cloud would extend the possibilities of procurement, accounting and sustainability of the service. |
| **Customer/user of the demonstrator** | Universities, International Research Centres and Earthquakes Monitoring Centres. |
| **Scenario** | The Competence Centre focused its activities on a Science Gateway for Computational Seismology (VERCE) offering tools, workflows, data-management and computational services in support of Forward Modelling analysis. The application allows the use of earth models for the simulation of seismic events and their evaluation and improvement through the misfit analysis with the observational data. |
| **Success criteria** | The main Criteria were the evaluation and integration of the FedCloud resources within the current science gateway, which is based on the WS-PGRADE gateway. After a successful integration, the MISFIT workflow should be migrated to the new resources. Unfortunately, the evaluation activity required an intense investigation of several shortcomings that prevented the gateway to configure the access and to connect to the remote cloud services. Solutions were identified and applied, in cooperation with the WS-PGRADE team, fostering the overall update of the gateway technology. However, the pilot could not reach the full integration and we will address the migration of the workflows onto cloud within the EOSC pilot framework |
| **User Documentation** | <http://www.verce.eu/Training/UseVERCE.php> |
| **Technical Documentation** | [http://www.verce-project.eu](http://www.verce-project.eu/) (Available. However, its access requires authentication) |
| **Developer team** | Fraunhofer SCAI, INGV, ULIV, KNMI, CNRS |
| **License** | MIT License |
| **Source code** | <https://github.com/KNMI/VERCE> |

## Architecture

The VERCE virtual research environment (VRE) requires communities driving the research, collections of relevant data, application software tuned and maintained to meet the latest research requirements and to exploit hardware advances, and teams of ICT experts maintaining the VRE’s advanced capabilities. VRE usability and sustainability is key to attracting these researchers, enabling their collaboration, creating effective interplay between ICT experts and domain-focused researchers, gaining access to the resources required and amortising costs over sufficiently broad communities. The design of the VERCE platform embraces this diversity.



### High-Level Service architecture

The platform offers a number of application specific services to perform operations such as simulation, raw data pre-staging from FDSN archives, pre-processing and MISFIT. These high level services are organised into independent workspaces, which operate through an interaction with platform’s webservices and workflows. The infrastructure serves these interfaces with computational and data resources. We provide below a high-level overview of the architecture including user interaction and data-flows. The efforts of the team are focusing on extending the Data-Intensive component shown by the illustration with EGI Federated Cloud capabilities.

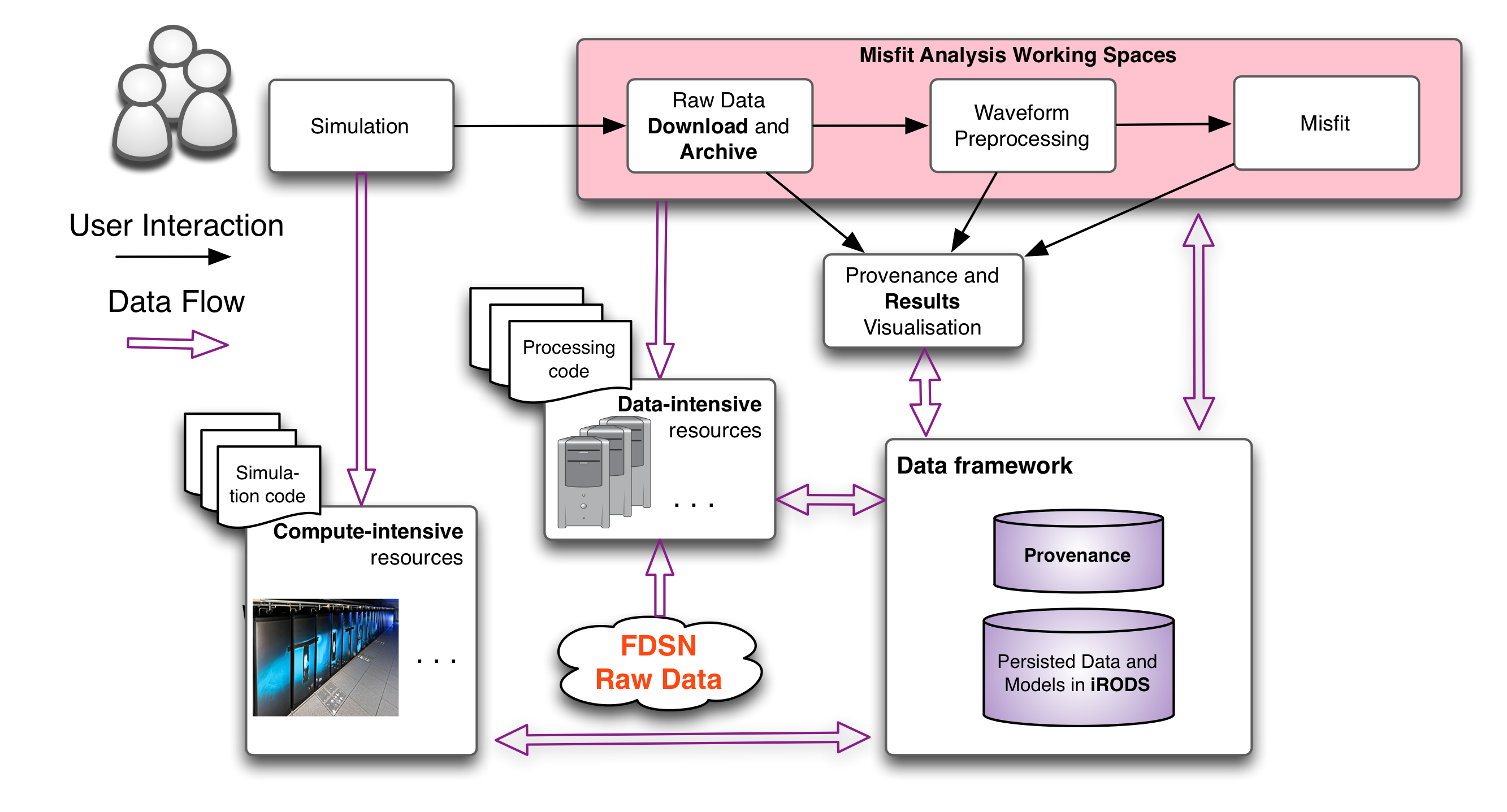


Figure: 6. The original VERCE VRE architecture (without EGI Cloud connected)

### Integration and dependencies

This section provides a system-centric overview of the VERCE integrated software and dependencies.

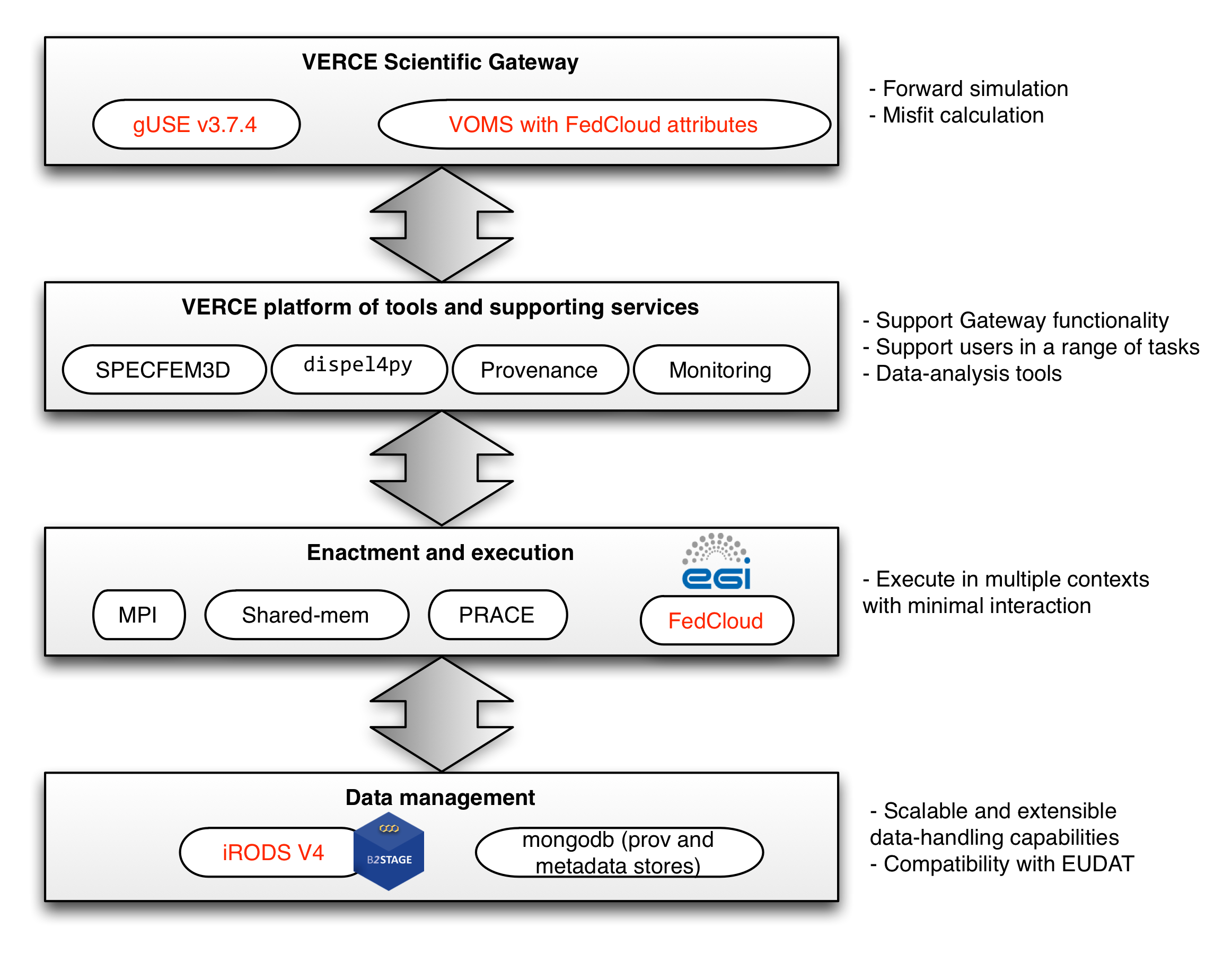


Figure: 7.VERCE integration and dependencies: In red the updated components

**VERCE Science Gateway**

The gateway is based on the gUSE technology, a Liferay-powered science-gateway framework. The support of VOMS and x.509 certificates is a clear security requirement imposed by many providers of computational resources providing facilities for the Enactment and Execution layer. The current gateway/portal technologies demonstrated to ease the management of user accounts and certificates, which required minor efforts from the development team.

**Platform tools and supporting services**

* **SPECFEM3D:** a widely used tool for simulating seismic wave propagation at regional scale. As requested by the seismology research community, the workflows have been extended with the support of an additional simulation wave propagation code (**SPECFEM Globe**), which addresses global seismology studies.
* **dispel4py:** Python framework for describing abstract data-intensive workflows. It is used for all the data processing tasks in shared-memory as well as MPI-powered clusters. Incremental updates have been performed on the library, in order to improve the overall performance of the analysisand provide a better management of lineage information.
* **S-ProvFlow provenance service:** It is used on the gateway and it interacts with a number of architectural components, such as the corresponding data management elements shown in a later section, (“Data Management”). Its conceptual model is based on the W3C-PROV standard[[12]](#footnote-12), which is also offered as export format. The service, (S-ProvFlow[[13]](#footnote-13)) includes a database, a WEB API and a visualization interface. Access methods are inspired by real usage scenarios within the science gateway, but are also extended according to the analysis of cross-disciplinary use cases. The latter is conducted with special attention to the use cases identified by the RDA (Research Data Alliance) Provenance WG.
* **Monitoring of execution:** A functionality offered by the gateway through the combined access to the jobs’ status and extensive lineage information collected at runtime. A number of monitoring interface have been developed targeting different access modes and visualizations. The latter is obtained thanks to the provenance.

**Enactment and Execution**

These are architectures and computational services that the VERCE solution is able to exploit, thus adding to sustainability and standardization. We extended this portfolio with the resources provided by the **EGI Federated Cloud**, which required to evaluate the impact on the existing implementation, identifying and addressing, when possible, the shortcomings that prevented its technical integration

**Data Management**

The Data Management leverages on a collection of technologies and workflows that control the exchange of data and metadata between the computational clusters and the user data-space. One of the fundamental technologies is the iRODS[[14]](#footnote-14) system, a rule-based data-management system enabling storage and access at scale with configurable access policies. In the current setup we updated to the iRODS v4. This brought improvements also to the GridFTP support for data-staging operations to and from iRODS, thanks to the integration of the **EUDAT B2STAGE** technology. iRODS stores the physical data produced by the simulations and analysis and pre-stages the observational data required for the MISFIT. Provenance and metadata are stored within MongoDB[[15]](#footnote-15), a distributed and scalable document database, which is exposed to the other components of the platform through a web-service API.

## Demonstration

### Scenario and current status

The use case will run on the SCAI Federated Cloud site, which is already configured in the VERCE Gateway’s DCI Bridge as "SCAI-FedCloud" resource. Cloud resources can be accessed by users who are registered members of the *verce.eu* VO. The simulated and observed data will be staged by accessing pre-computed collections. These may be already available within the VERCE data-management layer or downloaded on-demand from the FDSN[[16]](#footnote-16) web services, through the execution of a Data-download workflow. The parameters of the Data-download workflow are automatically configured thanks to the information obtained from the lineage collected during the simulation. This will make sure that time-window and geographical extents of the observational records match the synthetic data. The next step will allow the interactive configuration and execution of Processing and Misfit workflows. Users are enabled to control and monitor the workflows through dedicated workspaces.

However, the deployment of the new version of the **gUSE** gateway, is still presenting technical shortcomings. At the current status, although the gateway reaches the successful instantiation of the FedCloud VM, connectivity issues are still experienced during the execution of the workflow’s job. This issue will be tackled, involving the partners responsible of the middleware technologies, evaluating changes and suggesting improvement to the current service and software.

### Feedback

The demonstrator will be finalized later this year. The team will address the pending issues with the additional contribution of the EOSC initiative (European Open Science Cloud). EOSC has selected the VERCE gateway as one of the beneficiaries for financial support as the outcome of the recent call for the implementation of cloud-based pilots. This should be positively interpreted as the continuative interest of the target community in pursuing the Cloud integration and the acknowledged technical value of the service from the European e-infrastructure. The pilot will pursue also the application of the FAIR-data principles through the generation and management of provenance during the execution of each workflow and the incremental exploitation of e-infrastructure services for a more general-purpose identification and management of the scientific artifacts (e.g. B2HANLE, B2SHARE). This will open the data to additional scientific stakeholders. Once the cloud integration will be completed, evaluation will be conducted together with the target user community, possibly within dedicated trainings and webinars.

## Future plans

More communities within EPOS are looking with interest at the capabilities of the platform and at the outcome of the EGI use case. Currently a plan is developing to include Volcanological research applications in the VERCE framework, with the scope of scaling all the generic components to a larger user-base and moving towards the realization of the EPOS Computational Earth-Science strategy. With the support of the EOSC, which will offer resources for the developments (1FTE for one year), the team aims at offering a production deployment of the service for the EPOS community. This can be seen as an immediate outcome of the pilot. The pilot will address challenges that emerge from real-life computational scenarios in seismic assessment, where each experiment involves hundreds of GB of data and metadata (aiming at ~20TB storage). The heterogeneous datasets will be generated in HPC and Clouds, acquired from federated institutional archives, and moved across computational resources and dedicated storage sites. The demonstrator wants to expand the portfolio of resources with additional institutional providers, for instance increasing the access to more EGI FedCloud nodes.

# Satellite Data (P-SBAS InSAR Sentinel-1 TOPS)

## Overview

|  |  |
| --- | --- |
| **Name** | P-SBAS InSAR Sentinel-1 TOPS |
| **URL** | https://geohazards-tep.eo.esa.int/ |
| **Description** | The demonstrator connected the thematic exploitation platform with the EGI Federated Cloud to support a specific use case, named P-SBAS.Such a use case aims at making Sentinel-1 data accessible on a scalable compute platform.  P-SBAS stands for Parallel Small BAseline Subset and it is a DInSAR processing chain for the generation of Earth deformation time series and mean velocity maps, which uses as Input SLC (Level-1) Sentinel-1 data. |
| **Value proposition** | Surveillance service for automatic and systematic generation of earth surface displacements. The integration of the Geohazard TEP with the EGI Cloud resources enables the massive processing of Sentinel data at an unprecedented scale and the sharing of the outputs between many research scientific communities. |
| **Customer/user of the demonstrator** | Scientific community representatives, Governments, and Industries. |
| **Scenario** | The P-SBAS InSAR Sentinel-1 TOPS processing chain will be executed over the EGI Federated Cloud through the ESA Geohazard exploitation platform. The EGI Federated Cloud will provide all the computing and storage resources needed for the pilot.  Pilot results will be advertised within the EPOS and EO community. EO community will be reached via ESA. |
| **Success criteria** | The pilot will be actually executed exploiting EGI Federated Cloud resources that will provide enough compute resources to perform a massive computation of all data related to the monitored areas enabling surveillance services.  The EPOS EO community will access the Geohazard Exploitation platform to exploit the results. |
| **User Documentation** | n.a |
| **Technical Documentation** | n.a. |
| **Developer team** | CNR-IREA |
| **License** | Closed source license |
| **Source code** | n.a. |

## Architecture



### High-Level Service architecture

The P-SBAS InSAR Sentinel-1 TOPS demonstrator will run on top of the Geohazard TEP that has been developed outside the EGI-Engage project. The Geohazard TEP (shortly GEP) implements the ESA Thematic Exploitation Platform architecture[[17]](#footnote-17), see the figure 8 below.

Terradue is the lead partner of the consortium that has developed the GEP and is also the provider of the service.

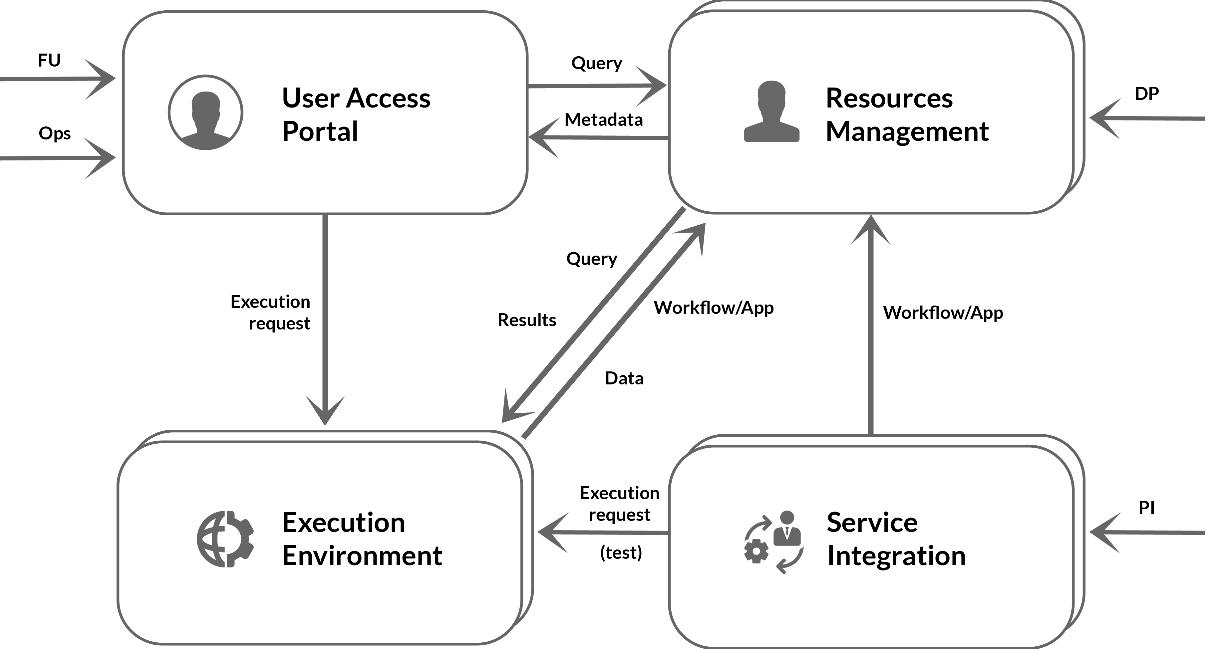


Figure: 8. ESA Thematic Exploitation Platform architecture

Figure 9 is an example of data visualised through the GeoBrowser of the Geohazard exploitation platform.

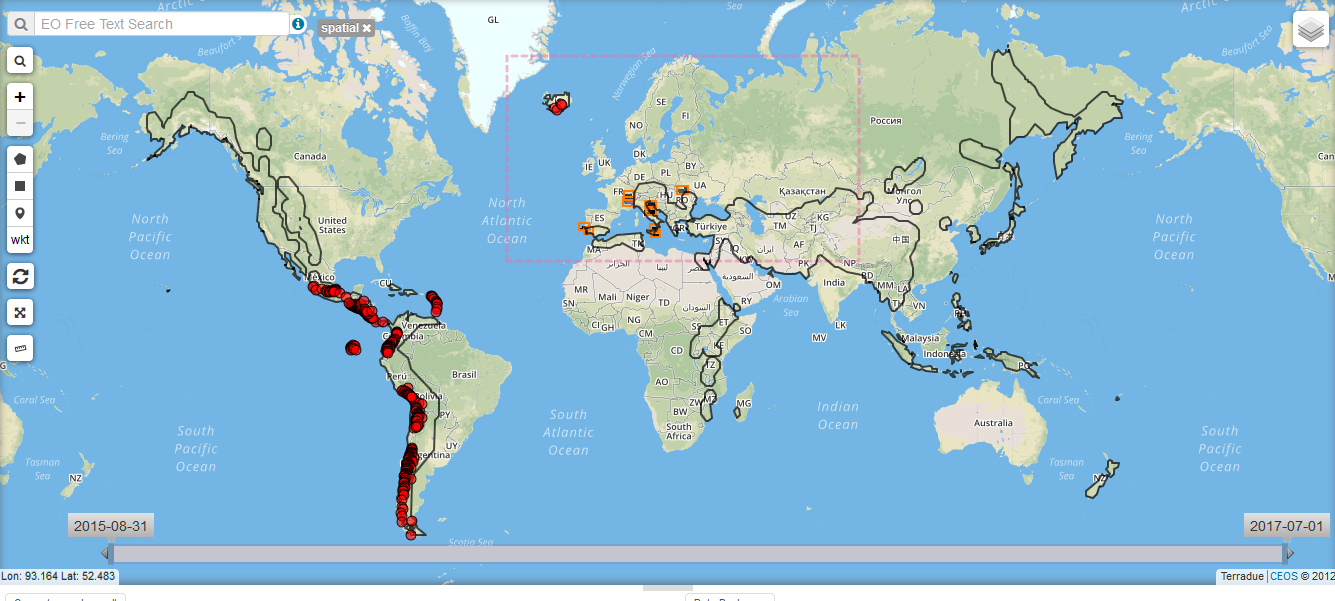


Figure: 9. Seismic and volcanic areas showed in the GeoBrowser of the GEP

## Demonstration

### Scenario

P-SBAS stands for Parallel Small Baseline Subset and it is a DInSAR processing chain for the generation of Earth deformation time series and mean velocity maps, which uses as Input SLC (Level-1) Sentinel-1 data. Such processing chain generates an output in CSV[[18]](#footnote-18) format with the following structure: Line Of Sight (LOS) Displacement time series, Mean LOS Velocity, Temporal Coherence, Average scattered elevation (Topography). The service can also generate wrapped and unwrapped interferograms that are delivered in geoTiff format.

The P-SBAS InSAR Sentinel-1 TOPS processing chain will be executed over the EGI Federated Cloud through the ESA Geohazard exploitation platform. The EGI Federated Cloud will provide all the computing and storage resources needed for the pilot.

This processing chain will perform a massive computation of Sentinel-1 data of the geographical areas of interest. The chain will continuously process most recent Sentinel-1 data providing a full picture over the time of the monitored zones. The data will be copied to the EGI Federated Cloud infrastructure when needed and deleted after the computation.

For the processing, a dedicated cluster was already deployed in the BEgrid-BELNET cloud resource. The cluster consists of:

* 1 VM with 2 CPUs, 8 GB RAM, 100 GB of local disk
* 4 VMs with 8 CPUs, 32 GB RAM, 100 GB of local disk
* 4-6 TB of shared block storage

VMs were instantiated using resources of the production VO geohazard.terradue.com and are managed via the Geohazard TEP middleware stack.

Currently, CNR-IREA is fine-tuning the processing chain to release a new improved version, which should be available in the future. Once available, it will be deployed in the cluster and the massive computation will start.

Pilot results will be advertised within the EPOS and EO community. EO community will be reached also via ESA events and the Thematic Exploitation Platform collaboration framework.

### Feedback

The demonstrator will run later this year for the EPOS community. A feedback on satisfaction will be requested to the users that will try the service.

## Future plans

The P-SBAS InSAR Sentinel-1 TOPS demonstrator will run as soon as the CNR-IREA completes the deployment of the processing chain over the already available cluster in the EGI Federated Cloud.

The VO SLA agreement between Terradue[[19]](#footnote-19) and EGI guarantees the availability of the cloud resources until the end of 2017, consequently the pilot could be executed, at least, until the end of this year. Terradue intends to request a renewal of such agreement in order to continue the pilot in 2018.

During the execution of the pilot, we will invite people of the EPOS community to test and validate the service. Pilot results will be also advertised via ESA.

In addition, the integration of the Geohazard Exploitation Platform with the EGI Federated Cloud will be further exploited scheduling and running new Satellite Data pilots according to the needs of the EPOS community.

As next steps of the collaboration with Terradue, EGI is studying a solution to directly provide the Copernicus Data running on top of the GEP in its infrastructure. This would improve the current deployment with a direct benefit also for the EPOS use cases.

# Conclusions, lessons learnt, future work

The prototype of the EPOS Authorization Authentication Infrastructure (AAI) will be integrated with the EPOS ICS-C by fall 2017. The prototype can be improved and reused by the EPOS TCS communities. One of the main lessons learned concerns the complexity of EPOS: involving the actual data providers is an action that should be undertaken involving both the EPOS main coordination node, the ICS-C, and the communities of the TCS. To do this appropriate procurement policies are required. EPOS will be able to define such policies only after the EPOS-ERIC is established (the ERIC process is planned to start in 2018). ). However, the useful demonstration of AAI capability needs to be enhanced for the EPOS production phase to include stronger attribute-based authentication using the EPOS catalog.

As for the Earthquake simulation use case, communities within EPOS are looking with interest at the capabilities of the platform and at the outcome of the EGI use case, as in the case of the Volcanological community. To sustain this interest it is important to find appropriate means to scale all the generic components to a larger user-base and going towards the realization of the EPOS Computational Earth-Science strategy. A possible context where this action can be further explored is the EOSC, where such use case found a few resources to be carried on, and where team will aim at offering a production deployment of the service for the EPOS community, building on the outcome of the pilot. The need for continued work on this CES demonstrator indicates the complexity of integrating a single ICS-D environment into the EPOS ICS framework (s well as the complexity of the CES environment). Integration of multiple heterogeneous ICS-ds (with or without a CES workload) will prove challenging. The ‘touch points’ should be documented to guide this future work.

As for the satellite data use case, it will run until the end of the project and appropriate actions will be undertaken in order to involve people of the EPOS community to test and validate the service. The Geohazard Exploitation Platform with the EGI Federated Cloud will be further exploited scheduling and running new Satellite Data pilots according to the needs of the EPOS community. Here, the integration of the ESA environment with the EGI environment has ‘touch points’ that should be documented to guide future work.

Generally, these ‘touch-points’ are:

1. AAI at the user interface (authentication) and subsequently at all levels of workflow deployment (authorization);
2. User interface attribute value collection for specifying the workflow;
3. Interface from EPOS catalog to assets if within an ICS-D (e.g. data, computational resources);
4. Partitioning of description of CES and ICS-d within the EPOS catalog when they are closely inter-related;

Based on the achievements of the three pilots, the following opportunities emerged for EPOS to organize the separate components into a single infrastructure:

1. Establish a VO in EGI to pool resources from EGI providers for the EPOS community. This EPOS VO could include the SCAI and BEGRID providers which were used already in the 2nd and 3rd pilot, and also additional partners from EGI and EPOS. This resource pool could potentially become an Earth science-specific computational infrastructure in the European Open Science Cloud.
2. Connect the clouds of the EPOS VO to the EPOS AAI system as Service Providers. This would empower EPOS community members to access those resources seamlessly and in a harmonized way, using institutional/EPOS accounts and attributes.
3. Promote the integrated cloud-AAI setup for the EPOS sub-communities. Articulate the system as a platform where communities can host and share VMs and data with each other. The EGI Applications Database and DataHub can serve as the ‘Shop window’ and ‘Registry’ for these applications and datasets. High level science gateways can utilize the applications, datasets and the cloud to offer customized, thematic environments.
4. Define and run additional pilots to cover relevant scenarios/features that were not covered by the EGI-Engage CC pilots (such as AAI demonstrator where EPOS accounts are used to access EGI resources; AAI demonstrator to test EduGAIN, LDAP and x509 certificates in service access.)

1. ICS-D is a terminology used in the EPOS infrastructure architecture. [↑](#footnote-ref-1)
2. A current assumption is that compatibility between UNITY IDM and EGI CheckIn is enough to state that EGI services can operate as ICS-D. Clear requirements for being an ICS-D still need to be drafted by EPOS. [↑](#footnote-ref-2)
3. The selected demonstrator is P-SBAS InSAR Sentinel-1 TOPS developed by CNR-IREA. P-SBAS stands for Parallel Small BAseline Subset and it is a DInSAR processing chain for the generation of Earth deformation time series and mean velocity maps, which uses as Input SLC (Level-1) Sentinel-1 data. [↑](#footnote-ref-3)
4. The e-Collaboration for Earth Observation (e-CEO) platform: <https://wiki.egi.eu/wiki/EGI-Engage:WP5#TASK_SA1.3_Integration.2C_Deployment_of_Grid_and_Cloud_Platforms> [↑](#footnote-ref-4)
5. <https://www.epos-ip.org/> [↑](#footnote-ref-5)
6. <https://www.epos-ip.org/ride> [↑](#footnote-ref-6)
7. EGI Applications On Demand service: [http://access.egi.eu](http://access.egi.eu/) [↑](#footnote-ref-7)
8. The pilot focussed on using an EPOS service with an EGI user account. In general, bidirectional relationship is expected, i.e. being able for a user to use an EGI service with an EPOS user account. [↑](#footnote-ref-8)
9. <http://portal.verce.eu> [↑](#footnote-ref-9)
10. <https://geodynamics.org/cig/software/specfem3d> [↑](#footnote-ref-10)
11. <http://www.fdsn.org> [↑](#footnote-ref-11)
12. <http://www.w3.org/TR/prov-overview> [↑](#footnote-ref-12)
13. <https://github.com/KNMI/s-provenance> [↑](#footnote-ref-13)
14. [http://irods.org](http://irods.org/) [↑](#footnote-ref-14)
15. [http://www.mongodb.org](http://www.mongodb.org/) [↑](#footnote-ref-15)
16. <http://www.fdsn.org/> [↑](#footnote-ref-16)
17. <http://go.egi.eu/EP-OpenArchitecture> [↑](#footnote-ref-17)
18. Comma Separated Values [↑](#footnote-ref-18)
19. VO SLA Terradue: <https://documents.egi.eu/document/2763> [↑](#footnote-ref-19)