

**EGI-Engage**

Report on evaluation of EPOS EGI pilot studies

D6.23

|  |  |
| --- | --- |
| **Date** | 16 August 2017 |
| **Activity** | SA2 |
| **Lead Partner** | INFN/INGV |
| **Document Status** | DRAFT |
| **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 earth quake simulations and (3) making Sentinel-1 data accessible on a scalable compute platform. The document also draws conclusions and lessons from the pilots and outlines possible future work for future 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:** |  |  |  |
| **Reviewed by** |  |  |  |
| **Approved by:** |  |  |  |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
| **v.1** | 18/08/2017 | First full draft 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 |
| **...** |  |  |  |
| **...** |  |  |  |
| **v.n** |  |  |  |

**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

1.1 Background and motivations 6

1.2 EPOS infrastructure 6

2 Prototype of the EPOS Authorisation Authentication Infrastructure (AAI) 7

2.1 Overview 7

2.2 Architecture 8

1.1.1 High-Level Service architecture 9

2.3 Demonstration 10

2.3.1 Scenario 10

2.3.2 Feedback 13

2.4 Future plans 13

3 Earthquake simulation (MISFIT) 14

3.1 Overview 14

3.2 Architecture 14

1.1.2 High-Level Service architecture 15

1.1.3 Integration and dependencies 15

3.3 Demonstration 17

3.3.1 Scenario and current status 17

3.3.2 Feedback 18

3.4 Future plans 18

4 Satellite Data 19

4.1 Overview 19

4.2 Architecture 20

1.1.4 High-Level Service architecture 20

4.3 Demonstration 21

4.3.1 Scenario 21

4.3.2 Feedback 22

4.4 Future plans 22

5 Conclusions, lessons learnt, future work 22

**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 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, as well as the demonstrations that were given to the EPOS community based on the pilots.

The pilots highlight how EPOS could benefit from EGI services and provide the first suggestions about how services from the EGI infrastructure could be connected to the EPOS Research Infrastructure as Distributed Integrated Core Services (ICS-D)[[1]](#footnote-2). 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-3). The EPOS AAI prototype was developed based on the UNITY IDM technology, and interfaced with EGI CheckIn.
* Earthquake simulation (MISFIT): this pilot shows how an existing application, MISFIT from the field of Computational Seismology, can be improved by integration with the EGI Federated Cloud. Furthermore, it also 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: 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-4) 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-5)).

The demonstrators used the CheckIn security proxy service (1st demonstrator), and two federated cloud site 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-6), 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 already existing computational resources and specific services coordinated and technically 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 EPOS and EGI infrastructure to properly interoperate.

Members of the CC were selected by their competence and involvement in the main field EPOS wanted to pilot, that is to say harmonised provision of AAI solutions, for which Cyfronet was reckoned to be one of the main players, provision of computational resources in the framework of computational seismology, which triggered the participation of the main stakeholder in the previous VERCE project (i.e. KNMI, IPGP - CNRS, SCAI Fraunhofer, GRNET), and finally some of the actors already involved in joint EPOS / EGI development (satellite use case) (CNR).

## EPOS infrastructure

The EPOS functional architecture is composed of three connected technical and organizational elements: NRIs, TCS, ICS:

* 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 as they become operational can be integrated in future. The existing solid Earth science NRIs that support the EPOS integration plan are listed in the RIDE database[[6]](#footnote-7). 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 need not 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) replicates/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

|  |  |
| --- | --- |
| **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-8) 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-9). |
| **Value proposition of the demonstrator** | *Describe how the  demonstrator/pilot alleviates specific user pains and/or would support future users/customer(s) to exploit new opportunities.* |
| **Customer/user of the demonstrator** | Whole EPOS community, since Integrated Core Services, through Thematic Core services up to national (Earth science) research infrastructures. |
| **Scenario** | *Describe the scenario in which the demonstrator/pilot was used.* |
| **Success criteria** | *What were/are the success criteria for the demonstrator/pilot? Did the demonstrator/pilot meet these criteria?* |
| **User Documentation** | *Provide url if applicable* |
| **Technical Documentation** | <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, after the description of the use case, requirements were gathered from the community, and general consideration were done:

1. Taking into account dispersed 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 some AAI solutions. These will have to be connected to the general EPOS AAI in order to avoid break-up of already existing services.
3. To connect existing AAI approaches in EPOS into one ecosystem, AAI hub is needed which will assure interoperability between existing technologies. Hub technologies passing logins, passwords, etc. have to be abandoned by principle. Instead a modern, attribute based, solution is advised. 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 IdPs that EPOS AAI is considering to support are EduGAIN, LDAP, OpenID, x509 certificates.

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

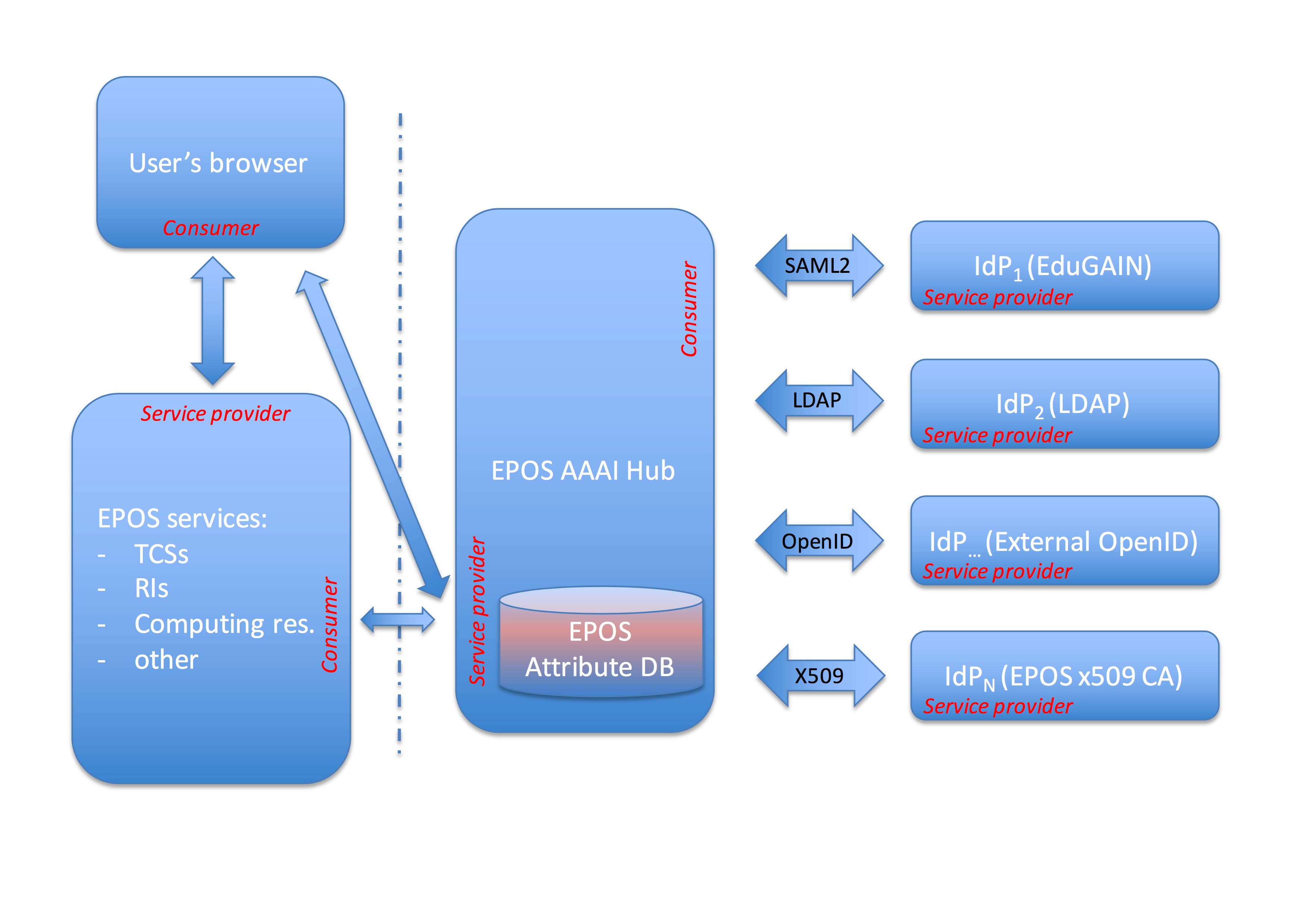


Figure 1. The proposed AAAI scheme for EPOS

### High-Level Service architecture

The prototype has been built according to the scheme sketched on Figure 2. 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. I.e. this includes job submission with data analysis, processing, advanced visualisations, etc.

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.

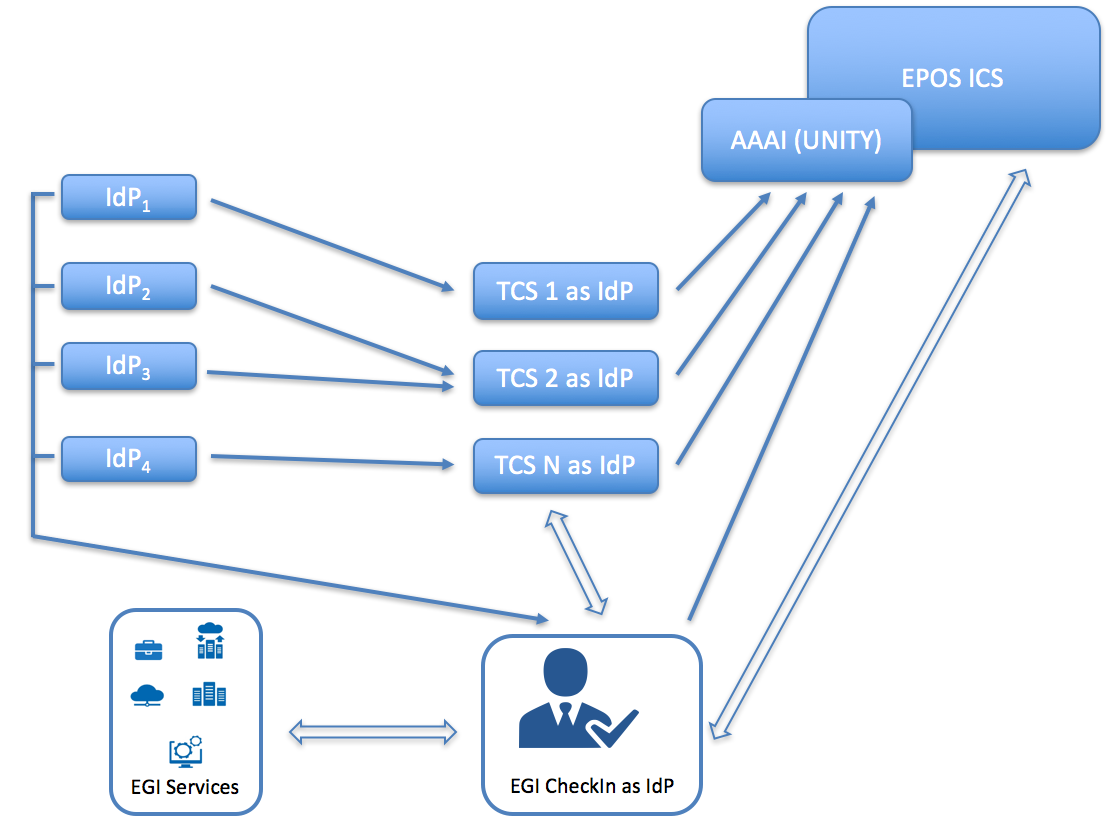


Figure 2. EPOS AAI prototype scheme

## Demonstration

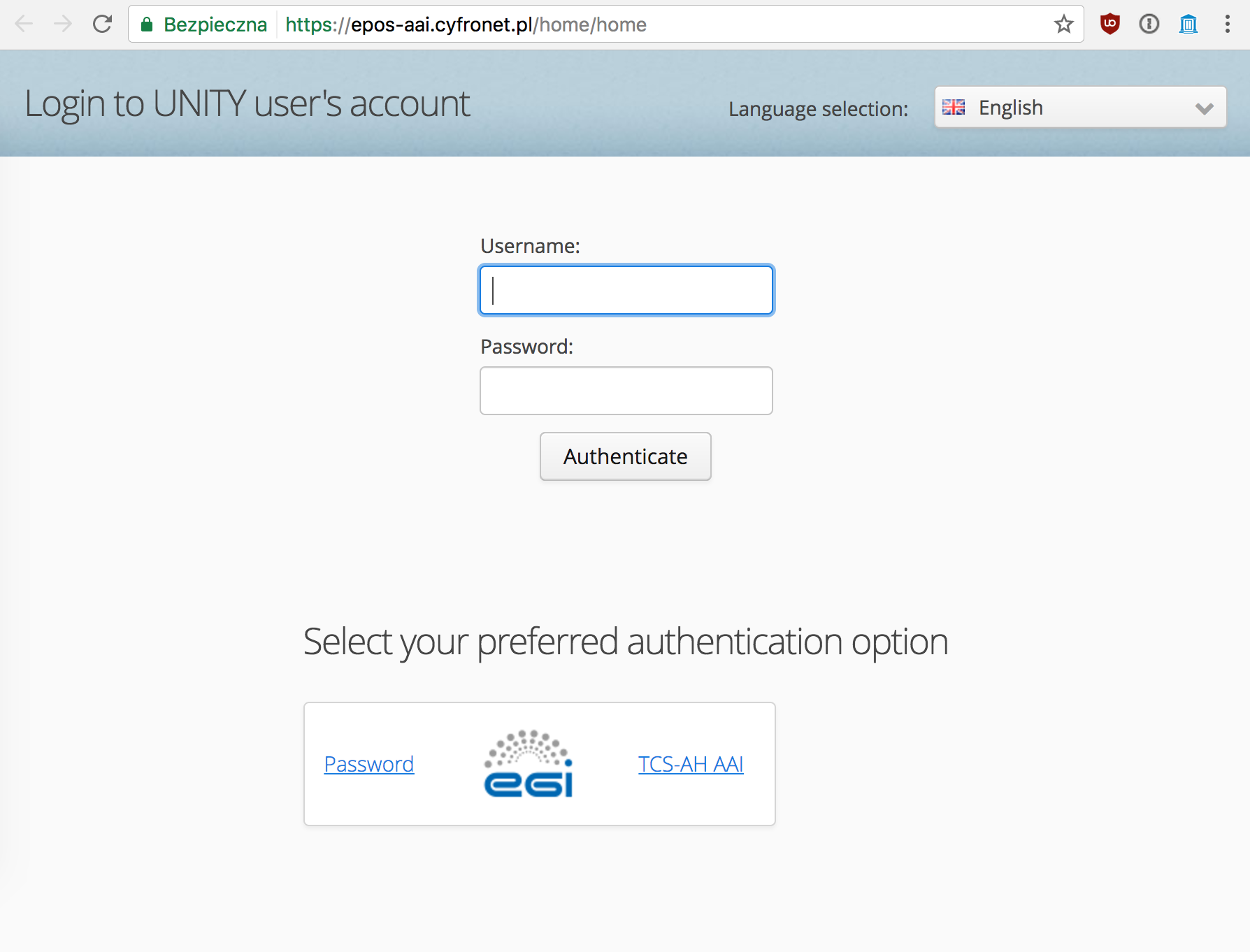
### Scenario

For the sake of simplicity, a basic scenario has been set. In general, 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. Three possible paths can be utilised then:

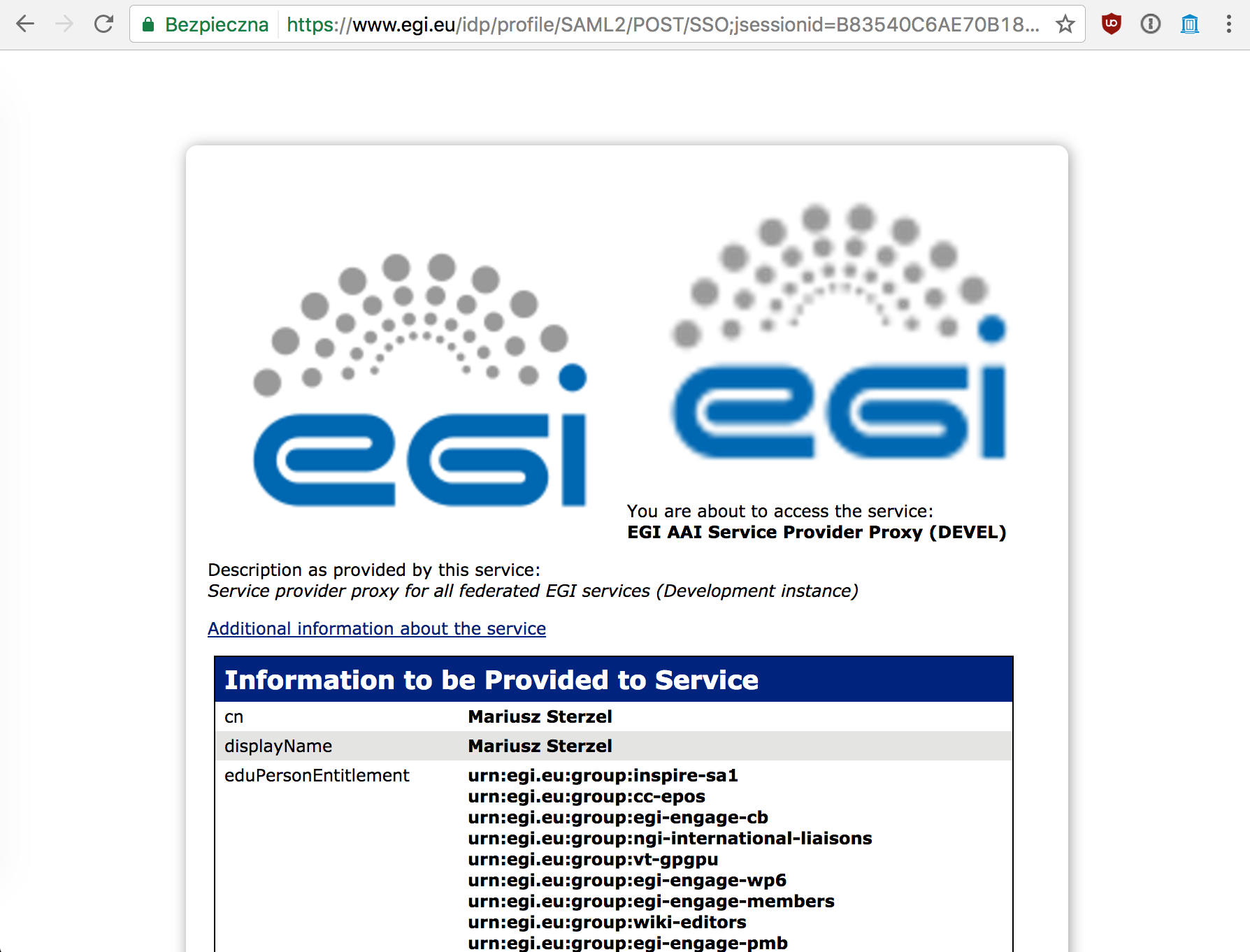
* Local authentication (not considered for the pilot example)
* Authentication via credentials of one of the EPOS Thematic Core services (Anthropogenic Hazards, TCS AH in this example)
* Authentication via EGI Check-In service.

Once the user has been authenticated and his/her identity conformed by one of the IdPs he/she can gain access to ICS-C resources and, going deeper into EPOS e-infrastructure also gain access to the TCSs resources (data, data products, software and services, DDSS).

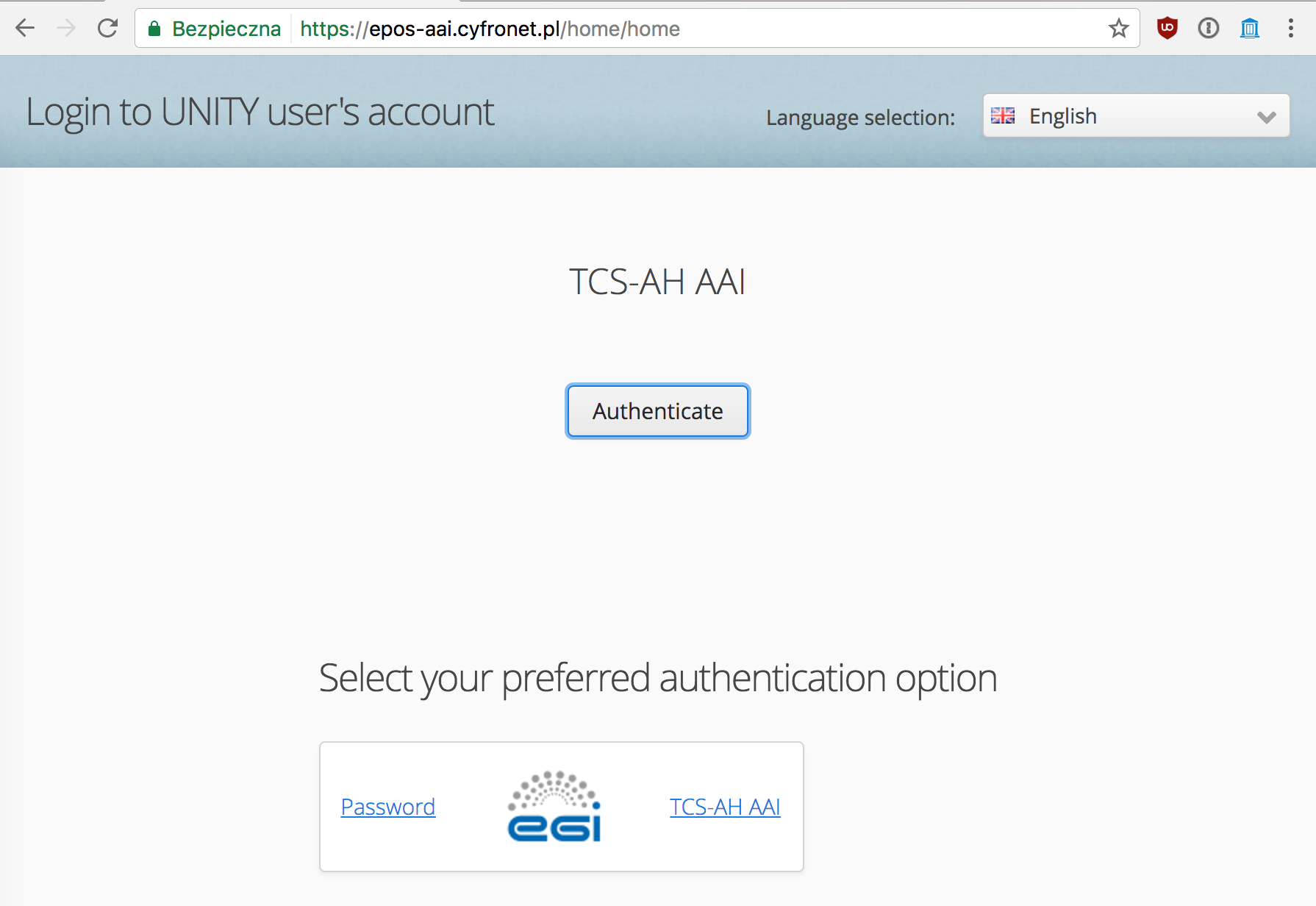
Authorisation to the specific EPOS resources is based entirely on tokens passing/translation (if needed). The same scheme will be utilised in case of EPOS ICS-D (distributed part of ICS, responsible for computational tasks including data analysis, visualisation, computing, etc.)



Picture : The pilot's main page



Picture : Authentication walk-through step via EGI Check-IN



Picture : Authentication walk-through via TCS AH path

### Feedback

The prototype has been 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 has been accepted by EPOS WP6 and WP7 teams as a general solution of EPOS AAAI. The first version, integrated with the EPOS ICS-C portal, is expected in a fall 2017.

## Future plans

The major planned step is integrating 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. Second stage will include integration of the remaining Thematic Core Services with the ICS-C platform. Final step will bind 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 be also provided ‘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.

# Earthquake simulation (MISFIT)

## Overview

|  |  |
| --- | --- |
| **Name** | EPOS Computational Earth Science – Seismology - VERCE Portal |
| **URL** | <http://portal.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** | *Describe the scenario in which the demonstrator/pilot was used.* |
| **Success criteria** | *What were/are the success criteria for the demonstrator/pilot? Did the demonstrator/pilot meet these criteria?* |
| **User Documentation** | <http://www.verce.eu/Training/UseVERCE.php> |
| **Technical Documentation** | [http://www.verce-project.eu](http://www.verce-project.eu/) (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 a 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.

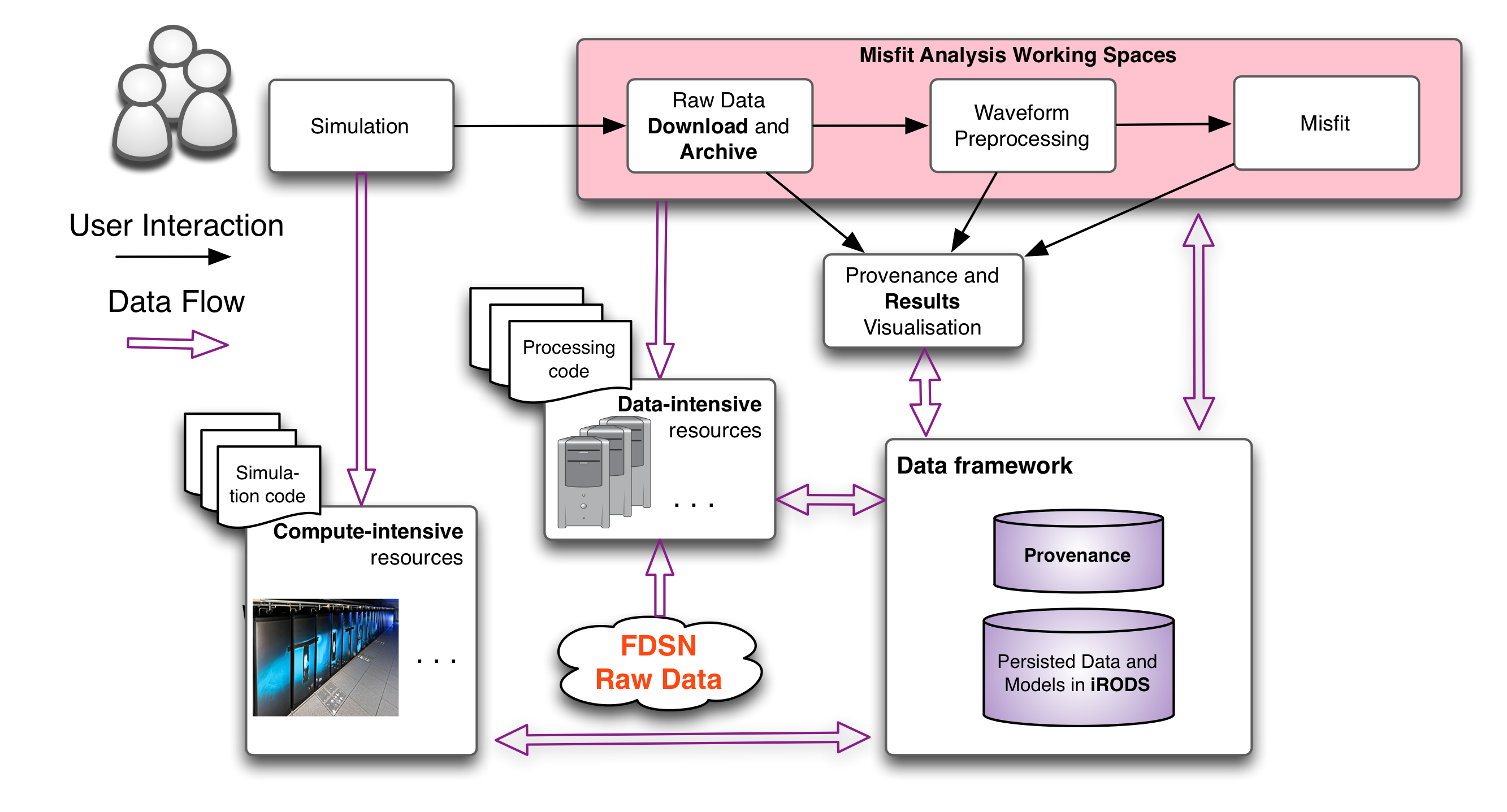


Figure . The original VERCE VRE architecture (without EGI Cloud connected)

### Integration and dependencies

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

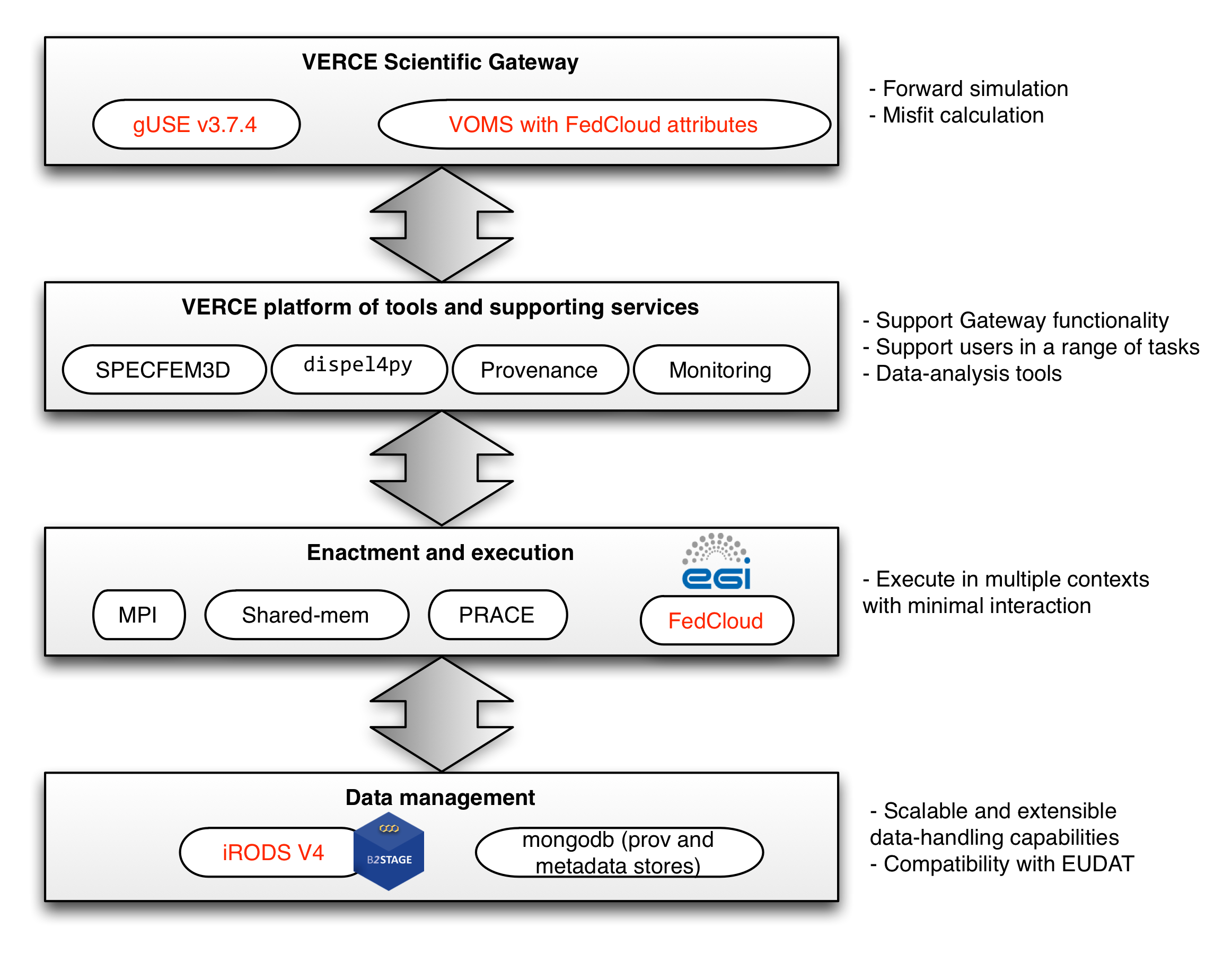


Figure 4. VERCE integration and dependencies: In red the updated components

**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[[9]](#footnote-10) 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 to the library, improving the overall performance of the analysisand 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 under “Data Management”. Its conceptual model is based on the W3C-PROV standard[[10]](#footnote-11), which is also offered as export format. The service, (S-ProvFlow: <https://github.com/KNMI/s-provenance>) includes a database, a WEB API and a visualisation 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 monitoring access The latter is obtained thanks to the provenance.

**Enactment and Execution**

These are architectures and computational services that the VERCE solution is able to make use of, thus adding to sustainability and standardization. We extended this portfolio with the possibility of integrating resources provided by the **EGI Federated Cloud.**

**Data Management**

iRODS[[11]](#footnote-12) is a rule-based data-management system order to address requirements to do with storage and access at scale, with configurable access policies. We have updated the current setup 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. MongoDB[[12]](#footnote-13) is a distributed and scalable document database used by the provenance and monitoring subsystems.

## 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 webservices, 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. Solutions to the issue will be tackled, trying to involving the partners delivering 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 remaining pending issue 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 more 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 same VERCE framework, with the scope of scaling all the generic components to a larger user-base and going towards the characterization of the EPOS Computational Earth-Science strategy. With the support of the EOSC, which will offer 1FTE effort of development for one year, the team will aim at offering a production deployment of the service for the EPOS community, as the immediate outcome of the pilot . It 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

## Overview

|  |  |
| --- | --- |
| **Name** | P-SBAS InSAR Sentinel-1 TOPS |
| **URL** | https://geohazards-tep.eo.esa.int/ |
| **Description** | The demonstrator connected Thematic exploitation platform with the EGI Federated Cloud to support a demonstrator use case, P-SBAS, by 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[[13]](#footnote-14), see the figure below.

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

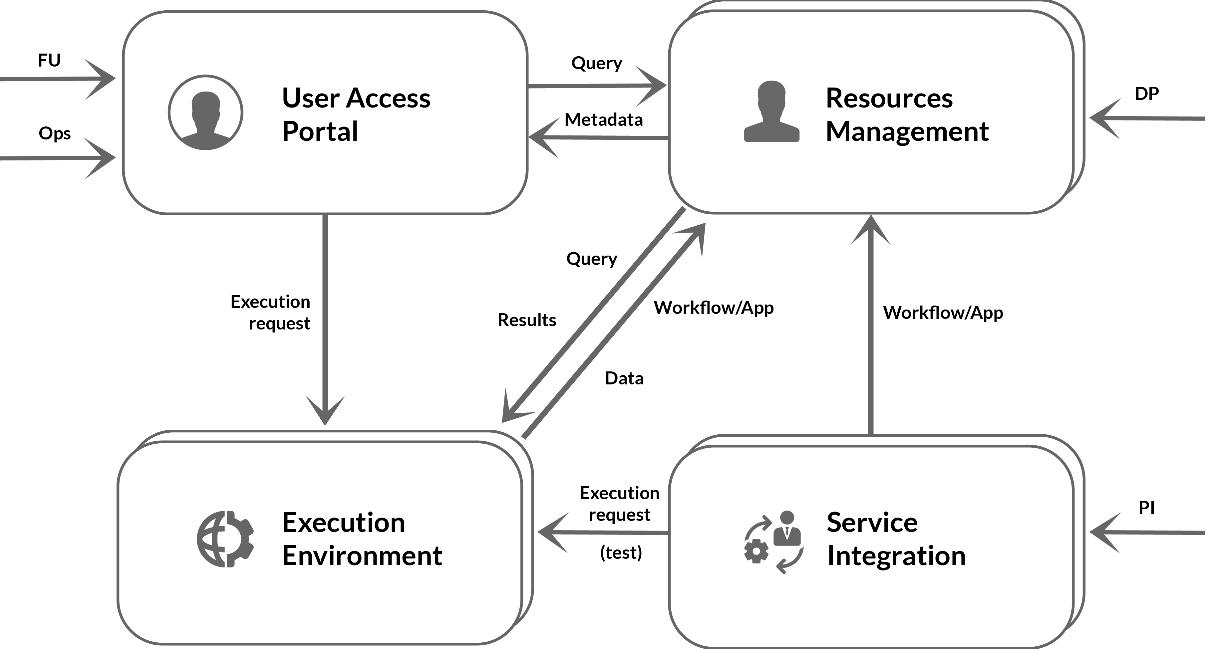


Figure 5. ESA Thematic Exploitation Platform architecture

The following image is an example of data visualised through the GeoBrowser of the Geohazard exploitation platform.

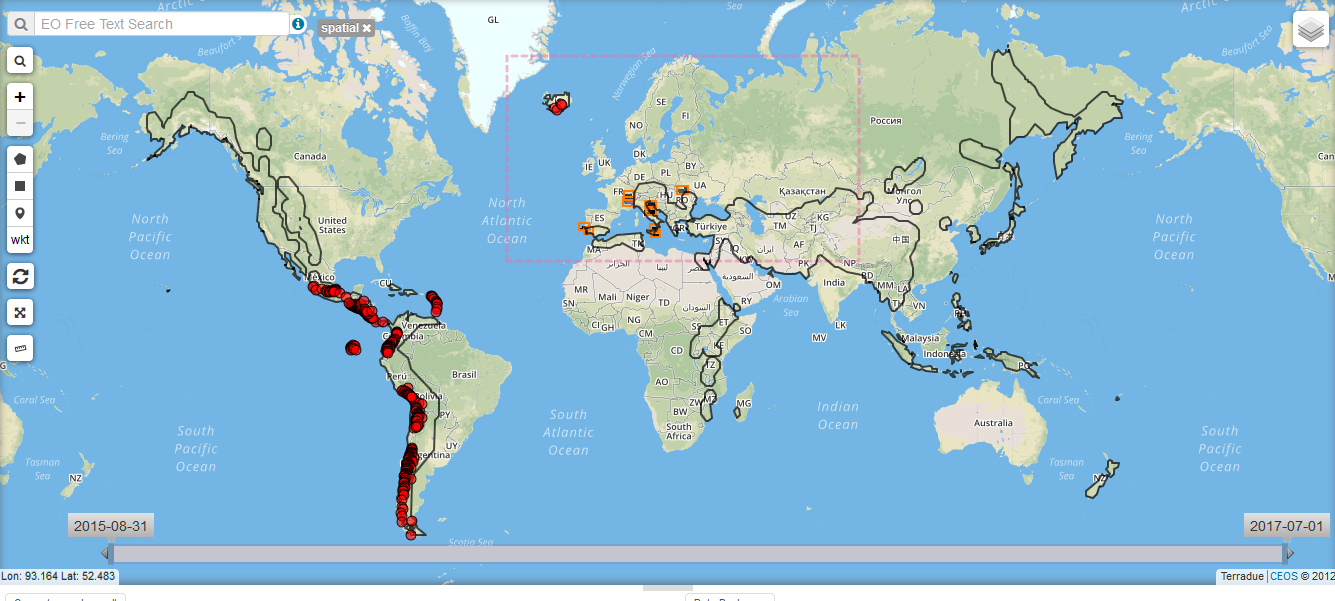


Figure . Seismic and volcanic areas showed in the GeoBrowser of the GEP.

## 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[[14]](#footnote-15) format with the following structure: Line Of Sight (LOS) Displacement time series, Mean LOS Velocity, Temporal Coherence, Average scatterer 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 advanced version, which should be available in the coming weeks. 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[[15]](#footnote-16) and EGI guarantees the availability of the cloud resources until the end of the 2017, and then the pilot could be executed, at least, until the end of the year. Terradue intends to request a renewal of such agreement so then the pilot could continue also in the 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 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

As for the Prototype of the EPOS Authorisation Authentication Infrastructure (AAI), its integration within the prototype with the EPOS ICS-C shall be done by fall of 2017. The work done can be also scaled and reused within the EPOS TCS community. With respect to this, one of the main lesson concerns the complexity of EPOS: involving the actual data providers is indeed an action that should be undertaken involving both the EPOS main coordination node, the ICS-C, and the Thematic Communities. The means to undertake such a challenge include the development of appropriate procurement policies that EPOS will be able to deal with after the ERIC is established (the ERIC process is planned to start in 2018).

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. It implies to find appropriate means to scale all the generic components to a larger user-base and going towards the characterization of the EPOS Computational Earth-Science strategy. A possible context where this action could be undertaken 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, as the immediate outcome of the pilot.

As for the Satellite Data use case, the demonstrator 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.

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 these 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 they 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, topical 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-2)
2. A current assumption is that compatibility between UNITY IDM and EGI CheckIn 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-3)
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-4)
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-5)
5. <https://www.epos-ip.org/> [↑](#footnote-ref-6)
6. <https://www.epos-ip.org/ride> [↑](#footnote-ref-7)
7. EGI Applications On Demand service: <http://access.egi.eu> [↑](#footnote-ref-8)
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-9)
9. <http://geodynamics.org/cig/software/specfem3d> [↑](#footnote-ref-10)
10. <http://www.w3.org/TR/prov-overview> [↑](#footnote-ref-11)
11. <http://irods.org> [↑](#footnote-ref-12)
12. <http://www.mongodb.org> [↑](#footnote-ref-13)
13. <http://go.egi.eu/EP-OpenArchitecture> [↑](#footnote-ref-14)
14. Comma Separated Values [↑](#footnote-ref-15)
15. VO SLA Terradue: https://documents.egi.eu/document/2763 [↑](#footnote-ref-16)