



**25-05-2018**

## **Deliverable DSA1.1: Results on Pilots with New Communities part 1**

### **Deliverable DSA1.1**

Contractual Date:	30-April-2018
Actual Date:	<b>Error! Reference source not found.</b>
Grant Agreement No.:	653965
Work Package:	SA1
Task Item:	1
Lead Partner:	GRNET
Document Code:	<DSA1.1>
Authors:	A. Terprtra (SURFnet), L. Florio (GÉANT), M. Reale (GARR), S. Visconti (Reti), C. Kanellopoulos (GÉANT), K. Koumantaros (GRNET)

### **Abstract**

This document provides a general overview of the goals and approach of the Pilots Service Activity 1 in AARC2. A detailed description including an outline of the use case and the results achieved to date is given for each of the nine Research Community pilots undertaken by SA1 Task 1 in Y1 of the project. The document concludes with some lessons learned so far.

© GÉANT on behalf of the AARC2 project.

The research leading to these results has received funding from the European Community's Horizon2020 Programme under Grant Agreement No. 730941 (AARC2).



# Table of Contents

Executive Summary	1
1 Introduction	3
2 Y1 results from Task1: Pilots with Research communities	4
2.1 Pilot approach	4
2.2 Overview of pilots carried out in Task 1	6
2.2.1 CORBEL (also known as Life Sciences pilot)	6
2.2.2 CTA	9
2.2.3 DARIAH	11
2.2.4 EPOS	13
2.2.5 EISCAT_3D	15
2.2.6 WLCG	18
2.2.7 LifeWatch	20
2.2.8 LIGO	21
2.2.9 HelixNebula / HNSciCloud	22
3 Conclusions	24

## Executive Summary

The goal of the Pilots Service Activity (SA1) is to enable e-Science communities, e-Infrastructures and research infrastructures to implement, deploy and operate AAI that follow the AARC Blueprint Architecture (AARC BPA) and that meets their needs. For interested research communities, SA1 also offers support for service delivery pilots, which help research communities design and choose an e-infrastructure provider that can deliver AAI services following the AARC BPA ‘as a service’ model.

At the start of the AARC2 project (April 2017), representatives of the pilot team conducted interviews with the research collaborations (selected during the preparation phase of AARC2) to review the use cases, and scope, design and plan the pilots. An additional use-case, DARIAH, emerged during Y1; this resulted in an additional pilot and it is also covered in this document.

The first section (Chapter 1) of this document provides an introduction and describes the methodology followed for the pilots in SA1 as whole. As indicated in the AARC2 description of work, the pilots are not meant to start from scratch, but they take as input existing implementations, tools and products and will only develop what is necessary to facilitate the integration of these components and to deliver an AARC blueprint-complaint AAI to support research collaborations. SA1 works very together with the research communities, in fact at least one representative of each research community is involved in the respective.

Chapter 2 reports on the nine pilots that started in SA1 that focus with the research collaborations (also referred to research collaborations). The list and the status of the pilots is depicted in Fig. 2.

The lesson learned so far are presented in the Conclusions (Chapter 3). The approach followed by the pilots meant however that AARC2 timelines could not always be met as planned. It was not possible, for instance, to stage the pilots to limit the number of active pilots in a given moment. This is due to the fact that research communities need to allocate their own resources, which may not always be available at the right time. The engagement of the research communities is a key aspect that also ensure that the resulting pilots are already in the hands of those that will be operating them.

In retrospective it would have been better to organise dedicated hands-on face to face meeting to discuss research communities’ requirements and possible technical solutions. This was addressed later during Y1 with a plug fest, which was very useful. A training on the key aspects of the AARC BPA and in particularly on the proxy, may have been useful to bring everybody at the same level; a more advanced training is planned for Y2 Q1. The pilots will continue and will be finalised during Y2.

The pilots so far show that the AARC BPA is scalable and can be deployed in ways to fit the research community’s needs. The pilots also show that the AARC BPA can be it can also be deployed and operated by one entity or by multiple entities each supporting different components.

In this document and in general in the context of AARC2, research collaborations, research infrastructures or research communities are used interchangeably to refer to communities (either in the form of real legal entity or a virtual collaboration) that undertake research in a specific field.

E-infrastructures henceforth refer to EGI, EUDAT, GEANT and PRACE, which are organisations that offer infrastructure services for the benefit of different research collaborations.

# 1 Introduction

Service Activity 1, Pilots (SA1) aims to demonstrate the feasibility of deploying Authentication and Authorisation Infrastructures (AAI) for research communities and e-infrastructures that fit the overarching AAI model defined by the AARC Blueprint Architecture (BPA). To this end, this activity demonstrates through (pre-)production pilots that:

- The AARC BPA and the related policy documents can be instantiated to fit research communities' requirements, and deployed and operated in production environments.
- Communities can choose one or more e-infrastructure providers that can deliver AAI services compliant with the AARC BPA, or operate the AAI themselves.
- User/group information can be retrieved from distributed group managements and attribute providers. This information in combination with the affiliation provided by the user Identity Provider is used for authorisation purposes.

The approach followed by the AARC2 project pilots differs from that used in AARC1. In AARC1, different technology-driven pilots were needed to test how selected components would meet the functional and technical integration requirements of research communities and e-infrastructures, whereas the focus in AARC2 is on implementing the AARC BPA in a (pre-)production environment. Thus, AARC2 pilots are driven by research community use cases and focused on deployment.

To achieve this, several research communities were brought into the project to work closely with the AARC BPA experts on developing and deploying their own AAI. During the preparation of the AARC2 project, research communities were invited to submit use cases they wanted to address in the AARC2 project. The response exceeded expectations, with more proposals than AARC2 could possibly manage. The selection process carried out resulted in the list of research community use cases that are currently included in the AARC2 description of work.

The AARC2 pilots are driven by three main use cases:

- Research infrastructures and/or e-infrastructures that need an AAI (that uses an IdP/SP proxy) to enable federated access to their (Web and non-Web) services. The AARC BPA fits these requirements; SA1 supports these communities to deploy their AAI in the most effective and interoperable way.
- Research communities that require access to services offered by different research infrastructures or e-Infrastructures and wish to use their existing credentials.
- Validating results from Joint Research Activity 1 (JRA1) and Networking Activity 3 (NA3) in a (pre-)production environment.

At the start of the AARC2-project (April 2017), representatives of the pilot team conducted interviews with the research collaborations to review the use cases, and scope, design and plan the pilots. The pilot approach followed is described in more detail in the next section. A pilot platform had been already built in the AARC1 project, which is also being used for the AARC2 pilots. As indicated in the AARC2 description of work, the pilots

are not meant to start from scratch, but they take as input existing implementations, tools and products and will only develop what is necessary to facilitate the integration of these components and to deliver an AARC blueprint-complaint AAI to support research collaborations. For most of the pilots, the sustainability model is already built in:

- The communities have an active role in building their own AAI (no hand-over is needed) with the support of the AARC2 team. In most cases, the communities also decide to operate the resulting AAI, and because the AAI is a critical element of service delivery, they will make effort and expertise available to ensure smooth operations.
- In one specific pilot, the LifeSciences AAI (LS AAI), the decision was made to outsource the AAI operations to e-infrastructures. In this case, EGI, EUDAT and GÉANT are working together within AARC to develop the LS AAI. Provisions are being made for the post-pilot phase that will end in February 2019.

To progress with the pilots, SA1 organised a plug fest (see this [blog post in 2018](#)), where the research collaborations' developers and the AARC2 pilot team could meet and discuss specific challenges.

## 2 Y1 Task 1 Results: Pilots with Research Communities

This section reports on the results achieved in SA1 Task 1 'Pilots with Research Communities' in the first year of the AARC2 project. Before the start of AARC2, eight research communities and related use cases were identified. During Y1, DARIAH, who was already participating in AARC2, provided a new and relevant use case, which resulted in an additional pilot being designed. The results of the DARIAH pilot are also included in this document.

For all nine pilots the goal is to implement an AAI that uses an IdP/SP proxy to support federated access and Single Sign On for different services the community needs; The AARC BPA provides a standardised approach on how to do that; it also provides policy and guidelines to ensure a smooth implementation.

### 2.1 Pilot Approach

It became quickly clear that it would not be possible to follow the staged approach (as described in the AARC2 Technical Annex) and have only a predefined number of pilots in execution. This is mainly due to the fact that the (pre-)production AAI development has dependencies with internal research collaboration resources which are outside of the AARC2 project. Rather than limiting the number of pilots for each cycle, it was decided to define four different phases for each pilot to go through (Figure 1) as described below.

#### 1. Analysis

- During this phase, the SA1 team conducted interviews with each research community to gather requirements which served as input for preparing detailed pilot proposals.

- The pilot proposal presented the AAI architecture and possible components to be used for its implementation to the research community. The result of this phase is an agreed pilot proposal with timelines.

## 2. Implementation

- Having drawn up an initial architecture and selected the technical components to be used, representatives of each user community started deploying these components and implementing the pilot infrastructure according to the proposed architecture.
- The technical implementation of the BPA takes place and an initial AAI is the result of this phase.

## 3. Evaluation

- In this phase, the basic AAI evolves and several production and/or test Service Providers as well as policies from NA3, if relevant, are introduced.
- This usually leads to new discussions, implementation and/or configuration.
- The (final) result of this phase is a (pre-)production AAI.

## 4. Finalisation

- During this phase, further testing is conducted and the results are packaged in such a way as to enable communities outside of AARC2 to understand the use case, the design of the AAI, and how it was implemented.
- The result of this phase is documentation, packaging and other (promotional) material.



Figure 1: Pilot approach

Figure 2 below shows that status of the pilots in Task 1 to date:

Community	Links	Topics/Focus	Status
		Connecting services & Brokering Leverage the work done by AARC on policies and architectural blueprints Implementing Sirtfi Using eduGAIN	CONCLUDED
		Cross infra use case integration with EGI/EUDAT/PRACE Controlled, granular access to resources. Need for a good LoA scheme for AuthZ	IMPLEMENTATION
		Cross infra use case integration with EGI/EUDAT/PRACE Delegated federated access (non-interactive) Workflows	IMPLEMENTATION
	CTA Cherenkov Telescope Array	Initial implementation of Community IdP/SP proxy, Group/Role based access to resources, SIRTFI and CoCo/GDPR compliance	IMPLEMENTATION
		Integration, access for citizen scientist	IMPLEMENTATION
	CORBEL LifeSciences AAI	Inter compatibility, share a common AAI shaping according to the ideas in Elixir. Also focus on sustainability and operational aspects    	TESTING
		Non webstuff (SAML-X509) Implementation of Sirtfi stuff Solution for a persistent unique ID (ORCID?)	ANALYSIS
		Non web scenarios + enrolment workflows	IMPLEMENTATION
		Implementing an AAI according BPA to allow communication between Dariah and other infrastructures	IMPLEMENTATION

Figure 2: Overview of pilots with research communities

## 2.2 Overview of Pilots in Task 1

### 2.2.1 CORBEL (also known as LifeScience pilot)

#### 2.2.1.1 Introduction

[CORBEL](#) is an initiative of 13 life science biological and medical research infrastructures (BMS RIs), which together create a platform for harmonised user access to biological and medical technologies, biological

samples and data services required by cutting-edge biomedical research. CORBEL communities are represented in the AARC2 project by BBMRI-ERIC, EMBL (and their third-party ELIXIR), INFRAFRONTIER and INSTRUCT.

In September 2017, CORBEL produced an [AAI requirements document](#) and published a Call for Proposals, urging the three e-infrastructures in AARC (EGI, EUDAT and GÉANT) to come together and design a solution for them based on the requirements contained in the document. The coordination of the pilot would take place under the AARC2 umbrella ensuring broader participation and benefiting from the technical expertise and guidance from the project.

EGI, EUDAT and GÉANT answered the call for proposals from life sciences research communities to deliver and operate a single AAI for them, as a more sustainable and cost-effective way to enable users' access to life science services. This pilot is particularly important for the following reasons:

- This is the first pilot in the AARC2 project in which a research community is not evaluating a technology in order to implement an in-house solution. This is a service delivery pilot in which life sciences communities evaluate a joint offering from the three e-infrastructure providers.
- This is the first time that different research infrastructures active in the same field have jointly agreed to run a single AAI to serve the entire life sciences community.
- It demonstrates that the AARC BPA can also be deployed in a multi-operator environment.

Given the advanced nature of the pilot, that involves integration of different AAI components operated by different parties, it runs formally as part of Task 3 (Piloting Advanced use-cases). It is however reported here as CORBEL is one of the use cases that are part of Task 1.

### **2.2.1.2 Pilot Description**

CORBEL requests the ability to authenticate and identify users and link them to a unique 'LifeSciences Identity' (LSID) to support management of the accounts and their attributes, as well as to enable these users to access both federated services within the LifeSciences community and generic services using several technological interfaces (SAML, OIDC, X.509, represented as bridges or token translation services).

The CORBEL LifeSciences Combined Pilot comprises different components based on the AARC BPA:

- North-bound proxy – run by GÉANT.
- SAML South-bound proxy – run by EGI.
- OIDC South-bound proxy – run by EUDAT.
- User management and attribute service – run by EGI.
- Token translation service – run by EUDAT.
- Step-Up Authentication – run by GÉANT.

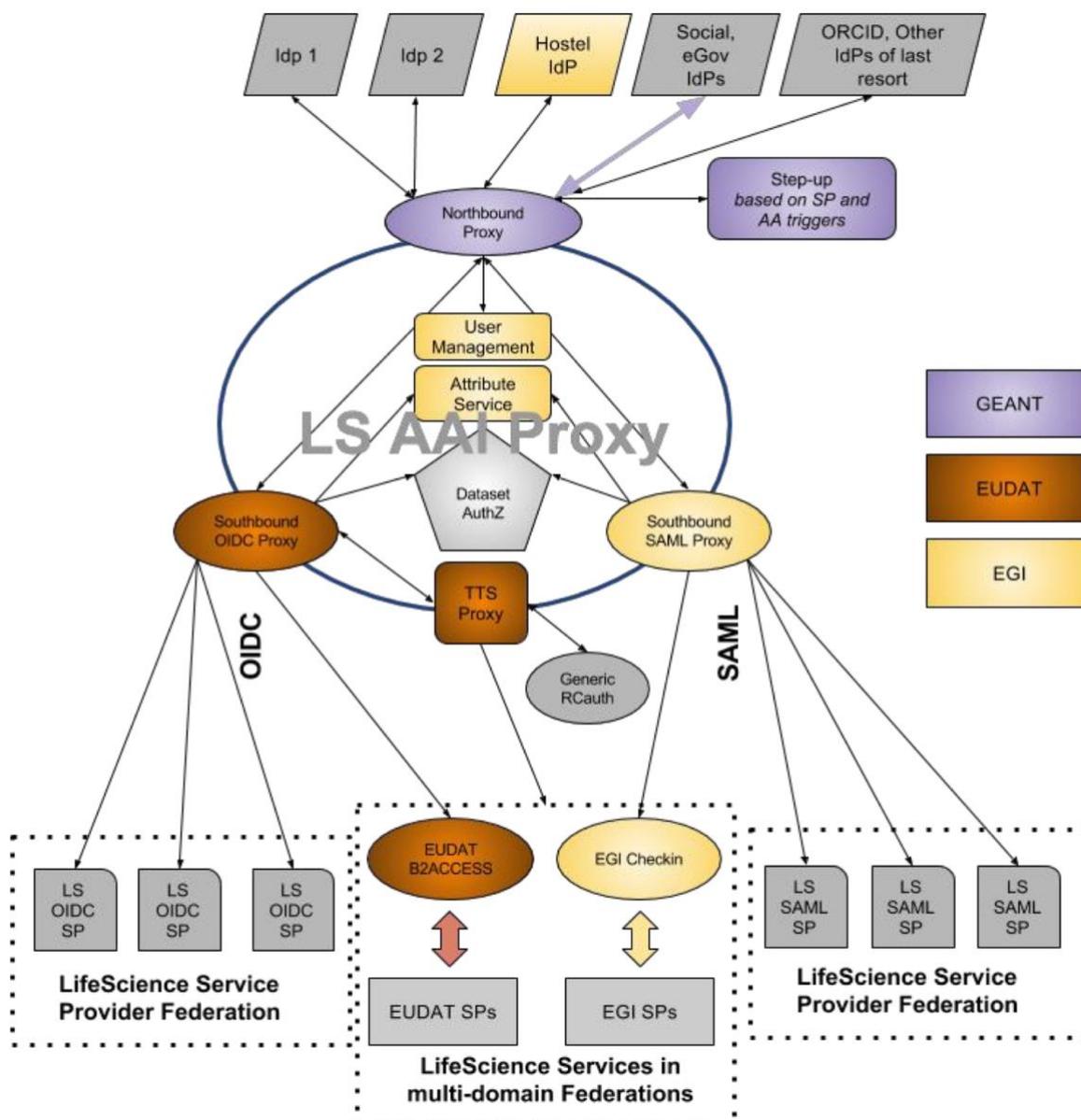


Figure 3: Architecture of LifeSciences pilot

The pilot started in December 2017, when the joint proposal from EGI, EUDAT and GÉANT was approved, and is being rolled out in three stages:

- *Phase 1* – December 2017 to end January 2018: concluded. Bootstrapped the AAI solution by putting proxy components together and defining the user registration process, attributes required by service providers and the authorisation flow.
- *Phase 2* – February to end May 2018: ongoing. During this phase the aim is to operate the dedicated AAI components and provide technical integration of identity providers and selected service providers. In particular, this phase improves security and trust and provides security incident response capabilities.
- *Phase 3* – June 2018 to January 2019: will consider all aspects of sustainability and drive the pilot to full production and operation.

After phase 3 of the pilot ends, the LifeScience AAI will move out of the AARC2 project. The Life Sciences community will contract an e-infrastructure (or a consortium of e-infrastructures) in 2019 for long-term AAI operations for the benefit of Life Sciences research communities.

## 2.2.2 CTA

### 2.2.2.1 Introduction

[The Cherenkov Telescope Array \(CTA\)](#) is a multinational, worldwide project to build a new generation ground-based gamma-ray instrument. Prior to the AARC2 project, CTA already had a partial AAI solution in place which did not address all their requirements. Therefore, this pilot provides a good example of how to address the needs of a community who already has an AAI in place, in this case a SAML stand-alone catch-all Identity Provider for all CTA users, integrated with a group management tool used for authorisation. The fact that the 'new' solution piloted in AARC must be compatible with CTA's current solution, in the sense that current users should be able to use CTA services in the same way they do now, increases the complexity of this pilot.

The CTA community requirements for the AAI are:

- Implement a user-friendly user enrolment flow.
- Enable access for users that login via eduGAIN as well as via the existing CTA Identity Provider.
- Link identities under administrator approval.
- Keep supporting Grouper as the main authorisation front end towards their services, but at the same enable account linking via COmanage.
- Enable access for guest identities through Social IDs (light requirement).
- Support OpenID Connect for services (light requirement).

The pilot carried out within AARC2 is aimed at expanding their existing AAI with a BPA-compliant solution and at allowing the CTA community to federate through eduGAIN. The core component of the new infrastructure is the an IdP/SP proxy (based on SATOSA). In addition, an external attribute authority (COmanage) has been plugged to the proxy to manage the user enrolment process and ensure injection of additional user authorisation attributes coming from Grouper. COmanage also allows for account linking whenever appropriate, requested by the users and granted by the administrator of the collaboration.

### 2.2.2.2 Pilot Description

The following categories of users were identified jointly with the CTA community:

#### 1. Existing users on the CTA catch-all Identity Provider.

Currently, all CTA users exist in the CTA LDAP and Grouper. These users need to be available to COmanage in order to provide the correct authorisation attributes to the proxy. This is achieved by using the LDAP as the source for COmanage identities.

**2. Brand new users for CTA (users in eduGAIN but not in the CTA catch-all IdP – not existing in both COmanage and Grouper).**

New users have to be registered through COmanage first before they are able to use CTA services. When approved by a CTA administrator, an identity will be created within the CTA LDAP.

**Users in CTA that can login via eduGAIN (but that have also a CTA identity)**

Users with both a CTA and an eduGAIN identity– can decide to link the two (COmanage offers support for this). By doing so, the user can log into CTA services using his/her institutional account instead of the CTA credentials.

For each of these categories of users the required workflow has been implemented to ensure access to CTA resources via the newly deployed AAI infrastructure.

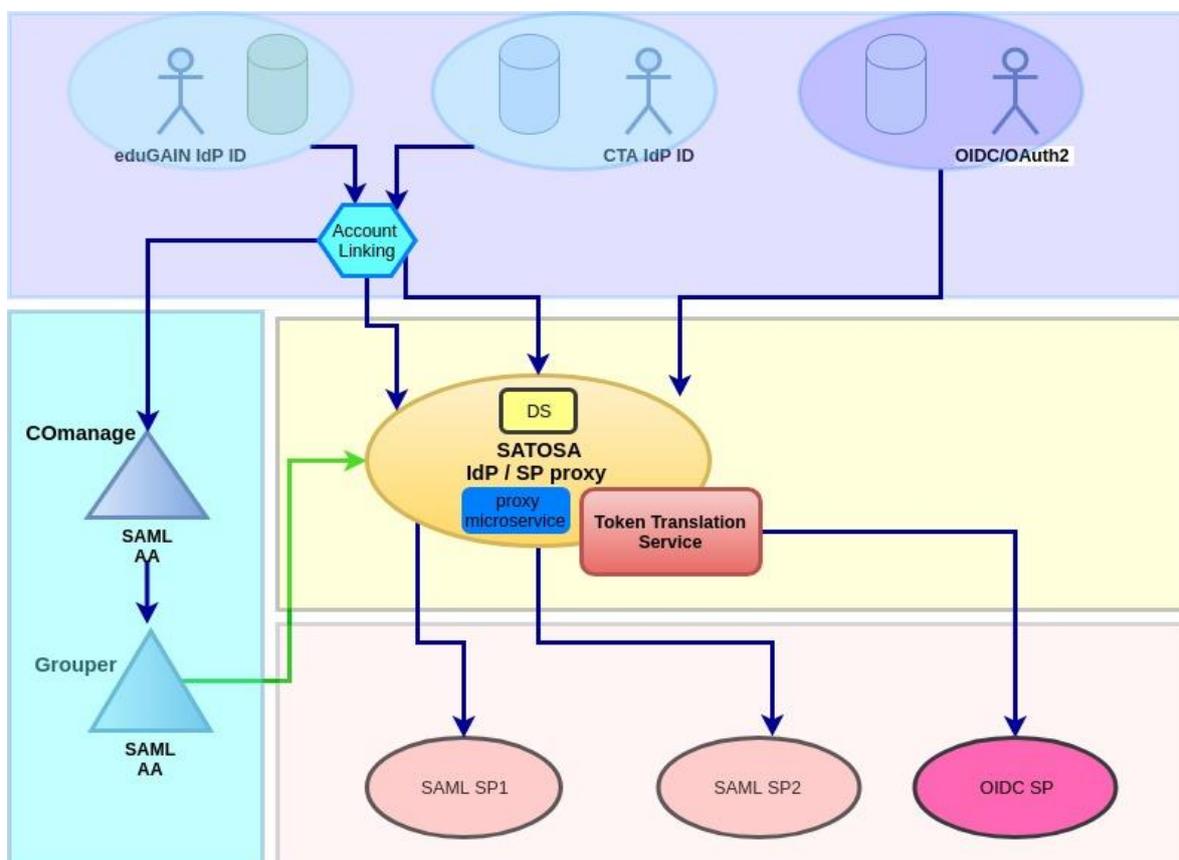


Figure 4: Architecture for the CTA pilot

The following main activities were carried out for this pilot:

- Defining clear user management processes to enable eduGAIN onboarding of all existing and new CTA users. This process defines the steps to interface and integrate COmanage and Grouper (their support was explicitly requested by the CTA community), the provisioning process of users inside COmanage, and account linking for users owning both (CTA local and eduGAIN) identities.

- Setting up all required infrastructural components required to implement the BPA-compliant architecture: CManage, SATOSA IdP/SP proxy.
- Piloting all foreseen steps for users to exploit all provider Authentication and Authorisation scenarios according to their needs.

The final goal of the pilot is to demonstrate the solutions designed and implemented to ensure full eduGAIN onboarding of the CTA user community. For this reason, the whole set of different flows foreseen for the users have been piloted and demonstrated.

## 2.2.3 DARIAH

### 2.2.3.1 Introduction

[DARIAH](#) is a European Research Infrastructure Consortium (ERIC) for arts and humanities scholars working with computational methods. It supports digital research as well as the teaching of digital research methods. It currently connects several hundreds of scholars and dozens of research facilities in the 17 European DARIAH member countries. In addition, DARIAH connects to several cooperating partner institutions in non-member countries and has strong ties to many research projects across Europe. Participants in DARIAH provide digital tools and share data as well as know-how. They organise learning opportunities for digital research methods, such as workshops and summer schools, and offer training materials for Digital Humanities. In AARC2, DARIAH is represented by DAASI International. The tasks of DAASI International in DARIAH include:

- Constructing and operating an AAI (Authentication and Authorisation Infrastructure).
- Integrating new technologies such as the management of virtual organisations via attribute aggregations.
- Developing a long-lasting and comprehensive operating unit.
- Drawing up a concept for and developing the DARIAH storage infrastructure.

### 2.2.3.2 Pilot Description

This AARC2 pilot is divided in two parts:

#### 1. Implementation of an SP-IdP proxy within the DARIAH AAI

According to the AARC Blueprint Architecture (BPA), communication between infrastructures should happen through dedicated infrastructure proxies. During this pilot, DARIAH implemented their own proxy solution based on Shibboleth. This proxy will be compliant to all relevant recommendations and guidelines developed within AARC, therefore this pilot can stand as a real-world example of the architecture work within AARC. As a side benefit DARIAH-internal services will also gain from this solution as much of the previously necessary complexity will be transferred away from the individual services to the central proxy component.

#### 2. Interoperability pilot between EGI and DARIAH

To showcase the successful implementation of the DARIAH SP-IdP proxy, the second part of this pilot deals with interoperability between the DARIAH research infrastructure and the EGI e-infrastructure. The goal is to allow

DARIAH users to transparently access EGI resources through EGI's own proxy solution (EGI CheckIn). As an initial use case, selected DARIAH users should be able to deploy and access virtual machines in the EGI infrastructure.

The following (slightly simplified) diagram shows the interaction between the various components of the DARIAH AAI (green), home organisation IdPs (yellow) and EGI (red):

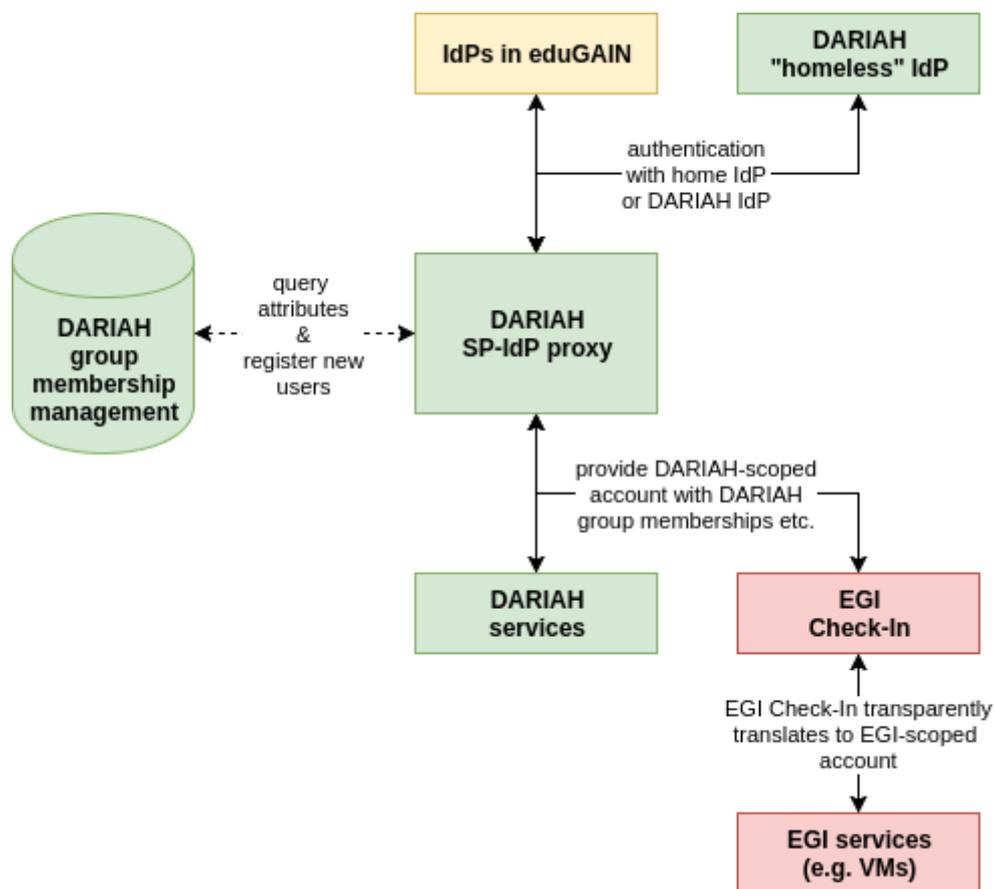


Figure 5: Architecture for the DARIAH pilot

The central component, which is being implemented in this pilot, is the DARIAH SP-IdP proxy. The proxy implements the AARC BPA and serves as an AAI gateway for two scenarios:

1. To allow users to authenticate at DARIAH services using their preferred authentication method (e.g. eduGAIN IdPs or the DARIAH homeless IdP).
2. In the inter-infrastructure use case the DARIAH proxy connects directly to the proxy of EGI (or other infrastructures in the future).

In both of these cases the user is assigned a DARIAH eduPersonUniqueId at the DARIAH proxy and the identity of the user is enriched with additional attributes (e.g. group memberships within DARIAH).

From a technical point of view the proxy is implemented using Shibboleth software. The Shibboleth IdP is the service-facing component of the proxy, while the Shibboleth SP deals with communication with the upstream IdPs. Since Shibboleth is not able to serve as a proxy out of the box, some additional glue is needed to connect the two components.

The other components involved are already being used in the current version of the DARIAH AAI (which is not using a central proxy) and must be slightly modified to work with the proxy. This includes the DARIAH homeless IdP (Shibboleth IdP), the DARIAH group membership management and self-service portal (LUI) and the DARIAH services, which are mostly based on Shibboleth SP.

Within the remaining timeframe of AARC2, the second use case (interoperability with EGI) will also be completed.

## 2.2.4 EPOS

### 2.2.4.1 Introduction

[EPOS](#) is a pan-European collaboration which aims to establish a comprehensive multidisciplinary research platform for the Earth sciences in Europe. It spans 25 countries, and encompasses 4 international organisations and 256 national research infrastructures. The expected number of users will likely grow to around 2000. EPOS already has established an AAI prototype with the EGI CheckIn service as an IdP and Unity-IdM as its core. The aim of this pilot is to vastly extend this prototype to meet the full EPOS requirements concerning AAI and achieve a more mature, production setup.

The ultimate goal for EPOS is to implement SSO for their users while accessing EPOS services: the so-called EPOS Thematic Core Services (Web-based services in specific Earth Science domains) and the Integrated Core Services (General, cross-domain computing and storage resources, user management, metadata catalogue). TCS and ICS are interconnected by an EPOS interoperability layer.

The abstraction, interoperability layer, will ensure interoperation between the Integrated Core Services and the Thematic Core Services (TCS). As an example, Cloud services can be provided at the ICS level, but in some cases need to be accessed by a TCS. National Research Infrastructures will contribute to the provisioning of TCS services.

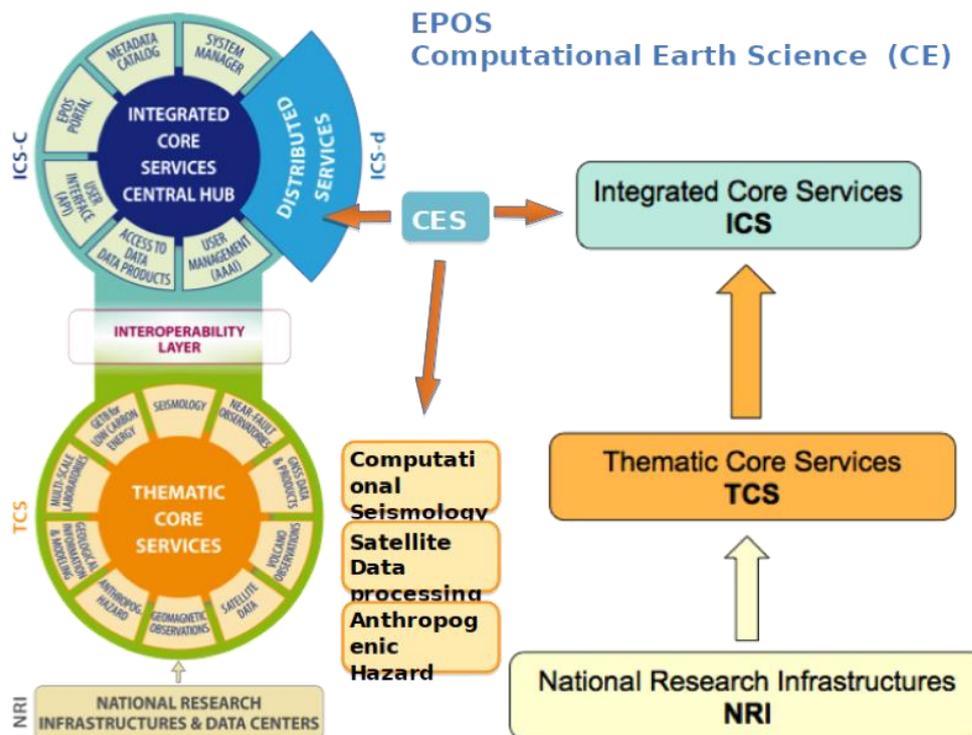


Figure 6: Schematic overview of EPOS

### 2.2.4.2 Pilot Description

The pilot will focus on the implementation of an architecture according to the AARC BPA, for which the central element is an IdP/SP proxy acting as a central AAI hub, based on Unity. Furthermore, Service Providers will be integrated using SAML or OIDC/OAuth2, and Identity Providers using X.509, SAML or OIDC.

Thematic Core Services (such as the domains: Anthropogenic Hazard and Computational Seismology) will be accessed via the central proxy service.

In addition, a core reference scenario for the pilot will consist in a user being able to access both EGI and EUDAT services by means of one account and one login process (SSO). Such a use case implies for example being able to retrieve seismological data on the EUDAT One Data service and being able to compute such data on the EGI infrastructure. This implies a token exchange for the user either internally, within the given science gateway, or while obtaining a token from One Data, using B2SAFE or B2STAGE EUDAT services to move the data where they can be accessed by an analysis program running on an EGI Federated Cloud Virtual Machine.

Overall, a possible simplified workflow related to the above-mentioned scenario is the following: Seismological Cloud Services will login via EPOS AAI and based on the users' credentials can get data stored in EUDAT and process them via EGI services.

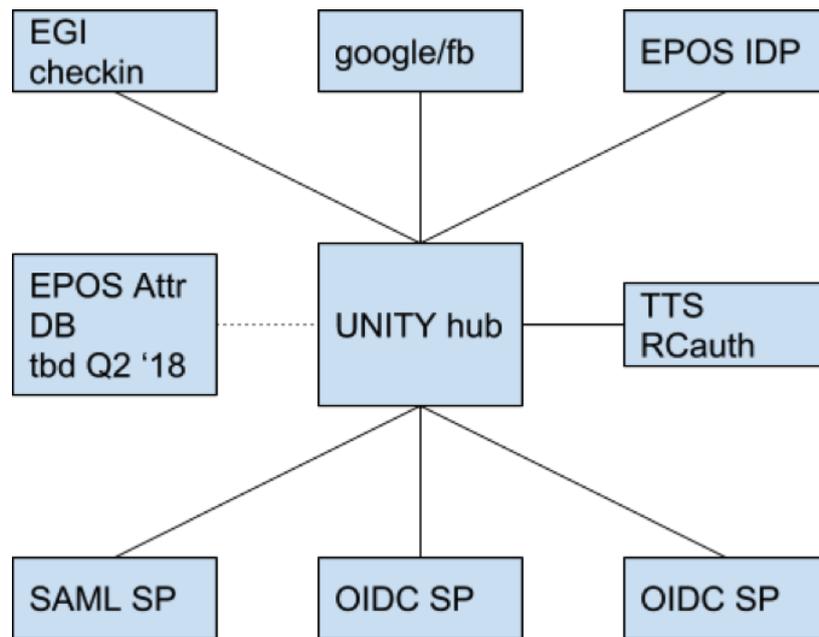


Figure 7: Architecture for EPOS pilot

A training session for EPOS was provided by AARC NA2 Task 2 and held on March 14, 2018 in Lisbon, entitled *Introduction to AAI concepts and federated access for the EPOS community*.

The current AARC AAI setup comprises several scattered services, non-usable by means of a Single Sign On procedure, and consists of:

- Unity IdP/SP Proxy.
- EPOS Attribute DB.
- LDAP.
- RAuth as a Token Translation System.

## 2.2.5 EISCAT\_3D

### 2.2.5.1 Introduction

[EISCAT 3D](#) is building an international research infrastructure that will use radar observations and the incoherent scatter technique for studies of the atmosphere and near-Earth space environment above the Fenno-Scandinavian Arctic, as well as support scientific research related to the exploration of the solar system and radio astronomy. This radar system is designed to investigate how the Earth's atmosphere interacts with space but will also have a wide range of other scientific applications for e.g. space weather forecasts and detecting space debris.

EISCAT\_3D's users are expected to access the User Analysis Facility through a user portal (Web) or a command-line interface to the virtualised resources. The metadata searches for analyses may also be performed through either the EISCAT\_3D portal or command line interface. The data to be analysed must be accessed from the fast and slow data stores of datacentres and transferred to the computing resources where the analysis code will run. As EISCAT\_3D users will access the computing e-infrastructure from different countries (including, it is expected, from outside the Nordic area), a common means of authenticating (identifying) users and authorising access is needed.

### 2.2.5.2 Pilot Description

EISCAT currently provides access to their resources to their partners through a web portal, using **IP addresses** and **country codes**. The current setup consists entirely of software components written in Python, including:

- CGI portal under Apache.
- Separate data download service, IP based.
- Processing services, IP based.
- Schedule request service, open.

The intended AARC AAI setup consists of:

- SATOSA IdP/SP Proxy.
- COmanage.
- EISCAT and eduGAIN Identity Providers.
- Plugins for Social Identity providers (OIDC/OAuth2).
- SAML to OIDC/OAuth2 TTS.

The main goal of the pilot is to have EISCAT\_3D move away from the IP-based Authentication model they are currently using for [their portal](#) and to embrace the federated access. An AARC BPA compliant AAI that will manage access to all EISCA\_3D resources will be deployed; this will ensue that all EISCAT\_3D users can authenticate using their federated credentials.

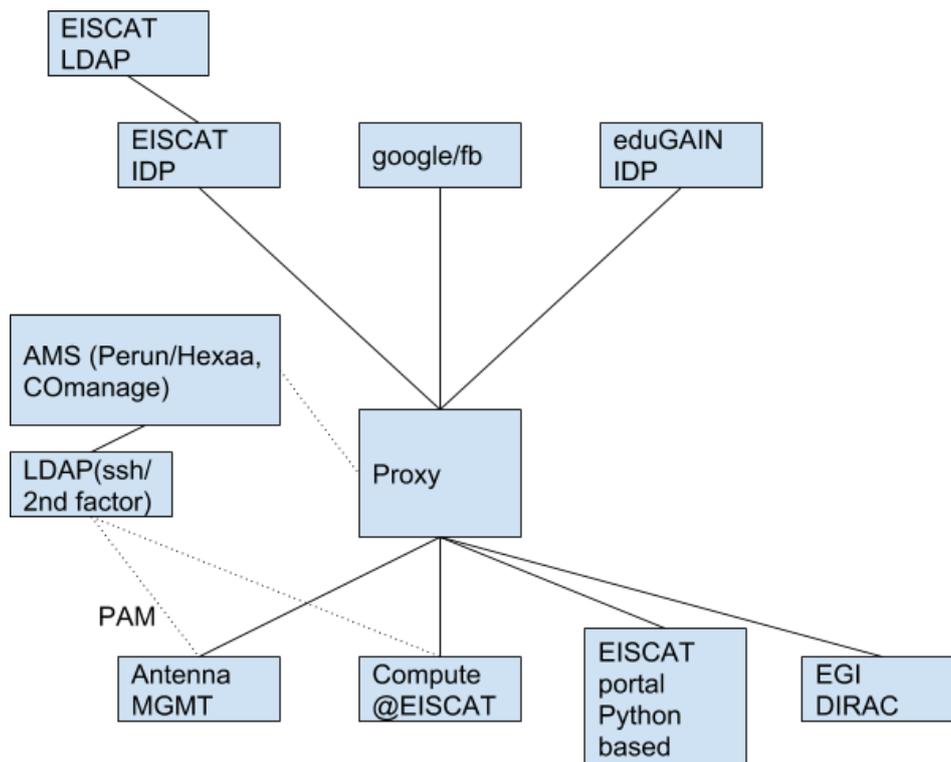


Figure 8: Architecture for EISCAT\_3D pilot

The main implementation phases designed for this pilot are the following:

1. Cloning of the current EISCAT\_3D production environment, including the main portal, to provide pilot/test instances for the AARC2 pilot.
2. Turning the main EISCAT\_3D data access portal to a SAML Service Provider (offer the functionality currently provided by the IP-based portal).
3. Defining key attributes to map users' roles to the new AAI.
4. Implementing a test Identity Provider for the EISCAT\_3D community for testing/piloting purposes, to be used jointly with the test Service Provider. Support for social identities will also be provided.
5. Setting up a CManage instance to enable account linking and group management.
6. Setting up a IdP/SP proxy (SATOSA has been chosen for this pilot).
7. Testing the new integrated AAI.
8. Planning and Implementing the migration of the current user community to the new AARC BPA AAI.

To support the implementation of the whole pilot in its various phases, a specific training for the EISCAT\_3D community will be organised in Y2 by AARC NA2, in close collaboration with SA1.

## 2.2.6 WLCG

### 2.2.6.1 Introduction

[The Worldwide LHC Computing Grid \(WLCG\)](#) project is a global collaboration of more than 170 computing centres in 42 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources to store, distribute and analyse the ~50 Petabytes of data expected annually, generated by the Large Hadron Collider (LHC) at CERN on the Franco-Swiss border.

The goal of the AARC2 WLCG pilot is to demonstrate the usage of WLCG Grid services by users through eduGAIN without the need to make use of personal X.509 certificates, which are the current method of authentication for WLCG users. Extensive discussions have been held within the WLCG Authorization Working Group to define the community requirements for the architecture of the future AAI infrastructure for the WLCG Grid. The AARC pilot on WLCG has been liaising with the WLCG AuthZ WG to tune the pilot around the main steps required to provide a demonstration of such a use case.

The main goal for this pilot can thus be summarised as follows: *“Demonstrate a pilot solution for a researcher without a certificate to register in a WLCG VO and access a grid service”*, while at the same time:

- Introducing the minimal required new components to allow a smooth user experience:
  - Central IdP/SP proxy;
  - Token Translation Service;
  - Attribute Authority.
- Managing authentication and authorisation to comply with WLCG requirements and standards:
  - HR-db integration;
  - Acceptable level of assurance in line with IGTF profiles.
- Minimising the number of new developments required by WLCG Services.

### Requirements

The WLCG community is currently operated by means of Virtual Organisations, which represent a community of users belonging to large, usually world-wide groups, with similar needs and requirements in terms of Job processing and data management. The current AAI makes use of the VOMS (Virtual Organisations Management System) server, which stores relevant user attributes to issue an X.509 temporary proxy certificate acting on behalf of the users to resources. The VOMS server adds to the X.509 proxy an extension which specifies the user's VO affiliation, and his/her role within the VO. Based on this extension, authorisation decisions are taken by the Grid infrastructure.

An overall assessment of the requirements by WLCG for a new SAML-based AAI infrastructure resulted in the following list of items:

- VO Membership Management:
  - Attributes – VO ID, ID of credential, Name, Email, Authorization;
  - Support multiple federated credentials & their linkage;
  - Active role selection;
  - Token management achievable by the standard user.

- Service Requirements:
  - Attributes – Authorisation plus traceability / Groups/Roles;
  - Ease of implementation;
  - Use standard approaches;
  - Token integrity and validity verifiable without connecting to the issuer;
  - For non-web, users should not have to manage identities in addition to their login session.
- General:
  - Support for fine-grained suspension;
  - Smooth transition from current X509-based to token-based AAI.

### 2.2.6.2 Pilot Description

An overall AARC BPA-compliant architecture has been proposed to satisfy the needs of WLCG AAI, which consist of a central proxy element:

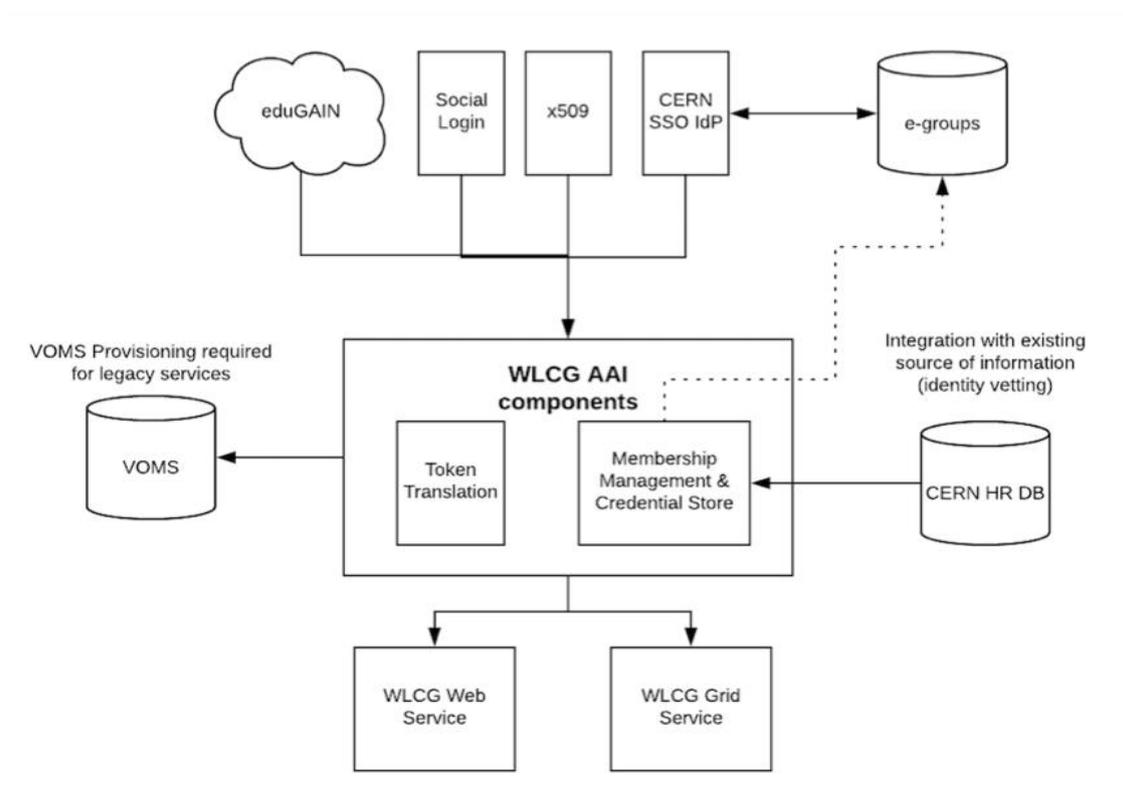


Figure 9: Architecture of WLCG pilot

The core component of the proposed architecture is the proxy element. The proxy will see CERN SSO as an IdP, and further plug additional Authentication Methods.

One implementation option for the IdP/SP proxy is the [Indigo IAM](#), which represents a complete solution with southbound OIDC, X.509 and SAML and northbound OIDC/OAuth2.

Another approach that will be evaluated within this pilot is the EGI CheckIn service, acting as an IdP/SP proxy, implemented in SimpleSAMLphp, and already integrated with the COmanage attribute authority.

The following main requirements (for both options) have been identified:

1. Integration with CERN SSO (as an option) and CERN's Identity Vetting DB;
2. Users should not have to request or manage X.509 certificates or other identity tokens themselves in addition to their login session (required token should be provisioned transparently), i.e. token translation, token provisioning;
3. Authorisation attribute selection by the user must be possible (i.e. active role selection);
4. Step-up for critical services e.g. 2FA.

## 2.2.7 LifeWatch

### 2.2.7.1 Introduction

[LifeWatch ERIC](#) is a key Research (e-)Infrastructure for EU-Latin America and the Caribbean Cooperation on Research Infrastructures that provides access to data (from different domains), analytical tools and computational facilities to support environmental research.

The purpose of this pilot is to demonstrate how a LifeWatch user can access a specific service, using the LifeWatch proxy based on AARC BPA. The solution deployed must be able to manage the different types of roles defined: Infrastructure Managers, Developers, Researchers and Citizen Scientists.

In order to cater for this list of different users, the system needs to support both roles and group management.

### 2.2.7.2 Pilot Description

The LifeWatch AAI will be used to:

- Give access to restricted LifeWatch services. The services may be restricted because of processing power or storage demands.
- Protect user data and scripts that are stored on the infrastructure (e.g. Unix home folders).
- Give access to data not yet in the public domain (data in databases, project moratorium period).
- Distinguish between users uploading data to the system (RvLab, eLab, data explorer).
- Give access to OpenStack configuration interface and computing resources at infrastructure layer.
- Manage roles/groups and authorise them to access specific services.

At least two components have been identified to be included in the AAI infrastructure: a proxy (one or more components depending on the solution selected, to manage groups/roles, authorisation) and a Token Translation System to allow access to non-web services.

The proxy component needs to satisfy the following requirements:

- Federation of 1-N institutions.
- Support Citizen Scientists via Social IDs).
- OpenID Connect for LifeWatch services (priority); SAML for LifeWatch services (optional).
- Roles Management. Role mapping (e.g. Google users to Citizen Scientist).
- Group Management.
- Identity linking (optional).
- Distributed, clustered; High availability; Via Database.

INDIGO IAM has been tested for supporting this, but there are some limitations in IdP federation.

The intended AARC AAI setup consists of:

- Proxy based on Keycloak or a different solution satisfying the requirements.
- WaTTS – configured to link to HPC resources.

The current pilot setup consists of:

- Keycloak instance – federates IFCA SSO, Google (for citizen scientists). Ongoing – eduGAIN, VLIZ.
- WaTTS – deployed but not yet configured/tested.

## 2.2.8 LIGO

### 2.2.8.1 Introduction

[The LIGO Scientific Collaboration](#), along with the Virgo collaboration, analyses data generated by the LIGO, Virgo and Geo600 detectors to search for gravitational waves. It enables collaboration between approximately 1000 researchers in 20 countries. Data, documents, and other resources need to be shared amongst many different combinations of groups of researchers. LIGO uses SAML and X.509 certificate technologies with a single internal Identity Provider.

This pilot will introduce an AAI architecture according to the AARC BPA. The pilot will also be used to motivate several service providers to make use of the AAI and therefore federated identities. SATOSA will be used for the SAML proxy and PyFF for metadata aggregation and discovery.

### 2.2.8.2 Pilot Description

The project is deploying an IdP/SP proxy based on SATOSA. The IdP/SP proxy will be registered in the LIGO metadata and pushed up to eduGAIN. A number of SPs will be connected to this proxy.

## 2.2.9 HelixNebula / HNSciCloud

### 2.2.9.1 Introduction

[HNSciCloud](#) is a Pre-Commercial Procurement (PCP) tender for the establishment of a European hybrid cloud platform to support the deployment of high-performance computing and big-data capabilities for scientific research. The project is building a public-private partnership between public research actors and cloud service providers.

As part of the procurement process, the HelixNebula consortium worked with cloud service providers to make their services available through eduGAIN via federated access.

The Helix Nebula use case was added to the list of potential pilots in AARC2 mostly to provide support for AAI aspects. The AARC2 team was involved in the design phase to present options on how to enable federated access to the procured services.

It was not possible to engage in an active pilot with this community due to the fact that the service delivery model was not completely defined at the time and development work was to be performed by the commercial cloud providers themselves rather than the consortium. The participation of the research collaboration in the development is a prerequisite for a pilot in AARC2 and the project does not have sufficient manpower to deliver a solution on demand.

#### Service Providers Needs

Federated access with SAML was a requirement; service providers were therefore asked to implement a SAML SP to connect to a federation via which to gain exposure to eduGAIN, and in parallel become a relying party of ELIXIR. This is already a challenging task for many service providers.

Some SPs decided to base authorisation decisions not only on affiliation (which is generally provided by the SPs) but also on group information. This information is not available in eduGAIN, but has to be retrieved by a group management system. Consequently, a group management solution had to be chosen from one of the well-established offerings available.

The choice of some of the service providers to rely on entitlements for authorisation added an extra layer of complexity, as such an entitlement has to be populated by at least the identity providers that participated in the pilot, which is not a trivial process and does not support the long-term aim for integration with existing identity federation practices.

Because each federation has its own documentation, and entities (i.e. IdPs and SPs) cannot join eduGAIN directly, the service providers were confronted with multiple sets of documentation, whereas they would have expected to find all the necessary information in a central point. There was a general mismatch of expectations regarding the level of operational support (including a help desk) eduGAIN would provide versus how much is provided in the eduGAIN model via federations. It would have been useful to include a support person from eduGAIN (or from one of the participating federations) to help during the integration phase and to act as one single point of contact. A testing environment would have been extremely beneficial for troubleshooting.

Although there are multiple proxy implementations and service offerings (e.g. EGI CheckIn, GÉANT eduTEAMS, EUDAT B2ACCESS) for different communities, there is no clear understanding of the added value of each, or how sustainable the service offerings are. Of the two commercial cloud services selected one opted for developing/maintaining their own proxy service to ensure long-term sustainability, while the other chose to bypass a proxy deployment. A publicly available IdP/SP proxy that could have been leveraged by the Helix Nebula consortium would have greatly facilitated the process by:

- The services would have had a central point, the IdP/SP proxy, to connect to, rather than trying to get into a federation or doing bilateral testing with ad-hoc IdPs.
- The proxy would be operated in a manner that complies with best practices and minimises problems with authentication and attribute flow.
- The IdP/SP proxy would have offered the same interface and procedures as well as centralise support and documentation.
- The proxy would connect to eduGAIN as an SP.
- Authentication would be managed via attributes provided via eduGAIN in combination with information provided via a group management.

Non-web use cases are critical for several scientific research and e-infrastructures. Although the SAML ECP (Enhanced Client or Proxy) profile is available it is not always supported by IdPs, which makes federated access to non-web resources still challenging. If services in a research/e-infrastructures are managed via an AAI that uses an IdP/SP proxy (as proposed by the AARC BPA), then a token translation service is generally supported; the latter enables federated access for non-web applications. However, not all research infrastructures have at the moment such a sophisticated AAI in place. In the case of Helix Nebula, there was no pre-existing community that could offer such an AAI to enable access to the services at large; Elixir AAI was successfully used, but it only serves Elixir users. The objective of the procurement, however, is that any interested research collaboration should be able to purchase and access the selected commercial cloud services.

### 2.2.9.2 *Pilot Description*

Two commercial providers developed a prototype to provide commercial cloud services to institutions, leveraging SAML authentication. One provider chose to deploy a proxy, the other did not. In the absence of a proxy, the approach followed by the consortium was to connect various services to different federations via which they would be made available in eduGAIN. However, different federations follow distinct practices, so that the experience varied for each service provider. In the process of testing configurations with selected IdPs, SPs discovered that the various IdPs may release different attributes, despite policy frameworks (REFEDS R&S entity categories) being in place to streamline the process.

The selected providers have now implemented their own solutions which they will continue to use. There are several recommendations to take away from this exercise:

- The service delivery model should be defined during if not before the procurement phase.
- For similar use cases (services procured for multiple international research communities), a platform (typically an IdP/SP service + group management) should be made available to support connection to eduGAIN.
- Effort to deploy the platform should be allocated.

### 3 Conclusions

During Y1 of AARC2, a total of nine research community pilots have started and are well under way in SA1. Each of the communities involved shares the goal of providing an AAI for their users and services according to the AARC BPA. Thus far, one of these communities has explicitly requested to have that AAI run by the e-infrastructure providers.

The pilots will be completed during AARC2 Y2. SA1 will close the pilot cycle by gathering the feedback received from members of the community and lessons learned, and producing the relevant manuals, and hand over the results of successful pilots to NA1 to work on sustainability aspects as needed. It is important to note that sustainability aspects were already considered in the design phase and, in light of AARC2 strategy, operation of the services always rests in the hands of the research infrastructures and e-infrastructures.

So far, SA1 has demonstrated that the AARC BPA is a very good fit for the community use cases. Additionally, some lessons learned were:

- In AARC2, at least one representative of each research community is involved from the beginning. This is crucial for the pilots to be successful since they are based on actual use cases and current (production) infrastructure. Moreover, from the point of view of the communities, participating actively in the AARC2 project and other tasks ensures an up-to-date view of developments within AARC2.
- In retrospect, conducting interviews through VC with representatives of research communities in order to be able to provide detailed pilot proposals took longer than anticipated. It turned out that, specifically for this phase, face-to-face meetings were much more efficient, due to the possibility to interactively draw examples and discuss. Moreover, research communities find the consultancy provided by the AARC2 AAI experts to be highly valuable, in the sense that existing infrastructures and use cases could be translated to specific AAI proposals.
- Despite the plans made, it was not possible to stage the pilots to limit the number of active pilots in a given moment. This is due to the fact that research communities need to allocate their own resources, which may not always be available at the right time, to take part.
- The scope of the CORBEL pilot was initially much smaller than what it is currently. This required more effort on the part of the AARC2 team than planned.

## Glossary

AAI	Authentication and Authorisation Infrastructure
AARC/AARC2	Authentication and Authorisation for Research and Collaboration
BPA	Blueprint Architecture
eduGAIN	education Global Authentication INfrastructure
eID	Electronic identification
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
IdP	Identity Provider
IP	Internet Protocol
LSID	LifeSciences Identity
OIDC	OpenID Connect
R&E	Research and Education
REFEDS	Research and Education FEDerations group
SA1	Service Activity 1 in AARC2
SAML	Security Assertion Markup Language
SATOSA	IdP/SP proxy
SP	Service Provider