20-09-2018

Deliverable DJRA1.1: Use-Cases for Interoperable Cross-Infrastructure AAI

Deliverable DJRA1.1

Contractual Date:	28-02-2018
Actual Date:	220-09-2018
Grant Agreement No.:	653965
Work Package:	JRA1
Task Item:	1.1
Lead Partner:	EGI
Document Code:	DJRA1.1
Authors:	Diego Scardaci (EGI Foundation), Licia Florio (GÉANT), David Hübner (DAASI), Michał Jankowski (PSNC), Jens Jensen (STFC), Christos Kanellopoulos (GÉANT), Daniel Kouril (CESNET), Nicolas Liampotis (GRNET), Mikael Linden (CSC), Shiraz Memon (Jülich), Mischa Salle (Nikhef), Arnout Terpstra (Surfnet)

Abstract

The researchers' need to access online services and resources offered by different research and e-infrastructures has increased over the last years. Through federated access, researchers should be able to seamlessly and securely access resources across these infrastructures using their existing credentials from their home organisations. AAI interoperability a key requirement to support this. The AARC blueprint architecture has been designed to address this need, aiming to improve the user experience when accessing and sharing resources provided by different infrastructures. To this end, this document analyses research community use cases that require access to services and resources across infrastructures. The research community specific use cases have been mapped to a set of generic use cases of cross-infrastructure AAI flows. These flows will serve as input for further refining and complementing where needed the AAI interoperability aspects of the AARC Blueprint Architecture.

Table of Contents

Table of Figures

Executive Summary

Nowadays, researchers need to leverage services and resources distributed across multiple service providers that are part of different administrative domains in order to collaborate in an effective way. These services are often provided by research infrastructures and general-purpose e-infrastructures, each implementing their own AAI.

Researchers should be able to access, exploit and combine all these services in a seamless way. Users should login once and, then, benefit from all the available features without worrying about the peculiarities of the infrastructure that is providing a specific service. This requires cross-infrastructure, interoperable AAIs that enable the transfer of authentication and authorisation information about the users in a transparent way.

This document provides an analysis of research community use cases that deal with access to services provided by more than one infrastructure. Use cases were identified from various types of research communities, including large collaborations (e.g. EPOS, ELIXIR), small research groups and single researchers (the so-called Long Tail of Science). Based on the analysis results, generic use cases have been extracted. These generic use cases represent common flows that require cross-infrastructure AAI interoperability. These flows will serve as input for the evolution of the AARC Blueprint Architecture aiming to enable researchers to seamlessly access and share federated services and resources offered by different infrastructure providers using existing credentials from their home organisations.

1 Introduction

An increasing number of researchers requires to use online collaboration tools and scientific applications and to be able to access distributed resources provided by multiple service providers that belong to different administrative domains. These use cases include, for example, accessing storage services and computing services provided by different general-purpose e-infrastructures in workflows orchestrated by platforms provided by a research infrastructure.

The AARC Blueprint Architecture has been proposed to address these use cases by allowing researchers to access, exploit and combine these distributed resources in a seamless way using their existing credentials, typically, from their Home Organisations. This federated access model relies on interoperable AAIs that allow for the transfer of authentication and authorisation information about the users across infrastructures.

This document elaborates further on community use cases requiring cross research and e-infrastructures access to services/resources. The remainder of the document is structured as follows: Chapter 2 provides an analysis of community-specific use-cases. It should be noted that this document covers only community use-cases that put the focus on the use of services provided by multiple infrastructures, thus highlighting the requirements for cross-infrastructure interoperability. Therefore, only a subset of the communities participating in the AARC2 pilot activity has been included. The analysis of the community-specific use-cases lead to the extraction of generic/recurrent use cases for interoperable cross-infrastructure AAIs that are described in Chapter 3. Lastly, in Chapter 4, conclusions are drawn. An example cross-infrastructure interoperability use case involving non-BPA compliant AAIs is presented in Appendix A, while use cases related to communities not participating in AARC2 have been included in Appendix B.

2 Analysis of community use cases

This Chapter describes use-cases of cross-infrastructure federated access to services and resources. These use cases involve users accessing and sharing resources from a research infrastructure and an e-infrastructure, as well as more complex workflows that require delegated [AARC-G005] access to resources provided by multiple e-infrastructures. These use cases are derived by the respective AARC pilots.

2.1 DARIAH

2.1.1 Introduction

DARIAH is a pan-European infrastructure for arts and humanities scholars working with computational methods. It connects several hundreds of scholars and dozens of research facilities across Europe. DARIAH runs its own AAI, which was initially built to give researchers in the digital humanities access to services provided by DARIAH. Researchers whose home organisations operate SAML-based IdPs connected to their national federations, are encouraged to use them to authenticate to DARIAH AAI. To support users that do not have federated accounts, DARIAH provides an IdP of last resort.

2.1.2 In focus for the AARC pilot

The goal of the AARC pilot is to allow users of the DARIAH infrastructure to access resources/services provided by EGI. The main focus is on enabling DARIAH users to run virtual machines in the EGI infrastructure. In the future, the outcome of the pilot could be re-used to enable DARIAH users to exploit other EGI services.

As part of the AARC pilot, the DARIAH AAI is extended by introducing an SP-IdP-Proxy according to the AARC blueprint architecture. Researchers who authenticate via the DARIAH SP-IdP-Proxy are then assigned a unique DARIAH ID (eduPersonUniqueID) and group membership information, that is managed in the LDAP backend operated within the DARIAH AAI. The implementation of this proxy component is part of the AARC pilot and serves two purposes in a) simplifying and extending internal DARIAH use-cases (e.g. connecting new services to the DARIAH AAI, checking AUP and terms of use centrally, managing user registration and dealing with authorization) and b) allowing interoperability with other infrastructures according to the guidelines developed within AARC.

From a technical perspective, DARIAH users need to access EGI services via the EGI Check-in SP-Proxy using their community identity provided by the DARIAH AAI. Since authorisation to EGI services is based on Virtual Organisation (VO)/group membership information, a translation of the group information from the DARIAH AAI will be required. Specifically, DARIAH will define a set of groups that will then allow their member users to get appropriate entitlements within EGI, e.g. for deploying virtual machines via the EGI Applications Database (AppDB) VMOps dashboard.

Another aspect to consider is the level of assurance required for obtaining access to EGI resources. DARIAH will express assurance information according to the AARC guidelines [AARC-G021] and ensure that appropriate processes are followed for meeting the assurance requirements.

The technical complexity in a) redirecting a user through various proxy elements and b) mapping user and group membership information between the two infrastructures should be opaque to the user.

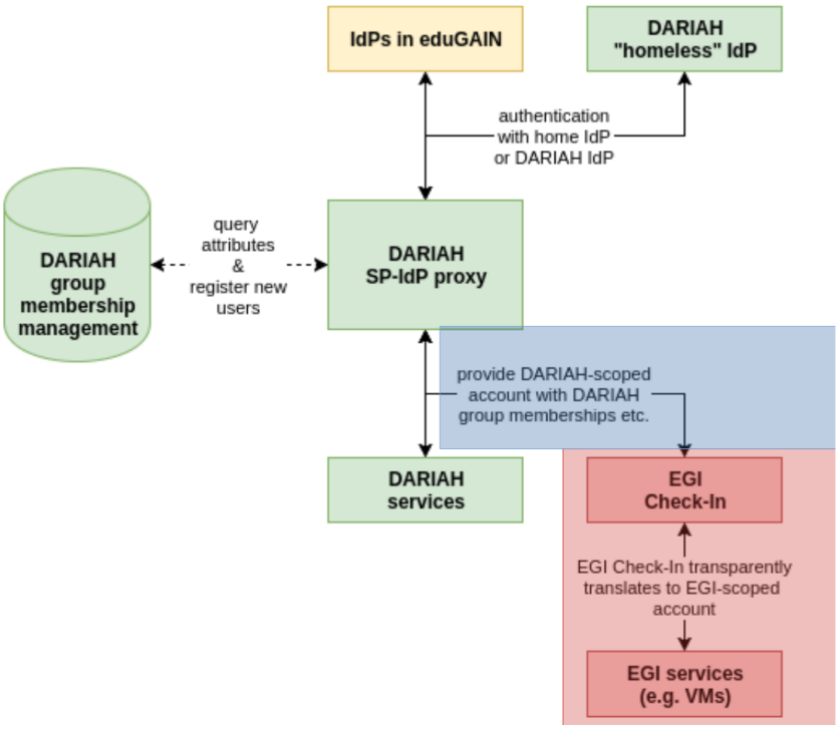

Figure 1. Architecture of the DARIAH AARC pilot.

2.1.3 Requirements

The more general use-case that can be deduced from the DARIAH AARC pilot is about research infrastructures users that need to access e-infrastructures services. For this particular case, key features that need to be enabled to allow DARIAH users to exploit EGI services are:

- When users try to access a service provided by an e-infrastructure (in this case EGI), the service will redirect the users to a Discovery Service, operated by the e-infrastructure, where they can find and select their own research infrastructure (in this case DARIAH)
- Users management is done within the research infrastructure, where users are assigned to groups. This information (user attributes) is translated within the e-infrastructure (EGI in this case) appropriately and is used by the e-infrastructure for authorisation purposes; a mapping between the attributes provided by the research infrastructure (DARIAH in this case) and the e-infrastructure that offers the service (EGI) is needed. The translation of the attributes should be transparent for the user.
- Access to e-infrastructure resources may require successful authentication with a high Level of Assurance (LoA). A procedure to allow users to increase the level of assurance needs to be in place.

2.2 CORBEL

2.2.1 Introduction

CORBEL is an initiative of 13 new biological and medical sciences research infrastructures (BMS RIs), which together will create a platform for harmonised user access to biological and medical technologies, biological samples and data services required by cutting-edge biomedical research. CORBEL and AARC are driving the design of the Life Science (LS) AAI, a common Authentication and Authorisation service portfolio for all the research infrastructures participating in the CORBEL project.

In 2016-2017, the Life Science research infrastructures, represented by the CORBEL project, gathered their AAI use cases and developed a requirements specification on a common Life Science AAI. Through the LS AAI users should be able to access federated services both within the LS community as well as generic services, using different technological interfaces (SAML, OIDC, X.509). These generic services include e-infrastructure services, e.g. those provided by EGI and EUDAT. In this case, the specific e-infrastructure SP Proxies are considered relying parties (service providers) in the LS AAI.

2.2.2 In focus for the AARC pilot

The AARC pilot is testing the deployment of a new AAI based on the AARC blueprint architecture in order to support the entire Life Science community. Although the main focus of the pilot is the delivery of the LS AAI from a consortium of e-Infrastructure providers (EGI, EUDAT and GÉANT), this pilot also highlights the interoperability aspects for cross-infrastructure AAIs as users registered in the new LS AAI, should be able to access also e-infrastructure services. In particular, for this pilot the EGI AppDB and the EUDAT B2SHARE services are used in order to investigate these challenges.

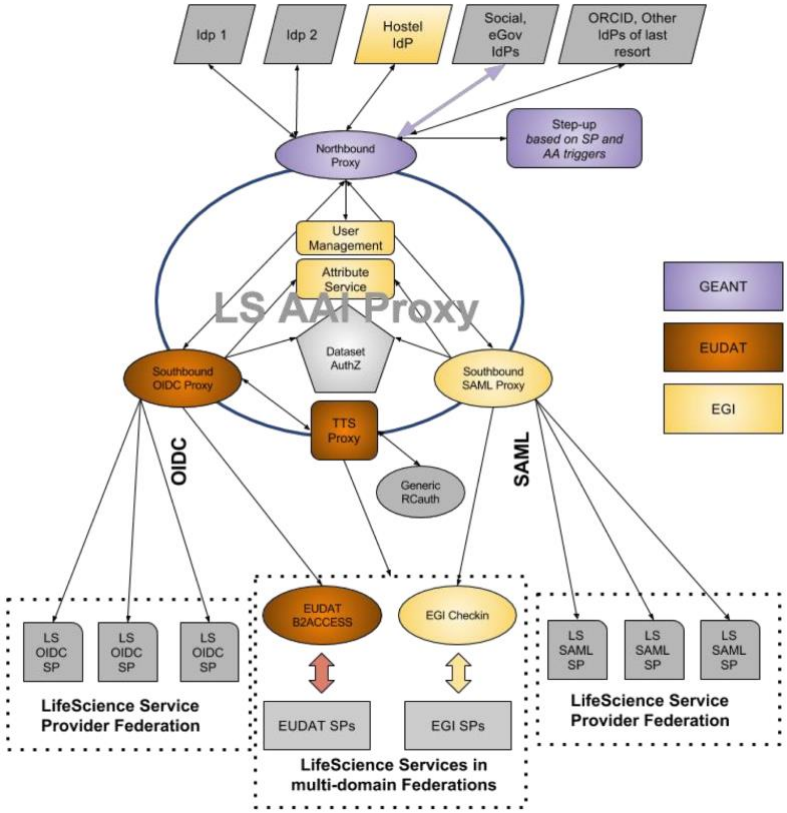

Figure 2. Architecture of the Life Science AARC pilot.

From a technical perspective, the flow to access the e-Infrastructure services is similar to that described in the DARIAH pilot. Specifically, in order to access EGI and EUDAT services, LS users need to go through the EGI Check-in and EUDAT B2ACCESS SP-Proxy respectively. Thus, the e-Infrastructure SP-Proxies can be considered service providers to the LS AAI. Contrary to the LS community specific services, the e-Infrastructure SP-Proxies are expected to present the user with a discovery page offering the choice to login via the LS AAI.

Another key requirement for the LS community is minimising the user interactions necessary to access the e-infrastructure services. The AARC2 Acceptable Use Policy (AUP) alignment study aims to draft a common minimum, or 'baseline', AUP text [AARC-AUP] to ease the flow of user information across AAIs, thereby satisfying this requirement. Authorisation attributes (e.g. representing group membership information in the LS infrastructure) should be also properly translated within the e-Infrastructures. A step-up service [AARC-G029] will be available to increase the LoA of LS users when they need to access services that require authentication with a high LoA. The pilot will be focusing also on aspects such as how to express LoA required for services behind e-Infrastructure SP-Proxies [AARC-G021].

2.2.3 Requirements

- Users already registered with a research infrastructure (LS AAI in this case) find their own research infrastructure from the WAYF (in this case LS AAI) when they try to access an e-infrastructure service (in this case an EGI or EUDAT service).

- When users try to access a service provided by an e-infrastructure (in this case EGI and EUDAT), the service will redirect the users to a Discovery Service, operated by the e-infrastructure, where they can find and select their own research infrastructure cluster (in this case Life Sciences)
- Users management is done within the research infrastructure, where users are assigned to groups. This information (identity and authorisation information) is translated within the e-infrastructure (EGI or EUDAT in this case) and should be used by the e-infrastructure for authorisation purposes; a mapping between the information provided by the research infrastructure (LS in this case) and the e-infrastructure that offers the service (EGI or EUDAT) is needed. The translation of the authentication and authorisation information should require minimum user interaction (ideally the process should be transparent for the user).
- Some e-infrastructure services may require a high level of assurance. A procedure to allow users to increase the level of assurance should be in place.
- Credential delegation for RI services accessing e-infrastructure services on behalf of a user: while accessing a service of a research infrastructure (a science gateway of the LS infrastructure in this case), a (LS) user wants to access resources provided by an e-Infrastructure (e.g. EGI Cloud Compute).

2.3 EPOS

2.3.1 Introduction

The main aim of EPOS is to coordinate, facilitate the integrated use of, and archive high quality Earth Science data across Europe. By nature, EPOS is a distributed Research Infrastructure where Data, Data Products, Software and Services (DDSS) are provided by different communities in the domain of the solid Earth sciences. In this context, EPOS envisages the construction of a central hub called “Integrated Core Services” (ICS) which aggregates all DDSS from various disciplines. From the technical viewpoint, DDSS are provided by a distributed network of endpoints (called Thematic Core Services, TCSs), which use heterogeneous authorization mechanisms. Users access the ICS querying for some data/dataproducts/software/service, and ICS is delegated to fetch the resources on behalf of the user. ICS can be connected to one or more ICS-D, the latter being an external infrastructure that could provide additional resources, such as computing and/or storage, to the EPOS RI.

ICS hub needs to offer an AAI that enables any EPOS user to access the ICS with his/her preferred authentication mechanism (e.g. OAuth, eduGAIN, X509 certificates etc.), delegate this identity to the ICS to fetch resources from the various endpoints (TCSs). These endpoints may implement heterogeneous authorization mechanisms, and the system should realise harmonised access to them. Since EPOS is a federation of federations, AAI interactions between ICS and TCSs can be seen as a case of interoperability between research infrastructures, where scientific tools belonging to different operational domains can be accessed by the EPOS users in a seamless way. A community-based identity will be used to identify the users of the EPOS RI.

The EPOS RI architecture is depicted in Figure 3.

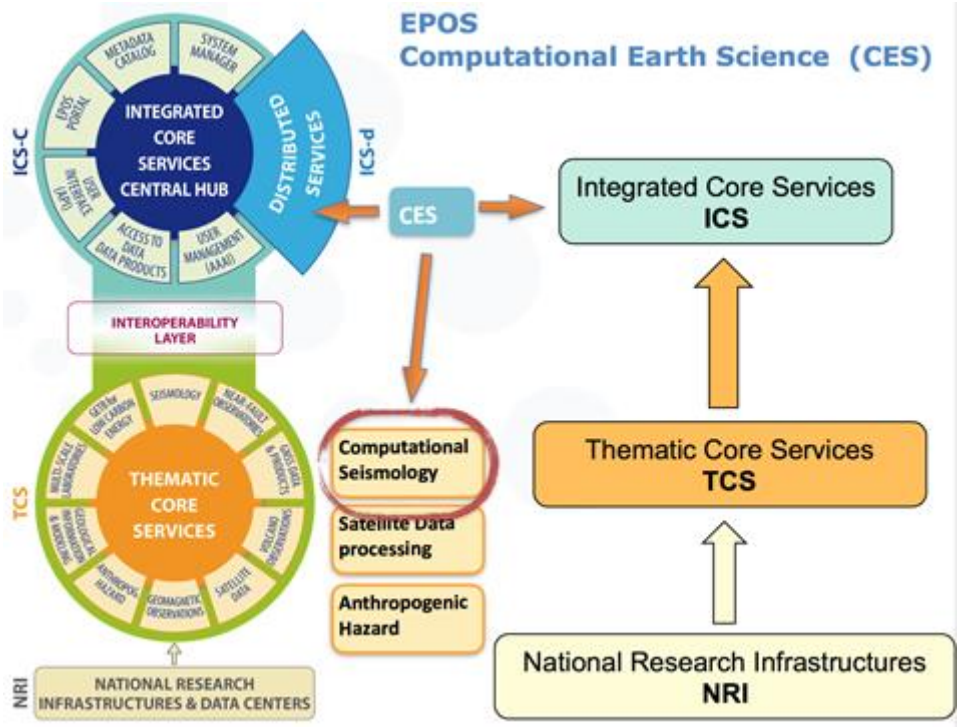

Figure 3. Architecture of the EPOS research infrastructure.

2.3.2 In focus for the AARC pilot

The AARC pilot aims at enabling EPOS TCS services to access e-Infrastructure resources offered by EGI and EUDAT (referred as ICS-D in the EPOS architecture). To analyse this use case, we describe a TCS in the field of Seismology, the VERCE Computational Seismology Platform. VERCE (Virtual Earthquake and seismology Research Community e-science environment in Europe) allows researchers to perform simulations of real earthquakes, allowing also comparison of simulated data with real observations. Users can use publicly available earth models as well as experimental and therefore private ones.

Figure 4 illustrates the interactions between the components of the VERCE platform and the EGI and EUDAT e-infrastructures. In the depicted flow, the VERCE Science Gateway is accessing computing resources (EGI Cloud Compute) and data-staging services (EUDAT B2STAGE/B2SAFE) on behalf the user, in a transparent way.

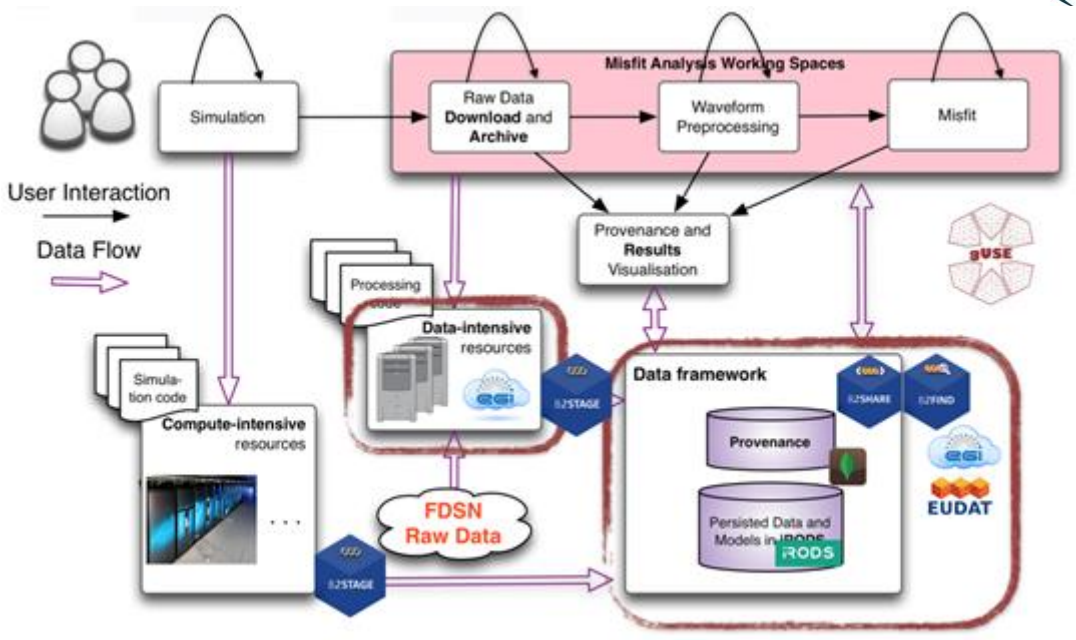

Figure 4. Architecture of the EPOS AARC pilot.

Besides the indirect/delegated access to e-Infrastructure resources via a TCS service, EPOS users also require direct access to e-Infrastructure services (e.g. EGI Cloud Compute). To this end, EPOS users need to log in through Check-in or B2ACCESS e-Infrastructure SP-Proxies using their community identity in order to access EGI and EUDAT services, respectively.

In both the delegated and the direct user access to e-infrastructure resources use-cases, authorisation attributes (e.g. group membership information managed in the EPOS infrastructure) should be properly translated at EGI/EUDAT level.

2.3.3 Requirements

The main requirements for supporting the delegated and direct cross-infrastructure access to resources are the following:

- User enrolment and group management should be handled by the research infrastructure (EPOS). Identity and group membership information from the research infrastructure should be communicated to the e-infrastructures (EGI and EUDAT in this case) for authorisation purposes. Thus, a mapping between the information provided by the research infrastructure and the e-infrastructures that offer the service (EGI and EUDAT) is needed. The translation of the authentication and authorisation information should require minimum user interaction (ideally it should be completely transparent for the user).
- Access to specific e-infrastructure resources may require authentication with a high LoA. A procedure to allow users to increase the level of assurance needs to be in place.

- Delegated access to e-infrastructure resources: Research infrastructure thematic services (TCS of the EPOS infrastructure in this case), should be able to access resources provided by different e-Infrastructures (EGI Cloud Compute and EUDAT B2STAGE/B2SAFE in this case) on behalf of a user of the research infrastructure (EPOS).
- When users try to access a service provided by an e-infrastructure (in this case EGI and EUDAT), the service will redirect the users to a Discovery Service, operated by the e-infrastructure, where they can find and select their own research infrastructure cluster (in this case EPOS).

2.4 EUDAT PRACE

2.4.1 Introduction

EUDAT (European Collaborative Data Infrastructure) and PRACE (Partnership for Advanced Computing in Europe) established a collaboration to support users with the management of data sets resulting from scientific computations by providing high-quality, federated, and seamlessly accessible data services [1].

The EUDAT-PRACE collaboration has been based on the implementation of community pilots resulting from periodic joint calls for proposals. In particular, since 2014 EUDAT contributes to PRACE Distributed European Computing Initiative (DECI) calls by offering data services to projects interested in managing data during or after the end of their PRACE grants. Some examples of PRACE projects supported by this collaboration are:

- Hybrid 3D simulations of turbulence and kinetic instabilities at ion scales in the expanding solar wind (Astro sciences);
- Multiscale simulations of nanoparticle suspensions (Engineering);
- High Resolution EC-Earth Simulations (Earth Sciences);
- Effect of rotation and surface roughness on heat transport in turbulent flow (Engineering).

While EUDAT has already implemented an AARC BPA-compliant AAI, at the moment PRACE is using IGTF compliant X.509 certificates for authentication, a centralised LDAP service for managing identities, and local gridmap files for mapping user certificates to local identities. PRACE users are usually coupled into small groups (per project rather than per community) who produce large volume of data and share it internally. The specifics of HPC and volume of the data to be stored usually requires bash or at least command line mode for the processing, which in turn requires non-web browser access to the data services. The data must be effectively transferred between the e-infrastructures (many EUDAT member institutions are also PRACE members).

2.4.2 In focus for the AARC pilot

EUDAT offers a number of data services, but the best fitting for most PRACE collaboration requirements, and the piloted ones in particular, were B2STAGE/B2SAFE. B2SAFE provides long-term archiving and preservation features configurable by elastic policies. The service is built on iRODS software that manages data access using local (system) accounts and groups.

B2STAGE works as a front-end providing data transfer capabilities. B2STAGE currently supports the GridFTP protocol and a RESTful HTTP API is in progress. The GridFTP implementation of B2STAGE uses X.509 certificates for authentication and B2SAFE maps the user identified by the subject of the certificate to the local iRODS account and groups. The certificate may be either a short-lived one, issued by the B2ACCESS portal for an authorized user, or a certificate signed by external authority (e.g. IGTF compliant). The HTTP API implementation uses an OAuth2 token obtained from the B2ACCESS portal for authentication. Then, B2STAGE uses this token to obtain the short-lived certificate from B2ACCESS which is used to interact with B2SAFE (map the user to local account and groups).

GridFTP supports third party transfers but existing implementations of the protocol require using the same certificate on both sides where the transfer takes place (e.g. PRACE site and B2STAGE). PRACE accepts only IGTF compliant certificates and certificates issued by B2ACCESS are not IGTF compliant, so they cannot be used in the EUDAT-PRACE use case. On the other hand, certificates signed by external authorities do not contain attributes issued either by PRACE or EUDAT (B2ACCESS) that could be used for authorisation. Thus, the authorisation requires background exchange of user information between all involved parties (PRACE LDAP – B2ACCESS – B2SAFE). For example, user attributes and placement in the PRACE LDAP tree may be used for making authorisation decisions and assigning users to groups in B2ACCESS. Then, the user's group membership and attributes in B2ACCESS can be used for making authorisation decision and assigning to groups in B2SAFE.

In the HTTP API case, the user needs to login to the B2ACCESS portal (e.g. using an IGTF compliant certificate) prior to accessing B2STAGE/B2SAFE to obtain the OAuth2 token. The background exchange of user information is also needed.

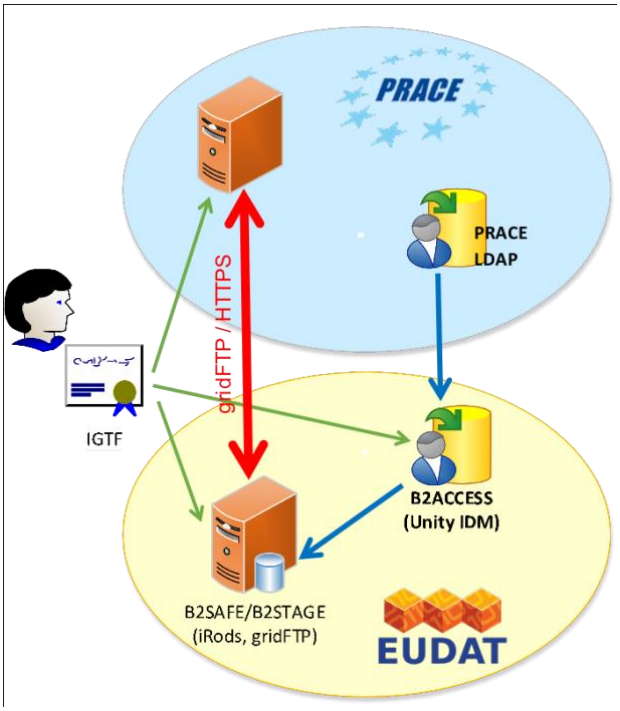

Figure 5. Architecture of the EUDAT PRACE AARC pilot.

The use case is illustrated in Figure 5. The red arrow is data transfer, the green ones are user credentials and the blue ones are background authorisation data flow. The PRACE AAI is not based on the AARC BPA, therefore cross-infrastructure AAI interoperability is currently achieved by integrating the PRACE AAI as an attribute authority for the EUDAT AAI. More details about this integration are available in 4Appendix A.

2.4.3 Requirements

The main requirements for supporting cross-infrastructure access to resources are the following:

- Identity and group membership information from PRACE should be communicated to the e-infrastructure (EUDAT in this case) for authorisation purposes. Specifically, the information maintained in the PRACE LDAP directory needs to be mapped to group membership information which is used for determining access to EUDAT services (e.g. B2SAFE). The translation of the authentication and authorisation information should require minimum user interaction (ideally it should be completely transparent for the user).
- When PRACE users try to access a service provided by the e-infrastructure (in this case EUDAT), the service will redirect the users to a Discovery Service, operated by the e-infrastructure, where they need to use IGTF-compliant certificates for authentication.

2.5 Long tail of science and scientific web applications and tools

2.5.1 Introduction

There is a large number of individual researchers and small research groups, the so-called Long Tail of Science (LToS), that do not belong to a particular Research Infrastructure. Long tail users are usually interested in accessing computing, storage and data resources for a short term in order to perform workflows via general purpose scientific web applications and tools. The need for interoperable cross-infrastructure AAI arises when these portals need to use resources offered by different e-infrastructures.

2.5.2 In focus for the AARC Pilot

The use case of accessing general-purpose scientific web applications and tools across different e-infrastructure could be addressed by the EGI-EUDAT interoperability pilot which started in AARC1 and was concluded in the current project in the context of the SA1.2 task[1]. Although this interoperability pilot was aiming at testing the harmonisation of user and group attributes without explicitly targeting a specific community, it can enable LToS users to access services/resources offered by EGI and EUDAT.

[1] https://wiki.geant.org/display/AARC/AARC2+SA1.T2%3A+Support+e-Infrastructures

2.5.3 Requirements

- Since LToS users do not belong to a specific Research Infrastructure, user enrolment and group management needs to be handled by an e-Infrastructure. Effectively, from a logical point of view, the e-Infrastructure should serve a dual purpose: i) managing the LToS community identity and ii) acting as a gateway to e-Infrastructure resources.
- The identity and authorisation information communicated across the infrastructures should be harmonised.
- Credential delegation: A scientific web portal should be able to access resources offered by different e-Infrastructures (e.g. EGI and EUDAT) on behalf of the users.
- Access to infrastructure resources may require a high level of assurance. A procedure to allow users to increase the level of assurance as requested by the target infrastructure needs to be in place.

3 Generic use cases

The research community specific use cases have been analysed in order to extract common patterns presented in this section. These patterns can be considered as generic/recurrent use cases for interoperable cross-infrastructure AAI. For each of the generic cases, we provide the basic flow of steps that describe the interactions between the user and the components of the involved infrastructure AAIs. Alternate flows are also described where necessary.

3.1 Research Infrastructure users accessing e-Infrastructure services

Primary actor: Researcher requiring access to a service offered by a generic e-infrastructure.

Preconditions:

- The researcher has already registered as a user in the AAI of their Research Infrastructure.
- The e-Infrastructure provider has established an agreement with the RI provider on how to interpret community-managed attributes (e.g. group membership and role information) for authorisation purposes.

Basic flow: Figure 6 illustrates the interactions between the components of the AAI systems of the involved infrastructures:

1. A user of the RI A requests access to a service of e-Infrastructure X.
2. The service forwards the request to the e-Infrastructure SP proxy of e-Infrastructure X. The user selects (or is automatically redirected to) the community Proxy, i.e. RI A.
3. The Proxy of e-Infrastructure X forwards the user to the Proxy of RI A.
4. The user selects (or is automatically redirected to) the IdP of his/her Home Organisation.
5. After successful authentication at the IdP of the user's Home Organisation, the user identity is returned to the Proxy of RI A.
6. The Proxy of RI A can query one or more attribute authorities to enrich the user identity with attributes.
7. The user identity enriched with attributes is returned to the Proxy of e-Infrastructure X.
8. The user is redirected to the service with the related authorisation attributes. The service provider either grants or denies access to the protected resources based on the authorisation attributes.

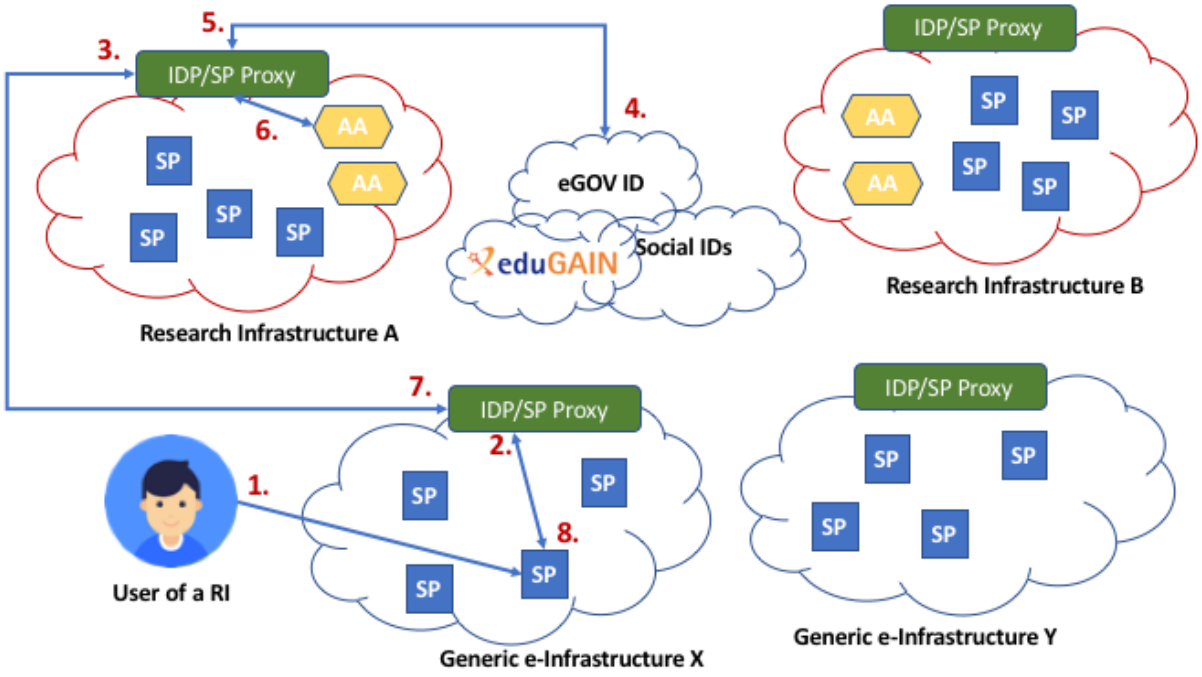

Figure 6. Generic use case: Research Infrastructure users accessing e-Infrastructure services

Alternate flow:

A high level of assurance may be required for accessing some e-Infrastructure services. In the case of services requiring strong authentication via multi-factor authentication (MFA), the requirements and a proposed flow are provided in guidelines document [AARC-G029].

Interested communities: DARIAH, Life Science Community, EPOS, Long tail of science and scientific web applications and tools.

3.2 Research Infrastructure services accessing e-Infrastructure resources on behalf of the user

Primary actor: Researcher requiring access to resources provided by different generic e-infrastructures in a workflow orchestrated by a service in the Research Infrastructure.

Preconditions:

- The researcher has already registered as a user in the AAI of their Research Infrastructure.
- The assurance information can be expressed by the Research Infrastructure so that assurance elements need not be re-asserted or re-computed by the recipient e-Infrastructure(s) (relevant for alternate flow)

Basic flow: Figure 7 illustrates the interactions between the components of the AAI systems of the involved infrastructures:

1. A user of RI A requests access to a service of RI A.
2. The service forwards the request to the Proxy of RI A.
3. The user selects (or is automatically redirected to) the IdP of his/her Home Organisation.
4. The user performs authentication at his/her own IdP and the user identity is returned to the Proxy of RI A.
5. The Proxy of RI A can query one or more attribute authorities to enrich the user identity with attributes.
6. The user is redirected to the service with the related authorisation attributes.
7. The service performs a workflow on behalf of the user which requires access to services provided by e-infrastructure X and Y. The execution of the workflow proceeds without the presence of the user. For more information on credential delegation and federated access to non-web-browser-based services across operational domains, please refer to [AARC-G005] and [AARC-G024].

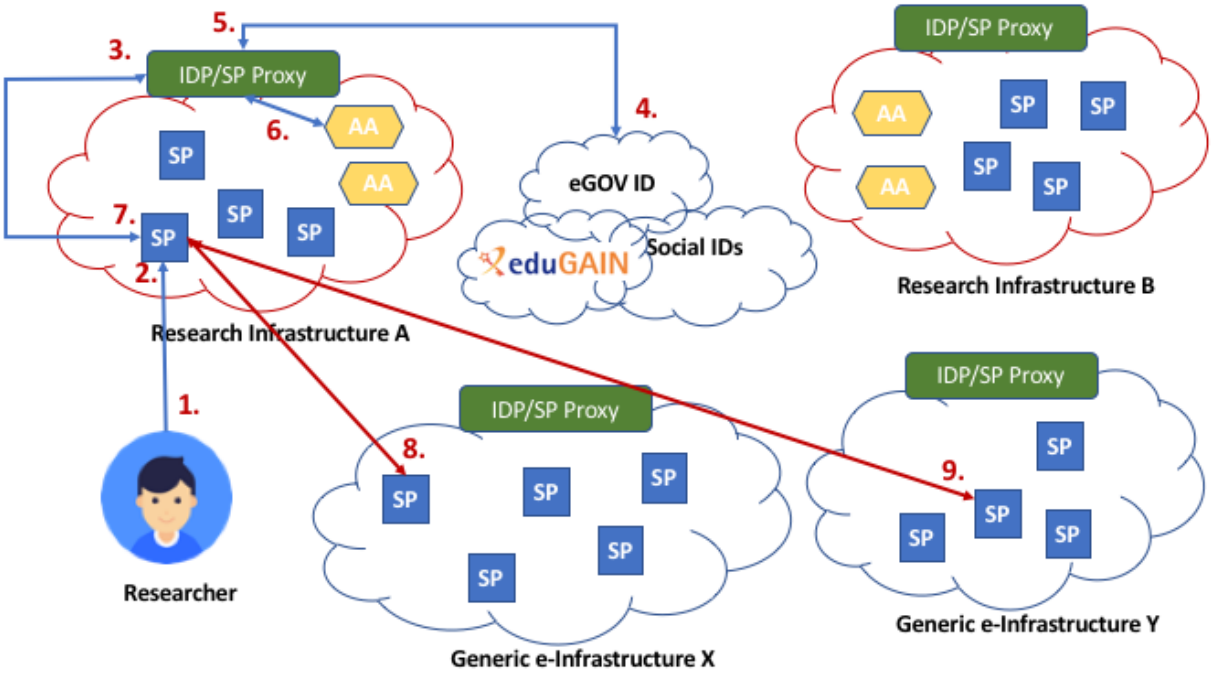

Figure 7. Generic use case: Research Infrastructure services accessing e- Infrastructure resources on behalf of the user.

Alternate flow:

A high level of assurance may be required for accessing sensitive resources. The assurance elevation mechanism should be triggered when one or more services involved in a given workflow impose assurance requirements.

Please refer to guidelines document [AARC-G029] for a proposed step-up authentication flow in the case of services requiring strong authentication via multi-factor authentication (MFA).

Interested communities: Life Science Community, EPOS, ICOS, Long tail of science and scientific web applications and tools.

4 Conclusions

This document provided an analysis of research community use cases that require cross-infrastructure access to resources. This analysis allowed to extract common patterns that can be considered as generic/recurrent use cases for interoperable cross-infrastructure AAI. Specifically, we identified two generic flows: i) Research Infrastructure users accessing e-Infrastructure services, and ii) Research Infrastructure services accessing e-Infrastructure resources on behalf of the user. For each of the generic cases, we provided the basic flow of interactions between the user and the components of the involved infrastructure AAIs. Alternate flows were also described where necessary. It should be noted that the identified generic use cases do not cover all possible cross-infrastructure flows. More complex scenarios requiring the combination of these cases can be envisaged, however such scenarios do not reflect the current research community use cases.

The identified generic cross-infrastructure use cases will serve as input for the evolution of the AARC Blueprint Architecture aiming to enable researchers to seamlessly access and share federated services and resources offered by different infrastructure providers using existing credentials from their home organisations.

Appendix A Example cross-infrastructure interoperability in the case of non-BPA compliant AAIs

This section describes a use case of accessing resources across infrastructures that involve non-BPA compliant AAIs. This example use case is based on the EUDAT/PRACE interoperability pilot analysed in Section 2. Achieving interoperability for that particular use case required integrating the (non-BPA compliant) PRACE AAI group management system as an attribute authority for the (BPA compliant) EUDAT AAI.

Primary actor: Researcher requiring access to a service of a generic e-infrastructure.

Preconditions:

- The researcher has already registered as a user in both the non-BPA compliant and the BPA-compliant AAI.
- The e-Infrastructure AAI has established a direct connection with the Attribute Authority (AA) of the RI AAI for retrieving community-managed attributes (e.g. group membership and role information) for authorisation purposes.

Basic flow: Figure 8 illustrates the interactions between the components of the AAI systems of the involved infrastructures:

1. A user of (non-BPA complaint) RI A requests access to a service of (BPA-compliant) e-Infrastructure X.
2. The service forwards the request to the Proxy of e-Infrastructure X.
3. The user selects (or is automatically redirected to) the IdP of his/her Home Organisation IdP.
4. The user performs authentication at his/her Home Organisation IdP and the user identity is returned to the Proxy of e-Infrastructure X.
5. The Proxy of e-Infrastructure X queries the Attribute Authority (AA) of RI A for group membership information about the authenticated user. For this particular pilot, the query is based on the distinguished name of the user's certificate, since the RI (in this case PRACE) relies solely on certificates for user authentication.
6. The Proxy of e-Infrastructure X incorporates the retrieved group membership information into the user identity.
7. The user is redirected to the service with the related authorisation attributes. The service provider either grants or denies access to the protected resources based on the authorisation attributes.

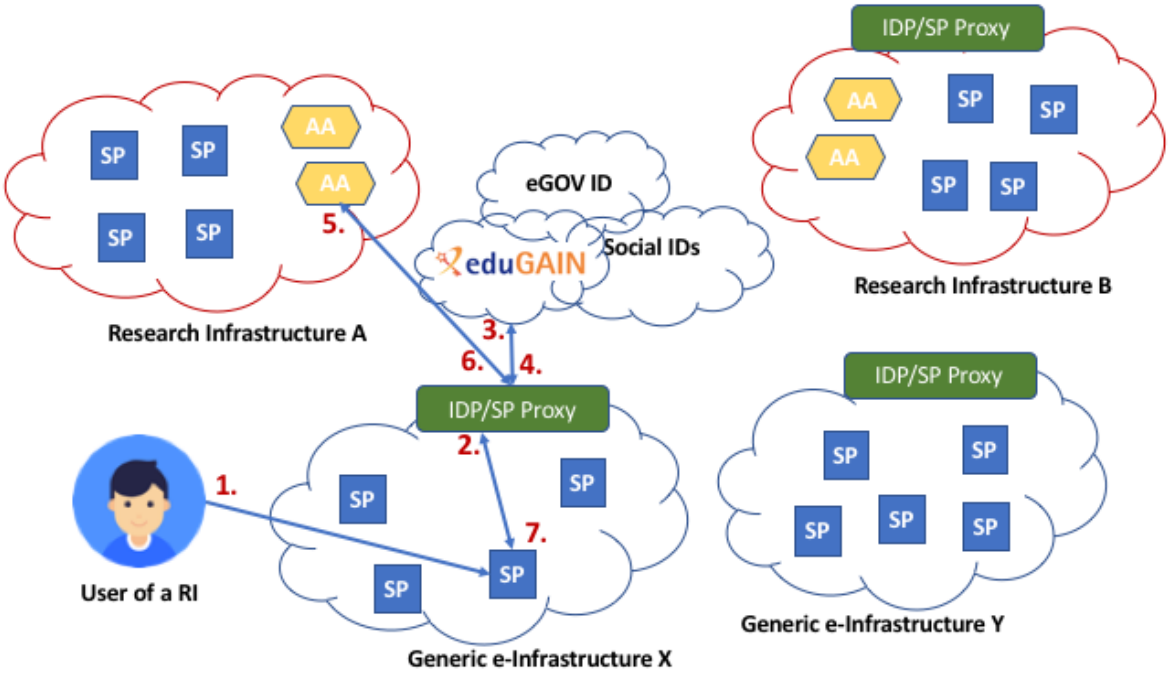

Figure 8. Example cross-infrastructure interoperability in the case of non-BPA compliant AAIs

Interested communities: EUDAT/PRACE.

Appendix B Example cross-infrastructure interoperability use cases beyond AARC

B.1 ICOS

B.1.1 Introduction

ICOS (Integrated Carbon Observation System, https://icos-ri.eu) is a European research infrastructure (RI) which aims to facilitate research to help understand the greenhouse gas (GHG) budgets and perturbations in Europe and adjacent regions. ICOS is based on the collection of high-quality observational data by measurement stations operated long-term (15+ years) as national networks in the RI member states. Apart from observational data products, which are distributed via the ICOS Carbon Portal (CP), also various "elaborated data products", i.e. outputs of modelling activities based on ICOS observations, are compiled and distributed by the CP. Furthermore, the CP facilitates the creation of such elaborated products by the research community.

B.1.2 In focus for the pilot

This pilot has been started in the EGI-ENGAGE and EUDAT2020 projects and is now continuing in the context of the EOSC-hub project. It is focussing on the integration of the "footprint tool" with EGI and EUDAT services. This tool is an interactive, on-demand service that computes and visualizes the sensitivity of GHG concentration signals at potential and existing ICOS atmospheric measurement stations to GHG emissions and fluxes from different sources.

It consists of a dockerized version of the atmospheric transport model STILT (Stochastic Time-Inverted Lagrangian Transport) and a web interface for the communication with the users. Both are run in Virtual Machines (VMs) in the EGI Federated Cloud. The work and data flow in this use case is illustrated in the next figure.

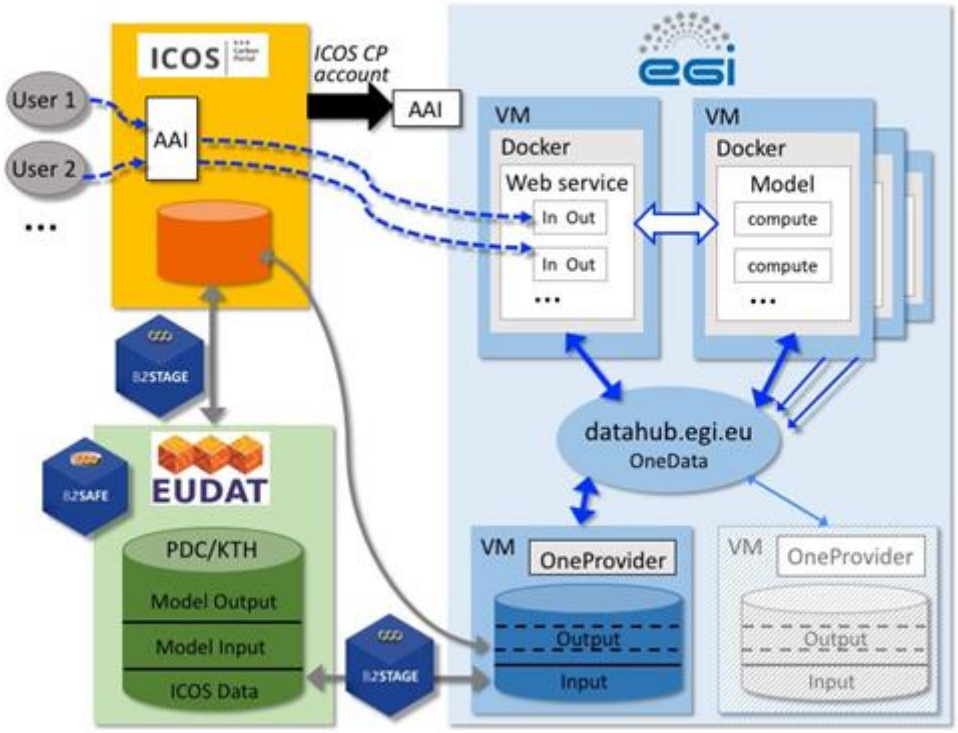

Figure 9. Architecture of the ICOS pilot.

Users of the footprint tool access the service via the ICOS CP website, where also authentication and authorization is handled by ICOS CP. In a similar way of the EPOS VERCE TCS case described in the previous section, all interactions with EGI and EUDAT services should be handled by ICOS CP on behalf of the user and a direct interaction, behind the scene, between an EGI service (EGI Cloud Compute) and EUDAT services (EUDAT B2STAGE/B2SAFE) is foreseen (data output of the processing are transferred from the EGI DataHub to the EUDAT B2SAFE, for long term preservation, via EUDAT B2STAGE). After the user is logged in the ICOS CP, and the application accesses EGI and EUDAT services on his behalf, he should not be requested to provide is credentials again. Login to the EGI and EUDAT services should happen in a transparent way for the user implementing a credential delegation (impersonation) between the ICOS CP and the e-Infrastructure services (EGI Cloud Compute and EUDAT B2STAGE/B2SAFE).

It should be noted that ICOS authorisation attributes should be properly translated at EGI/EUDAT level and a step-up service should be available to increase the LoA of ICOS users when they need to access EGI/EUDAT services that require a high LoA (e.g. EGI Cloud Compute).

B.1.3 Requirements

- Access to e-infrastructure resources may require authentication with a high level of assurance. A procedure to allow users to increase the level of assurance needs to be in place.

- Credential delegation between e-infrastructure services: while accessing a service of an e-Infrastructure (EGI Cloud Compute in this case), a (ICOS) user wants to access a service provided by another e-Infrastructure (EUDAT B2STAGE/B2SAFE in this case) without the need to login again.

- Credential delegation between RI and e-infrastructure services: while accessing a service of a research infrastructure (a TCS of the EPOS infrastructure in this case), a (ICOS) user wants to access a service provided by an e-Infrastructure (EGI Cloud Compute).

References

[AARC-AUP]	Baseline Acceptable Use Policy and Conditions of Use. URL https://wiki.geant.org/pages/viewpage.action?pageId=108013622
[AARC-G005]	Credential delegation (AARC-G005). URL https://aarc-project.eu/guidelines/aarc-g005/
[AARC-G021]	Exchange of specific assurance information between Infrastructures (AARC-G021). URL https://aarc-project.eu/guidelines/aarc-g021/
[AARC-G024]	Federated access to non-web services across different operational domains. URL https://aarc-project.eu/guidelines/aarc-g024/ (work in progress)
[AARC-G029]	Guidelines on stepping up the authentication component in AAIs implementing the AARC BPA (AARC-G029). URL https://aarc-project.eu/guidelines/aarc-g029/

Glossary

AA	Attribute Authority
AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration
API	Application Programming Interface
AUP	Acceptable Use Policy
BPA	Blueprint Architecture
eduGAIN	International interfederation service interconnecting research and education
IdP	Identity Provider
IGTF	Interoperable Global Trust Federation
LDAP	Lightweight Directory Access Protocol
LoA	Level of Assurance
OAuth2	Standard protocol for authorisation
OIDC	OpenID Connect, an identity layer on top of the OAuth 2.0 protocol
RI	Research Infrastructures
URL	Uniform Resource Locator
VM	Virtual Machine
VO	Virtual Organisation