22 April 2016

GN4-1 White Paper: Comparison of Authentication and Authorisation Infrastructures for Research

Work Package/Activity:	SA5
Task Item:	Task 5
Dissemination Level:	PU (Public)
Lead Partner:	SWITCH
Document Code:	GN4-1-16-37dcb3
Authors and Contributors:	L. Hämmerle (SWITCH/GN4-1), R. Sabatino (GEANT Limited/GN4-1), T. Lenggenhager (SWITCH/GN4-1), M. L. Mantovani (GARR/GN4-1), P. Pilt (HITSA/GN4-1), L. Toom (HITSA/GN4-1), J. Jensen (STFC/EUDAT), E. Torroglosa (University of Murcia), Stefan Paetow (JISC), Peter Solagna (EGI), Willem Elbers (CLARIN/EUDAT), Andrea Ceccanti (INFN/INDIGO), Bas Wegh (KIT/INDIGO), Marcus Hardt (KIT/INDIGO), Paul Millar (DESY/INDIGO), Johannes Reetz (MPCDF/EUDAT)

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 691567 (GN4-1).

Contents

Table of Figures

Table of Tables

1 Introduction

1.1 Overview

Today there are several international Authentication and Authorisation Infrastructures (AAIs) and technologies in existence, created to address the federated identity management needs of research and education in Europe as well as the rest of the world. While some of these AAIs and technologies were specifically built for particular research communities (such as DARIAH [DARIAH], ELIXIR AAI [ELIXIR] and CLARIN SPF [CLARIN]), others were built for a more general target group.

These more general purpose AAIs and technologies, on which the community-specific AAIs often rely, include eduGAIN [eduGAIN], EGI [EGI], EUDAT [EUDAT], Moonshot [Moonshot] and to some extent also STORK [STORK]. All of these are internationally or even globally available. They differ in characteristics, feature sets, coverage, governance and technology, even though they all share the same goal: to provide an infrastructure to facilitate the secure exchange of trusted identity data for authentication and authorisation. Some of these AAIs are part of an e-infrastructure that offers more than just authentication and authorisation.

This comparison of AAIs and technologies has been written by members of the GÉANT project with the collaboration and involvement of the AAIs described, which were invited to contribute to the document. The data was gathered in October/November 2015.

1.2 Motivation

As most of the AAIs use different and often complex technologies, it is difficult for people not yet familiar with them to know and understand the most basic concepts of the different AAIs. Even the operators of one particular AAI often have limited knowledge about the technical mechanisms, policies and needs of the main users of the other infrastructures. Obtaining a good overview of these aspects of the different AAIs and technologies is even more difficult for research communities, cloud providers or commercial services that are about to decide which one of these infrastructures to use for their own purposes and how.

1.3 Goals

This document aims at providing a simple overview and comparison of the available general-purpose authentication and authorisation infrastructures and technologies in academia. The goal is to make it easier for research communities and prospective users of these infrastructures to learn the most basic aspects and characteristics about them so that they can make use of one or several of them.

Another goal of the document is to make the operators of the infrastructures described more familiar with alternative infrastructures. This might lead to an increased awareness and willingness

to cooperate in areas where connecting the infrastructures might benefit research and education as a whole.

1.4 Scope and Limitations

This document describes international general-purpose AAIs, services and technologies that are used to provide access to data, services and networks for research and education. AAIs that provide network access exclusively, such as eduroam [eduroam], are not included.

This document does not "provide a requirement analysis from research communities regarding AAI", unlike the AARC document *Deliverable DJRA1.1: Analysis of user community and service provider requirements* [AUCSPR]. Neither does it "evaluate the feasibility of delivering an integrated Authentication and Authorisation Infrastructure" nor make any "recommendations for the delivery of an integrated AAI", as *Advancing Technologies and Federating Communities: A Study on Authentication and Authorisation Platforms For Scientific Resources in Europe* [AAA] does. Both documents are, however, recommended as providing complementary information.

Even though the infrastructures, services and technologies described all provide some form of authentication and authorisation, they are quite different and not easy to compare because they were created by different (academic/government) communities for slightly different purposes.

1.5 In this Document

This document contains overview descriptions of the existing general-purpose AAIs, each using the following structure:

- Introduction.
- History and Current Status.
- Intended Audience.
- Operation and Governance.
- Business Model.
- Underlying Technology.
- Major Benefits.
- Limitations.
- How to Join.

In addition to the individual overviews, the document provides a summary comparison table, considers the relationship between the AAIs, and outlines the known collaborations and opportunities for future collaborations.

It also provides an overview of the upcoming AAI INDIGO Data-Cloud [INDIGO] and summarises the community-specific AAIs before offering conclusions and recommendations.

2 Existing Authentication and Authorisation Infrastructures

This section contains overviews of five general-purpose authentication and authorisation infrastructures, services and technologies that are used internationally and that are general-purpose. This means that they are not limited to just one single research community or discipline but can be used by different (research) communities for different purposes. The five AAIs described are: eduGAIN, EGI, EUDAT, Moonshot and STORK.

2.1 eduGAIN

2.1.1 Introduction

What is eduGAIN?

eduGAIN [eduGAIN] is a so-called "interfederation" service developed within the GÉANT project [GÉANT]. It connects many different SAML-based academic Authentication and Authorisation Infrastructures of its – mostly national – member federations, which are mostly operated by the National Research and Education Network (NREN) of the respective country. Therefore, eduGAIN's actual components are mostly operated by the participant federations (e.g. national federation operators) and their federation members (e.g. universities, research institutions).

What does eduGAIN do?

eduGAIN provides a technical and policy framework to enable the exchange of trusted authentication and identity information across the borders of its member federations. Its goal is to extend the national Single Sign-On (Web SSO) to worldwide Web SSO, primarily for members of the research and education community. Service and Identity Providers, as well as their affiliated users, are enabled to access each other's services via their national identity federations.

Does eduGAIN provide anything else besides an authentication and authorisation function?

No, eduGAIN (together with its member federations) provides only information in the form of attributes that can be used for authentication and authorisation of users.

2.1.2 History and Current Status

When was eduGAIN created and by whom?

The work on eduGAIN started as a research activity in the EU GÉANT project GN2 (2004–2009), just when the first SAML-based national identity federations emerged in Europe and the US. The eduGAIN service activity that was started in the successor project GN3 (2009–2013) built upon the eduGAIN work that was developed in the GN2 project. Whereas the first attempt to develop eduGAIN during GN2 was still making use of gateways and protocol converters (as in STORK), the architecture was heavily changed to a distributed full-mesh model during the GN3 project. This was mostly due to the fact that SAML2 became the de facto standard for federated identity management in academia during this time, which made protocol converters obsolete. On 1 April 2011, eduGAIN became an operational service. From April 2013, funding of the service was continued as part of the GN3Plus and GN4-1 projects (2013–2015 and 2015–2016 respectively).

What is eduGAIN's current coverage (numbers of countries, organisations and users)?

As of March 2016 eduGAIN has 43 member Federations (including 5 voting-only members) and 8 candidates (for latest figures, see [eduGAIN-Status]). There are over 2,000 Identity Providers (IdPs) and 1'000 Service Providers (SPs) (for latest figures, see [eduGAIN-Statistics]). One Identity Provider can represent several dozen organisations of a country because – depending on the federation's architecture – it acts as a proxy. The exact number of users that have an eduGAIN-enabled account is unknown due to eduGAIN's distributed architecture, but counting the number of IdPs and assuming that most organisations have more than 1,000 staff and students, the number of users that have an eduGAIN-ready account is likely to exceed 30 million.

2.1.3 Intended Audience

For whom is eduGAIN operated primarily?

eduGAIN is a service that is open to research and education federations worldwide. It is operated for the benefit of students, lecturers and, in particular, researchers from the worldwide higher education community. Participation is not limited to Europe or to members of the GÉANT project. Therefore, the intended audience are the members of different research and education communities worldwide, particularly the students, staff members and faculty of higher education institutions (e.g. universities, research institutions) and research projects.

Are there further eligible intended audiences for eduGAIN?

Commercial companies can also offer their services in eduGAIN but they are typically not allowed to bring identities into national federations. For example, Microsoft can offer a service that allows students to download their software for free in a national federation, but Microsoft employees do not generally have an account in that federation. This is generally true for eduGAIN, even though the borders of commercial and research companies are not always so clear.

For whom is eduGAIN suited?

eduGAIN is suited to operators of web services that want to allow all academic users worldwide to log into their services while getting some identity and affiliation information about those users. This includes international research communities, universities collaborating in, for example, online courses, as well as e-journal and cloud providers.

2.1.4 Operation and Governance

Who operates eduGAIN?

The sponsor of the few central eduGAIN services is the GÉANT project and its main members, the NRENs. Everyday operations such as managing the eduGAIN Metadata Distribution Service (MDS) are managed by the eduGAIN Operations Team. However, most of the actual services (Identity Providers and Service Providers) are or will be mostly operated by universities, research institutions and commercial companies.

Who controls and governs eduGAIN?

The sponsor (GÉANT) nominates the eduGAIN Executive Committee (eEC), which is responsible for approving policies and profiles.

The eduGAIN Steering Group (eSG) [eduGAIN-SG] consists of one delegate and one deputy per eduGAIN member federation. The eSG is responsible for approving new member federations and appointing the eduGAIN Operations Team (eOT).

Where is support for eduGAIN available?

Generally, the local federation via which an IdP or SP has joined eduGAIN provides a helpdesk. These are usually the same helpdesks that also provide support for services and users that exclusively use services within the local federation.

Support for federation operators is provided by edugain-ot@lists.geant.org and third-level integration support for operators of services is provided by edugain-integration@geant.net.

In addition, the mailing list edugain-discuss@geant.net can be used to ask questions or exchange experiences.

2.1.5 Business Model

How is eduGAIN financed?

The development and operational costs of eduGAIN were funded by the GÉANT projects (GN2 – GN4-1). The operational costs to maintain the core eduGAIN services (in particular the MDS) are low. There is no central helpdesk as eduGAIN support has to be provided by the member federations for

their local Service and Identity Providers. Participating in eduGAIN is free for the member federations.

How sustainable is eduGAIN?

The structure of eduGAIN was designed such that only a little manpower is needed to operate the core components (mostly the MDS). The main work to provide interfederation login via eduGAIN is done by the eduGAIN member federations. For most of them their national AAI has already become a business-critical infrastructure[1]. Therefore, eduGAIN's sustainability depends on the sustainability of the different national AAIs that are part of eduGAIN.

2.1.6 Underlying Technology

What protocols and technologies are used by eduGAIN?

It was intended that the eduGAIN framework should be technology agnostic, such that multiple protocols could be supported in the future. However, as of 2015, Security Assertion Markup Language (SAML) version 2.0 is used almost exclusively for exchanging authentication and authorisation data between an Identity Provider and a Service Provider. eduGAIN mandates the use of the Web SSO profile as a minimum common denominator between the different member federations.

The syntax for representing user attributes follows the MACE-Dir SAML Attribute Profile. The eduPerson [eduPerson] and SCHAC [SCHAC] attribute schemas are preferred.

What does the architecture of eduGAIN look like?

eduGAIN consists of the following technical components:

- Identity Provider (IdP): Authenticates users and issues SAML assertions about the user containing user attributes.
- Service Provider (SP): Consumes SAML assertions to perform access control and make user attributes available to the (web) application that it protects.
- Discovery Service: Lets the user choose his IdP by selecting an organisation he is affiliated with. Often is directly integrated into an SP.
- Federation: A set of organisations that agree to interoperate under a certain rule set. Consists of SPs and IdPs. Most eduGAIN federations are operated by the NREN of a country.
- Hubs (H): Basically a special case of a proxy that acts as SP and IdP at the same time. Used by some so-called hub-and-spoke (H&S) federations with one central hub that is connected to all SPs and IdPs in a federation.
- Metadata Distribution Service (MDS): Validates, aggregates and republishes SAML metadata of all eduGAIN member federations and those of their entities that opted in for eduGAIN (or did not opt out).

[1] The motivation for and benefits if operating federations are described in *The Value Proposition for Identity Federations* [VPIF].

The architecture of eduGAIN is shown in Figure 2.1.

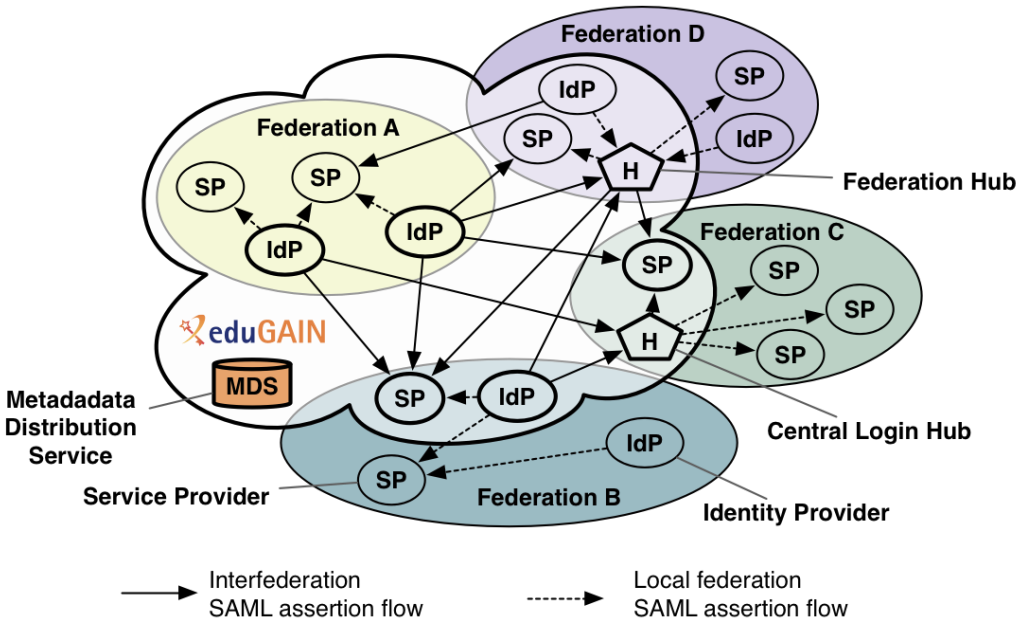

Figure 2.1: eduGAIN architecture

What does the eduGAIN login flow typically look like?

The typical eduGAIN login flow is described below and shown in Figure 2.2.

- A user wants to access a web service (1).
- He clicks on a "Login" button, which sends him to the IdP Discovery Service (2) where he chooses his IdP by selecting his home organisation (4). Often the Discovery Service is on the same host as or even integrated in the Service Provider (SP).
- The Discovery Service then sends the user's web client back to the service together with the identifier of the selected IdP (5).
- The SP then creates a SAML authentication request and sends it via the user's web client to the selected IdP (6).
- There, the user enters his credentials (8).
- The IdP then sends the user back to the SP together with a SAML assertion containing information about the user (9).
- Using the user information in the form of attributes, the SP can then recognise the user and make an authorisation decision.

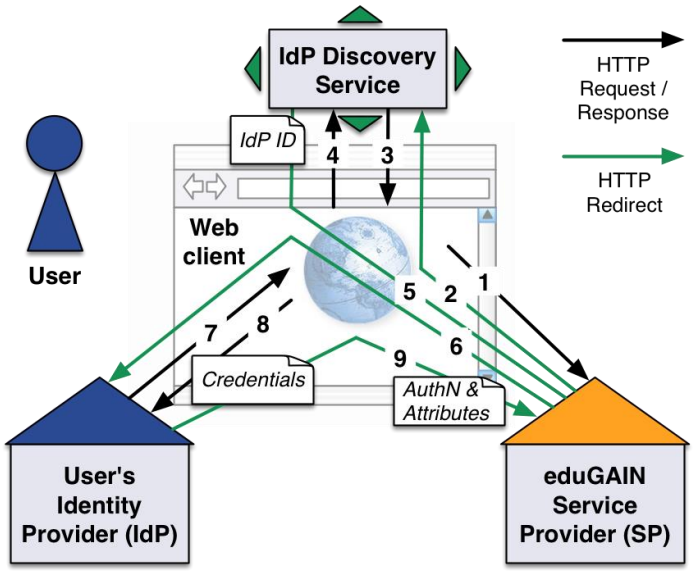

Figure 2.2: eduGAIN login flow

2.1.7 Major Benefits

- Easy collaboration: eduGAIN simplifies educational collaboration and supports distributed research infrastructure projects. Through eduGAIN, Identity Providers offer a greater range of services to their users, delivered by multiple federations in a truly collaborative environment; Service Providers offer their services to users in different federations, increasing their target market; and users seamlessly benefit from the wider range of services.
- Country coverage: eduGAIN has a large coverage when it comes to connected national identity federations. Because eduGAIN is based on a distributed architecture, there is no one, central user directory.
- Number of users: The number of connected users is very high compared to other infrastructures but also difficult to measure. Each of the several hundred Identity Providers is connected to one or several user directories. Therefore, knowing the exact number of users is not possible but assuming that all member federations add their users to eduGAIN in the coming years, the total number of eduGAIN users is likely to exceed 30 million [REFEDS-NUPF].

2.1.8 Limitations

- User coverage: National identity federations often do not cover 100% of all education and research institutions in a country. Therefore, there are still quite a few users that do not have a federated user account. Also, not all the federated organisations of a national federation are part of eduGAIN. Some eduGAIN member federations use an opt-in model for their federated organisations to join eduGAIN. The opt-in model results in a slow adoption.
- Data protection: Different data protection laws may restrict and hinder cross-border authentication because some organisations are hesitant to release identity information

about their users to services outside the same jurisdiction or federation. This is not an eduGAIN-specific problem.

- Focus on web-based services: Non-web services would require other SAML profiles, such as the SAML ECP profile, which is hardly deployed by eduGAIN Identity Providers. Therefore, non-web applications are difficult to support via eduGAIN as of September 2015.
- Branding issues: For users it is often not obvious that they are accessing a service via eduGAIN because there is no official branding for eduGAIN. To have no branding was an active decision taken at the beginning of eduGAIN, to ensure that the login process for end users was the same whether the service was provided by the national AAI or operated in another federation via eduGAIN.
- Only implicit level of assurances: The eduGAIN member federations have different rules regarding the identity vetting procedures and authentication policies. eduGAIN does not mandate such rules to the member federations and no explicit and widely used assurance information is available about users that access an eduGAIN service. Neither is there an agreed-on standard to express assurance levels. However, user identities are managed at or by research organisations that have their own interest in keeping identity data up to date and correct. The same accounts used to access eduGAIN services are also used within the same organisation or the same identity federation. For staff, faculty and student users, one can safely assume that the identity vetting as well as the identity management processes are much better than for affiliate (guest) users.

2.1.9 How to Join

How could a research community make use of eduGAIN?

To offer a service in eduGAIN or to allow users of an IdP to access services via eduGAIN, an entity must join an eduGAIN member federation[2] first that can then include the entity in eduGAIN. To add a single service to eduGAIN, it is generally best to join the local (mostly national) identity federation[3]. If there is no obvious member federation via which to join eduGAIN, the default process is described on the eduGAIN Wiki [eduGAIN-Wiki].

What are the conditions to join eduGAIN?

The joining process for eduGAIN member federations varies. In most cases, some form of agreement has to be signed. Generally, joining eduGAIN by offering a service for the benefit of the higher education and research community is free. Some federations charge commercial services. Bringing identities (e.g. by running an Identity Provider) into eduGAIN is often tied to certain conditions that depend on the federation. Most federations only accept an Identity Provider in their federation and in eduGAIN if operated by a higher education or research institution.

[2] Listed at [eduGAIN-Status].

[3] SAML-based academic identity federations worldwide are listed at [REFEDS-F].

2.2 EGI

2.2.1 Introduction

What is EGI?

The European Grid Infrastructure (EGI) [EGI] "is a highly distributed, multi-disciplinary resource infrastructure, integrating more than 300 resource centres (service providers) and almost 20,000 users grouped in 200 user communities called Virtual Organizations (VO). Currently, authentication and authorisation within EGI is enabled through an X.509-based Public Key Infrastructure (PKIX), based on the Interoperable Global Trust Federation (IGTF) and EUGridPMA Certification Authorities federation." (Source [AUCSPR])

What does EGI do?

EGI "gives European scientists access to the computing, storage and cloud resources and services they need for their research." (Source [EGI-FAQ])

It "creates and delivers open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries". (Source [EGI-About])

Does EGI provide anything else besides an authentication and authorisation function?

EGI provides first-line access to computing and storage resources. Therefore, the X.509-based authentication and authorisation function is only one component of EGI.

2.2.2 History and Current Status

When was EGI created and by whom?

"EGI.eu was created on 8 February 2010 to coordinate and maintain a sustainable pan-European infrastructure to support European research communities and their international collaborators. Its work builds on previous EU-funded projects which nurtured this goal, from the initial concept of a scalable, federated, distributed computing system.

"The distributed computing grid was originally conceived in 1999 to analyse the experimental data produced by the Large Hadron Collider [LHC] at CERN – the European particle physics laboratory located on the Swiss/French border.

"The European DataGrid Project [DGP], which started in January 2001, led the research and development of grid technologies. It established the organisational structure, gathered and analysed requirements, developed middleware (the software that links hardware resources), and provided training to its users. The project proved the grid's successful application in various research fields –

high energy physics, Earth observation and bioinformatics. Upon completion in March 2004, a new project called EGEE (Enabling Grid for E-sciencE) took over the grid's further development in what would result in three successive two-year phases." (Source [EGI-History])

What is EGI's current coverage (numbers of countries, organisations and users)?

"EGI brings together 27 national and 9 federated operations centres encompassing multiple National Grid Initiatives (NGIs) in Europe (IberGrid, NGI_NL and NGI_IT) and in the Asia Pacific and Latin America regions. During the past year, EGI engaged with two new partner infrastructures the South African Grid Initiative and the Ukrainian National Grid." (Source [EGI-IF]) EGI infrastructure also has interoperations agreements, to support worldwide research collaborations, with Open Science Grid and Compute Canada.

2.2.3 Intended Audience

For whom is EGI operated primarily?

EGI is operated for European researchers of all disciplines. (Source [EGI-About])

Are there further eligible intended audiences of EGI?

Given the highly heterogeneous nature of the infrastructure, some of EGI's resource providers can serve commercial users as well.

For whom is EGI suited?

EGI services are designed to serve both large and small collaborations. While the federated services enable distributed collaborations, EGI has also deployed access models from which individual researchers can benefit.

2.2.4 Operation and Governance

Who operates EGI?

EGI is coordinated and managed on behalf of its participants by EGI.eu [EGI-EGIeu], a Dutch foundation established in 2010. EGI's participants are National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs).

"EGI Operations handle the activities required to deliver services at agreed levels to the infrastructure's end users.

"EGI.eu coordinates the work of (about) 32 distributed Operations Centres, 22 of these managed at a national level and one at CERN.

"Locally, Operations Centres are responsible for supporting their Resource Centres, monitoring their performance, collecting requirements and representing them in EGI's Operations Management Board and its associated groups. Globally, the Operations Centre contributes to the development of the EGI operations roadmap and the evolution of EGI operations." (Source [EGI-Ops])

How is EGI controlled/governed?

EGI.eu is governed by the EGI Council, which is responsible for defining the strategic direction of the EGI federation. The Council acts as the senior decision-making and supervisory authority of EGI.eu. The Council participants are the NGIs and EIROs.

The Council delegates oversight of the day-to-day running of EGI.eu to the Executive Board, currently with seven members. The Executive Board devolves financial and organisational responsibility to the Director, who is supported by a staff of about twenty people based at the EGI.eu headquarters in Amsterdam. EGI.eu's work is supported by other workers spread across many organisations around Europe.

Where is support for EGI available?

The technical helpdesk is at [EGI-Helpdesk].

For more generic information, new users and communities can contact their national NGIs (reachable from the EGI.eu website [EGI]) or support@egi.eu.

The EGI Wiki can be found at [EGI-Wiki].

2.2.5 Business Model

How is EGI financed?

EGI.eu is a not-for-profit foundation established under Dutch law in the Netherlands. The foundation has participants and associated participants drawn from NGIs, EIROs, European Research Infrastructure Consortiums (ERICs), and other legal entities. These entities participate in the foundation independently or as the representative of a national e-infrastructure consortium. The main funding streams are: 1) yearly fees paid by the participants and associated participants; 2) income from competitive projects (e.g. Horizon 2020 funding instrument); 3) consultancy and training services.

Currently, the main funding project is EGI-Engage, which supports the evolution of EGI activities and services. The project involves EGI.eu, many EGI.eu participants or associated participants, and research infrastructures.

NGIs are organisations that have a mandate to represent a national e-infrastructure in all matters falling within the scope of EGI.eu. They represent the country's single point of contact for government, research communities and resource centres as regards ICT services for e-science.

(National Grid Infrastructure/Initiative and European Grid Infrastructure/Initiative are deprecated; now only NGI and EGI are used as names. See new EGI Statutes [EGI-Statutes].)

How sustainable is EGI?

EGI is coordinated by a not-for-profit foundation established in 2010, called EGI.eu, and is funded through a combination of participant fees and national funding for long-term operations as well as EC funding for service innovation.

The foundation has participants and associated participants drawn from national e-infrastructures as well as other legal entities that form the governing body (EGI Council). Participants and associated participants also provide the physical and human resources and shared services that enable EGI to deliver, improve and innovate services for research communities.

EGI.eu coordinates areas such as overseeing infrastructure operations, user community support, contact with technology providers, strategy and policy development, flagship events and dissemination of news and achievements.

EGI has stabilised its governance, evolved its business model and implemented management processes to ensure the service levels required for long-term operation.

2.2.6 Underlying Technology

What protocols and technologies are used by EGI?

"Currently, authentication and authorisation within EGI is enabled through an X.509-based Public Key Infrastructure (PKIX), based on the Interoperable Global Trust Federation (IGTF) and EUGridPMA Certification Authorities federation." (Source [AUCSPR])

What does the EGI architecture look like?

The EGI authentication and authorisation architecture is based on the IGTF certification authorities for the provisioning of the personal user certificates.

EUGridPMA produces a distribution of Certification Authority (CA) root certificates that are installed in every single EGI service. EGI actively monitors that the CA distribution is up to date in every production service. The services where the distribution is installed automatically trust certificates released by the CAs as part of the EUGridPMA distribution.

The authorisation in EGI is commonly based on the user's membership of a Virtual Organisation (VO). The most common service to manage VO membership in EGI is the Virtual Organisation Membership Service (VOMS). VOMS manages the community attributes associated to a user (to the certificate subject of the user) and signs the X.509 proxy certificate of the user – a short-lived "copy" of the certificate, which is used to interact with the services – adding the attributes about the VO, if the user is a member of the VO.

The services where the user VO is supported have configured the VOMS host certificates and recognise the information added by the VOMS as valid by validating them with the VOMS public credentials.

What does the EGI login flow typically look like?

Scenario 1: the user owns an X.509 certificate

VO membership:

1. User connects to the VOMS service with the X.509 certificate in their browser.
2. User requests VO membership, including roles and groups within the VO.
3. The Certificate used by the user provides the user's identity and affiliation.
4. VO Manager approves (or rejects) the user in the VOMS service.

Accessing EGI services:

1. The user – with their proxy – generates a short-lived X.509 proxy certificate.
2. The user uses the VOMS client to have the X.509 proxy certificate extended with the VO attributes and signed by VOMS.
3. The user sends a request to an EGI service, attaching the signed X.509 proxy certificate.
4. The service checks that the certificate proxy has been signed by a valid certificate from a trusted CA and that the VO information is signed by a trusted VOMS.
5. If all the checks are successful, the user request is approved and the user can access the service.
6. If requested by the user, the service can hold the user proxy certificate to be used to act on behalf of the user.

Scenario 2: the user does not own an X.509 credential

"To bridge different authentication technologies with X.509, the EGI partners and the user communities are deploying science gateways and portals where users can authenticate with username/password, and access the resources through web-based tools and interfaces. The portals are then generating short lived X.509 credentials that are used to access resources.

"The most common mechanism used to bridge between IdPs and X.509 are the robot certificates, which can generate programmatically short-lived X.509 proxy certificates […]. One of the drawbacks of this solution is that the real user identity is hidden behind the robot certificate. To partially address this issue, EGI is implementing an extension of the X.509 proxy certificate that contains an ID that can identify the user if needed. This is particularly useful for accounting purposes and, at the same time, improves the overall security of the implementation." (Source [AUCSPR])

EGI security policies limit the actions that a user can perform using a robot certificate.

1. The user connects with username/password in the science gateway (SG), possibly with their institutional credentials.

2. The SG – if the user is authorised as an SG user – generates an X.509 proxy certificate using a robot certificate. The SG signs the proxy with the VO proposed to be used. Usually SGs are mono-VO or have a mapping service to multiple VOs. Depending on what the users wants to use, the VO is selected.
3. The SG submits one – or many – requests to the service on behalf of the user using the X.509 proxy certificate generated in the portal.
4. The SG associates the activities run by the user with the user's identity. In this way a user is able to check the status of the jobs or virtual machines or retrieve output data, through the SG, only for the tasks that they submitted.

Note that the usual workflow of an SG prevents the user from downloading the X.509 credentials, which must always be handled by the SG.

EGI is also, however, looking into a "Catch-all IdP service (EGI sso), online CA, attribute authorities to support users without X.509 certificate". (Source [EGI-TaOSC])

2.2.7 Major Benefits

What makes EGI attractive and why?

- Web/non-web support: X.509 certificates can be used for web and non-web authentication. In fact, they can be used in many software products for authentication.
- Two-factor authentication: If the certificate is stored on an external device (such as a USB token) that protects the private key, X.509 login by default offers secure two-factor authentication.
- Scalability: Services do not need to contact the CA for every service call. A user can submit thousands of tasks without any scalability issue related to authentication.
- All the user information is local, sent with the X.509 proxy certificate.
- Delegation is a common requirement of many workflows. X.509 transparently supports delegation (impersonation).

2.2.8 Limitations

- Complicated to use: Retrieving and using X.509 certificates is non-trivial. Some users have difficulties using certificates properly. This problem is emphasised with the increased usage of easy-to-use web applications.
- Identity data: X.509 certificates contain a fixed number of attributes (depending on their profile). Therefore, they can contain too much information in the form of attributes. In some cases this can result in data privacy issues. Users must accept that the information in their certificates is shared with the service providers (but users are actively sending the X.509 proxy to the service – it's their choice).
- Level of assurance: EUGridPMA has different profiles for CAs supporting multiple levels of assurance (LoAs). What happens in practice is that only the highest LoAs are accepted uniformly, and this often translates into users being forced to use strong authentication for non-security-sensitive actions.
- Security: Using X.509 certificates on an external device (e.g. a USB token) complicates its usage because many applications expect the private key to be accessible on the user's

computer. Therefore, some users store certificates and private keys in software on a computer/server, which means that they can be copied and misused quite easily if they are not protected properly. Storing the private key unencrypted is, however, forbidden by the EUGridPMA and EGI policies.

2.2.9 How to Join

How could a research community make use of EGI?

Users are redirected to the NGI in their country, which will provide them with information about the registration authority that can support their institution. “There is a capillary network of certification authorities and registration authorities, distributed among the EGI partners, which can be contacted by users to obtain a certificate. EGI runs a catch-all CA to support users who – for any reason – cannot access another existing CA.” (Source [AUCSPR])

Many eduGAIN users can get their personal certificate from the TERENA Certificate Service (TCS).

What are the conditions to join EGI?

CAs are operated mostly by the NGIs and are free of charge. Usually CAs support only users from their respective countries, but there are no strict limits.

2.3 EUDAT

2.3.1 Introduction

What is EUDAT?

EUDAT [EUDAT] is a network of collaborating, cooperating centres, combining the richness of numerous thematic data centres with the permanence and persistence of some of Europe's largest scientific data centres. EUDAT offers common data services, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network of 33 European organisations. The AAI component of EUDAT is B2ACCESS [B2ACCESS]. This service is based on Unity, an identity management software that describes itself as "a complete solution for identity, federation and inter-federation management". (Source [UNITY])

What does B2ACCESS do?

The B2ACCESS service arbitrates access to other registered Service Providers, called Downstream Service Providers in the context of B2ACCESS. These downstream service providers consume attribute assertions which are provided by the B2ACCESS service when a user accesses one of these services. The role of B2ACCESS is to allow these downstream service providers to make their authentication and authorisation decisions, and carry out other processing required by the downstream service providers, when the user accesses these services. In turn, B2ACCESS may make use of and store the attributes provided by external primary IdPs for a certain period of time. B2ACCESS itself can also act as an Identity Provider to authenticate the users that have registered directly with the B2ACCESS service. In those cases, B2ACCESS assigns a dedicated username and the user defines his/her password.

EUDAT users create their specific EUDAT identity when they authenticate for the first time against the B2ACCESS service by using an existing primary Identity Provider (e.g. eduGAIN, Google, Facebook, GitHub, X.509, etc.). The unique EUDAT ID is then bound to the initially chosen primary identity. Using an external IdP is the recommended way to access and use EUDAT services. On the Service Provider side, B2ACCESS offers different protocols (OAuth2, SAML, X.509 SLCs).

A specific level of assurance (LoA) is assigned to the user identity, depending on the authentication method that was used when the EUDAT ID was created. Social identities imply a lower level of assurance; eduGAIN IdPs and the use of an IGTF X.509 certificate as the primary identity implies a high LoA.

Does EUDAT provide anything else besides an authentication and authorisation function?

As a Collaborative Data Infrastructure (CDI), EUDAT provides a range of data management, sharing and related services that support the full research data management lifecycle, including data transfer, data storage and workspaces, metadata support, data sharing, data preservation, data discovery and access (B2DROP, B2SHARE, B2STAGE, B2SAFE, B2FIND, etc.).

2.3.2 History and Current Status

When was EUDAT created and by whom?

EUDAT's roots are in the Partnership for Accessing Data in Europe (PARADE) initiative, which resulted in the PARADE White Paper (October 2009) defining a "Strategy for a European Data Infrastructure that should be persistent, multidisciplinary, and based on the need of user communities". The concept of a shared pan-European infrastructure was supported and further elaborated by a number of policy and experts bodies, for example, the e-Infrastructure Reflection Group (e-IRG) and European Strategy Forum on Research Infrastructures (ESFRI). The "e-IRG Blue Paper" (September 2010) recommended "to identify and promote common (long-term) data-related services across different RI". Also, the High-Level Expert Group (HLEG) report on Scientific Data (October 2010) called for a "Collaborative Data Infrastructure" for scientific data that supports seamless access, use, reuse, and trust of data. Based on these recommendations the implementation of EUDAT started in October 2012 as a three-year project. In 2015 the project was extended to run for another three years till 2018.

What is EUDAT's current coverage (numbers of countries, organisations and users)?

Any user from any country and organisation can use B2ACCESS. The service, operated by Forschungszentrum Juelich, was registered by the German AAI-DFN federation in October 2015. A data privacy statement [B2ACCESS-DPS] describes the characteristics of the B2ACCESS service as a Service Provider that arbitrates access to other registered Service Providers (Downstream Service Providers, which all have to comply with the GÉANT Code of Conduct). Therefore B2ACCESS is accepted by eduGAIN IdPs and EUDAT users can authenticate against EUDAT Service Providers using their (primary) eduGAIN identity. The Identity Management (IdM) part of B2ACCESS has been in production since November 2015 and since then further IdPs and EUDAT Service Providers have been included. There is currently (January 2016) nothing to report about usage statistics.

2.3.3 Intended Audience

For whom is EUDAT operated primarily?

EUDAT is operated for community data (repository) managers and for individual users from universities, research institutions and science organisations as well as citizen scientists.

It is operated in particular for any of the EUDAT communities, such as – but not limited to – its core user communities: CLARIN (linguistics), ENES (climate), EPOS (Earth observation), ICOS, LTER and VPH (bio-medical sciences). A full list of all communities involved with EUDAT can be found at [EUDAT-Communities]. In addition to these, a growing number of data projects are being implemented that originate either from the EUDAT calls for data pilots or as a result of data service provisioning requests from research infrastructures and communities.

Are there further eligible intended audiences of EUDAT?

Further eligible audiences include anyone who has data management needs and wants to use the EUDAT e-infrastructure, including users and data managers from other Horizon 2020 [Horizon2020] projects or e-infrastructures such as PRACE, EGI or OpenAire.

For whom is EUDAT suited?

EUDAT aims to cater for both non-technical researchers (who often have either institutional or social identities) and technical or operational ones (those who can use X.509 certificates in addition).

2.3.4 Operation and Governance

Who operates EUDAT?

EUDAT is a consortium of 35 EUDAT partners [EUDAT-Partners]. Half of them are Computing and Data Centres that operate the EUDAT services. Operations are governed by an operations coordination team that consists of representatives of these collaborating centres. The core of B2ACCESS, the Unity instance, is operated by Forschungszentrum Juelich (Germany).

How is EUDAT controlled/governed?

EUDAT is currently (January 2016) an EU-funded project led by CSC. The governance bodies are the General Council and the Executive Board. The project is currently in a transition phase, moving towards a sustainable partnership of centres and organisations. The latter will form an operational body that provides a few central services while most services will be distributed from the partner sites.

With regard to B2ACCESS, there is currently a service development team and an operations team. The latter is currently based at Forschungszentrum Juelich and the University of Oslo. Furthermore, there is a security team, for both operational tasks (including CSIRT) and software assessments.

2.3.5 Business Model

How is EUDAT financed?

The EUDAT project is currently funded by the European Commission.

How sustainable is EUDAT?

Currently, EUDAT is a project but there are dedicated tasks in EUDAT looking into this.

2.3.6 Underlying Technology

What protocols and technologies are used by EUDAT?

The main authentication technologies (both external and internal) are OAuth2, X.509 and SAML.

What does the EUDAT architecture look like?

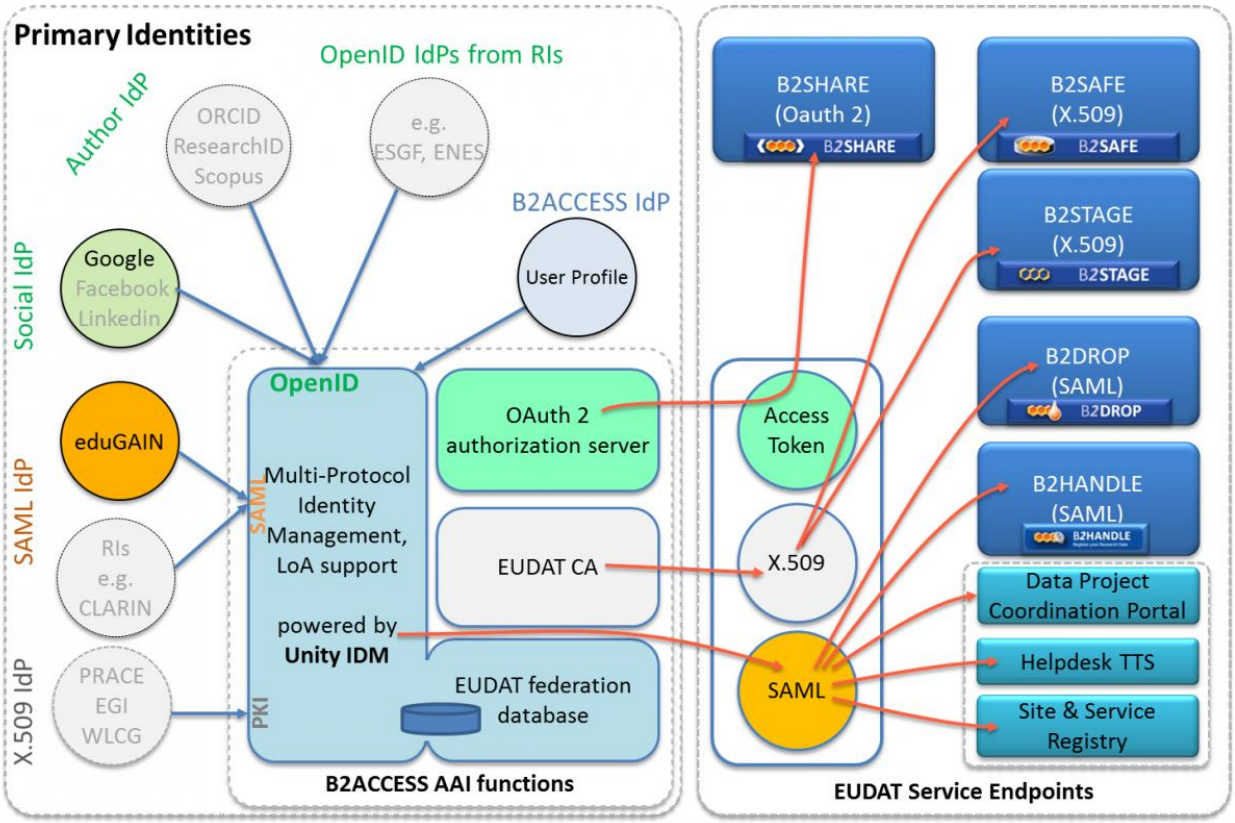

Figure 2.3: EUDAT architecture

As shown in Figure 2.3, primary Identity Providers are consumed using a range of technologies such as SAML, X.509 and OpenID. Identities and attributes, provided by these external IdPs, are mapped onto an EUDAT identity by the Unity IdM component of the B2ACCESS service. This EUDAT identity is then exposed, again using a range of technologies, to the EUDAT backend Service Providers hosting the EUDAT B2 services and internal project tools.

Any of the IdP technologies can be used to authenticate and access the backend Service Providers integrated using any of the support technologies.

What does the EUDAT login flow typically look like?

The login flow to an EUDAT service will look different depending on which authentication method is used and which service integration method is used. As can be seen in Figure 2.3, B2ACCESS supports logging in with many social Identity Providers as well as eduGAIN or X.509 certificates.

An example login flow for the SAML IdP scenario is as follows, and shown in Figure 2.4:

1. A user tries to access a protected resource in an EUDAT service – B2SHARE in this example.
2. Since the user is not authenticated, a redirect to B2ACCESS is issued.
3. Within B2ACCESS, the user can choose a home organisation IdP.
4. After the user selects an IdP, a redirect to that IdP login page is issued.
5. The user provides credentials to the IdP.
6. If the credentials are valid, the user is redirected back to the B2ACCESS service.
7. The B2ACCESS service consumes the information provided by the IdP.
8. The B2ACCESS service redirects back to the initial end point, providing the EUDAT identity and attributes over the technology used to integrate this service – OAuth2 in the case of B2SHARE.
9. The B2SHARE Service Provider consumes the information provided by B2ACCESS over the OAuth2 protocol and decides if the user is authorised or not.

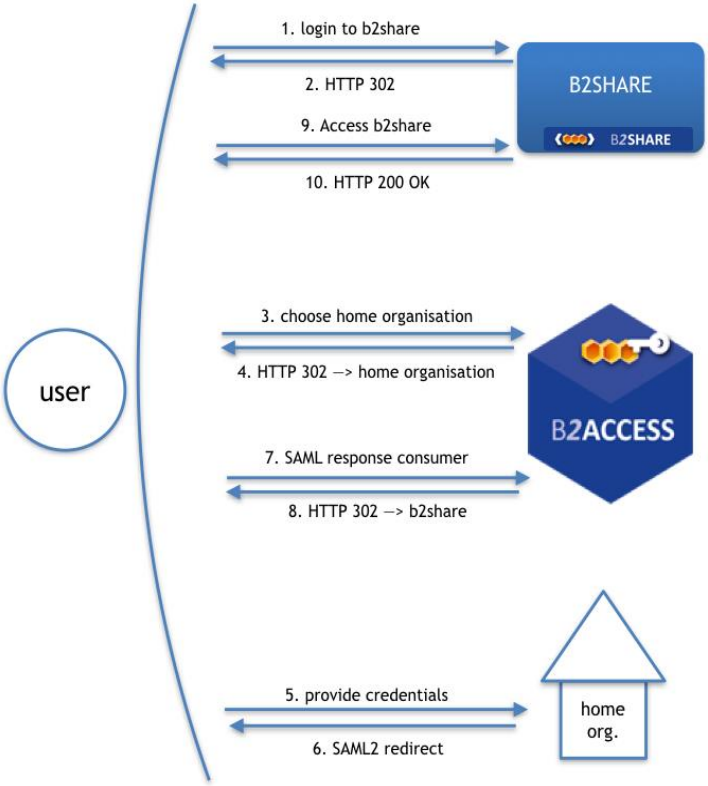

Figure 2.4: EUDAT login flow for SAML IdP scenario

2.3.7 Major Benefits

What makes EUDAT attractive and why?

- Single front end: EUDAT's B2ACCESS service provides a single front end for federated access to several EUDAT and research communities' services. In effect, it provides single sign-on to all of EUDAT.
- Multiple authentication protocols: B2ACCESS supports different authentication protocols such as SAML and X.509 certificate login and could support more in the future due to its architecture.
- IdP-of-Last-Resort included: EUDAT, as an e-infrastructure, by default allows the creation of identities without having to link them with another AAI. Therefore, EUDAT can already act as an IdP-of-Last-Resort for a research community.

2.3.8 Limitations

- Production status: The main problems at the moment are getting all the required features production ready, and fixing a large number of minor usability issues.
- Identity linking: There might be issues with linking different authentication credentials and identities to an EUDAT identity.
- Protocol translations: As a result of supporting multiple authentication protocols (web/non-web), there might be interoperability issues and limitations.

2.3.9 How to Join

How could a research community make use of EUDAT?

Currently research communities cannot bring in services to allow EUDAT users to access them. However, EUDAT is considering allowing this in the future (with a limited set of attributes released due to the privacy statement) as this has been a common request.

Therefore, EDUAT currently allows research communities to:

1. Join EUDAT as users and use EUDAT's own services.
2. Use EUDAT's B2ACCESS as own IdP. This means running own services via B2ACCESS and reusing only B2ACCESS from EUDAT.
3. Deploy EUDAT software on their own. All EUDAT software is open source.

What are the conditions to join EUDAT?

For (1), there will be conditions associated with being a user community. It helps if a research community can contribute some effort to the integration, as each community tends to have different requirements.

For (2), the research community will need to comply with certain policies, so that EUDAT can comply with the eduGAIN GÉANT Data Protection Code of Conduct.

For (3), there are no restrictions other than those specified by the software licences. Some limited manpower would need to be available if a research community wanted to take this step.

2.4 Moonshot

2.4.1 Introduction

What is Moonshot?

Moonshot[4] [Moonshot] is the name of a set of technologies that was developed by Jisc (formerly known as Janet), the UK's National Research and Education Network (NREN), in collaboration with a number of partners from around the world.

What does Moonshot do?

Moonshot aims at providing a single solution for effective management and access control for a wide range of web and non-web services. The Moonshot technology is the implementation of the IETF's Application Bridging for Federated Access Beyond web (ABFAB) [ABFAB] standards and makes use of proven and relatively widely deployed technology such as:

- EAP/RADIUS authentication, as used in eduroam [eduroam].
- SAML authorisation, as used in eduGAIN.

Does Moonshot provide anything else besides an authentication and authorisation function?

No, Moonshot only provides authentication and authorisation. For authorisation, the technology allows user attributes to be retrieved by a Moonshot Identity Provider from a SAML Identity Provider (via SAML ECP) or an LDAP directory. The attributes are then packaged in a standard SAML assertion and sent from a Moonshot IdP to a Moonshot SP.

2.4.2 History and Current Status

When was Moonshot created and by whom?

Project Moonshot was initiated by the TERENA Task Force on European Middleware Coordination and Collaboration (TF-EMC2) as part of its "Beyond Web SSO" work, and during the TERENA Networking Conference 2009 Jisc (formerly known as Janet) presented their initial work on the matter. Since then a lot of work has been done, both standardising extensions to existing protocols and developing software to implement these protocols.

From 2013 to 2015 two pilots were held for the project: a Jisc UK pilot and a GÉANT European pilot. In the latter, the participants were: Jisc, SWITCH, CARNet and SRCE, CESNET, NIIFI, RedIRIS and the University of Murcia, NORDUnet and CSC, and RENATER.

[4] Not to be confused with similarly named projects like Hewlett Packard's Moonshot System or Google[x] research projects.

October 15, 2014 marked the establishment of a pan-European Trust Router Network that enabled parties from different countries to communicate securely over this network.

As of March 2015, Jisc (formerly known as Janet) have completed and archived Project Moonshot and the UK pilot members have transitioned into production in Jisc's Assent service [Assent]. This service is primarily available to Jisc customers, but the underlying Moonshot technology can be set up by other NRENs, similar to the GÉANT pilot members, to offer an analogous service. According to Jisc, Assent is also open to research and education users outside the UK until the local NREN or GÉANT starts a Moonshot-based service. So, for example, an organisation in France could join Assent until RENATER or GÉANT start their own service, then the organisation would be encouraged to move over to that other service.

What is Moonshot's current coverage (numbers of countries, organisations and users)?

As of 2015 Moonshot is neither a global service nor an infrastructure. It is a technology that has been used in pilot infrastructures. In 2015, Jisc launched a production Moonshot service called Assent, for users in the UK.

2.4.3 Intended Audience

For whom is Moonshot operated primarily?

Moonshot is general-purpose technology that can be used in various setups and it is not limited to a specific audience. One aim of Moonshot is to unify and build on existing infrastructure, such as eduroam, and components (Identity Providers) of SAML identity federations to enable new applications for these technologies.

Are there further eligible intended audiences of Moonshot?

Although the Moonshot technology has been developed by an NREN (Jisc, UK), it could also be deployed in a commercial setting.

For whom is Moonshot suited?

For the time being, not all client-side software supports Moonshot out of the box, which means that Moonshot is most suited for organisations that have:

- Existing setup or practice to deploy custom software to all end-user computers, or
- End users capable (and willing) to install custom software themselves.

2.4.4 Operation and Governance

Who operates Moonshot?

Moonshot in itself is technology developed by Janet (now Jisc) and released as open source under the BSD licence [Moonshot-SA]. The intended Moonshot operation model is for every NREN to set up their own Moonshot-based federation (by reusing their SAML and eduroam federation as underlying layers) and then interconnect with other Moonshot federations by joining Trust Router Networks.

How is Moonshot controlled/governed?

In 2015 Jisc itself launched a Moonshot-based federation called Assent and as such is an interested party in developing the technology further.

Where is support for Moonshot available?

The main support resources are the Moonshot mailing list moonshot@geant.net and Moonshot Wiki [Moonshot-Wiki].

2.4.5 Business Model

How is Moonshot financed?

Moonshot as a technology was mostly developed by the UK organisation Janet (now Jisc). A Moonshot-based federation is intended to be run by an NREN or other central body that can assert trust between parties, much the same as in eduGAIN and eduroam, and any funding of these bodies is outside of the scope of the Moonshot project. At the time of writing there is no inter-federation governing body for the Trust Router Network, nor is there central funding, so any such work must be divided between participants in the network. Lack of central funding does not necessarily affect the sustainability of the model as most of the work in running the Moonshot federations is done by the federations themselves.

2.4.6 Underlying Technology

What protocols and technologies are used by Moonshot?

Moonshot builds upon existing and proven technologies:

- Authentication using EAP/RADIUS (the technology used for eduroam).
- Authorisation using SAML (the technology used for eduGAIN).
- Operating system security APIs (GSS-API, SSPI, SASL).

To these existing protocols Moonshot adds its own Trust Router technology, which allows automatic enabling of RADIUS trust between two parties with no previous direct setup.

What does the Moonshot architecture look like?

The Moonshot architecture can be divided into four main components:

- Client: Software on the end user's computer that is used for interfacing with the service. This software must include the necessary components to initiate a session request and provide an Identity Provider selection mechanism.
- Relying Party (RP): Consists of two parts: the Service the user is trying to access (SSH, Microsoft Exchange server, etc.) and the Relying Party Proxy, a RADIUS server that connects the Service to Identity Providers. When a new session is initiated, the Service contacts the local RP Proxy, which then uses its Trust Infrastructure to forward the authentication request to the Identity Provider.
- Identity Provider (IdP): Authoritative source of identity information. RPs and IdPs need to have set up a trust infrastructure that enables RPs to trust responses received from IdPs. The end user and IdP interact directly with each other through RADIUS secure tunnelling (EAP), which means that user credentials are never seen by any intermediate party. Upon successful authentication, the IdP responds to the RP with the success status and possibly a SAML message with attributes describing the user, such as name or membership information.
- Trust Infrastructure: Enables the RP and IdP to trust each other and is managed by the NREN. Trust Infrastructure can be either a classic hierarchical RADIUS network (such as in eduroam) or based on the Moonshot Trust Router Network. The latter enables direct communication between the first RP and IdP without intermediate proxying RADIUS servers.

Figure 2.5 below shows the basic Moonshot architecture. (Source [Moonshot-Wiki])

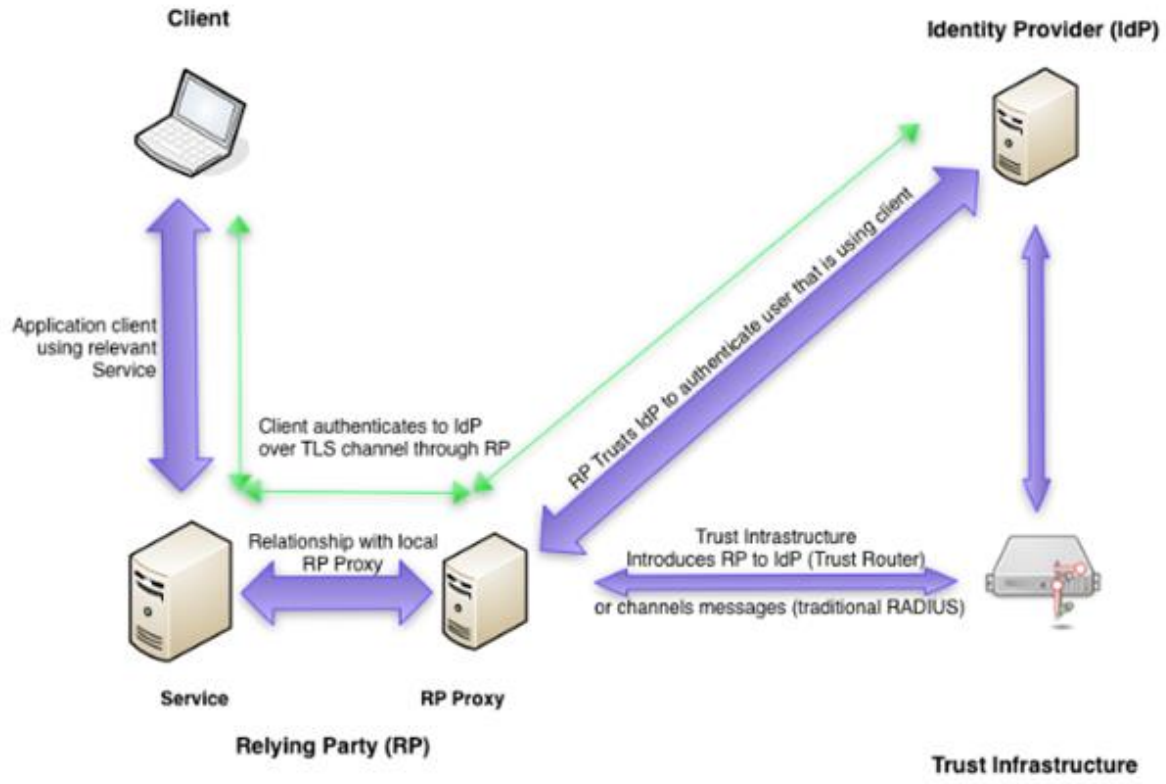

Figure 2.5: Moonshot architecture

Security Protocols

To authenticate Client to RP, Moonshot uses already existing protocols. A wide range of target applications support at least one of the three security protocols:

- Generic Security Service Application Programming Interface (GSS-API).
- Simple Authentication and Security Layer (SASL).
- Security Support Provider Interface (SSPI).

Both the Client and the Service part of the RP need to be written to support one of the above protocols.

Kerberos is a popular security protocol that acts as a mechanism for GSS-API. A lot of applications support Kerberos and therefore are compatible with GSS-API. Project Moonshot has created a GSS-EAP mechanism that enables EAP and RADIUS network to be used as authentication mechanisms for GSS-API. Any software that supports GSS-API and can be configured either directly or via negotiation to support the GSS-EAP mechanism can use Moonshot as an authentication and authorisation solution.

SSPI is conceptually similar to GSS-API and a range of Security Support Providers (SSPs) are available for applications. An EAP-SSP is needed to allow SSPI-enabled applications to use Moonshot for authentication.

Client Support

Currently several Linux distributions (Debian, Ubuntu and RHEL-based) and Windows versions are supported. Mac OS X is currently not supported. After installing GSS-EAP or EAP-SSP libraries to the client machine, many applications pick up Moonshot support out of the box with little to no configuration needed; such applications include MS Outlook, Internet Explorer and Firefox on a Windows platform and OpenSSH client and Firefox on a Linux platform. Some applications require patched versions to be Moonshot-enabled (e.g. PuTTY).

Server Support

Installing and configuring Moonshot on the server involves several steps:

1. Install and configure Moonshot libraries, configure RP and Trust Infrastructure.
2. Install and configure the service application to make use of Moonshot.

Moonshot documentation includes articles and step-by-step instructions for several platforms and service applications. Some applications support Moonshot out of the box (e.g. Microsoft Exchange Server); some require patched versions to be installed instead of vendor packages (e.g. OpenSSH server).

What does the Moonshot login flow typically look like?

A typical Moonshot login flow can be divided into three phases:

1. Negotiating a secure tunnel from the Client through the RP to the Identity Provider.
2. Client authenticating itself to the IdP.
3. Service authorising the Client based on the response received from the IdP.

The following is a detailed description of the steps involved to authenticate a Client, for example, to an OpenSSH server using GSS-EAP and EAP-TTLS protocols, adapted from the Architecture and Protocol Flows page of the Moonshot Wiki [Moonshot-A]. When different authentication mechanisms are used, the actual login flow can differ slightly, especially regarding the steps taken on the Client side (e.g. prompting for credentials vs using PKI certificates).

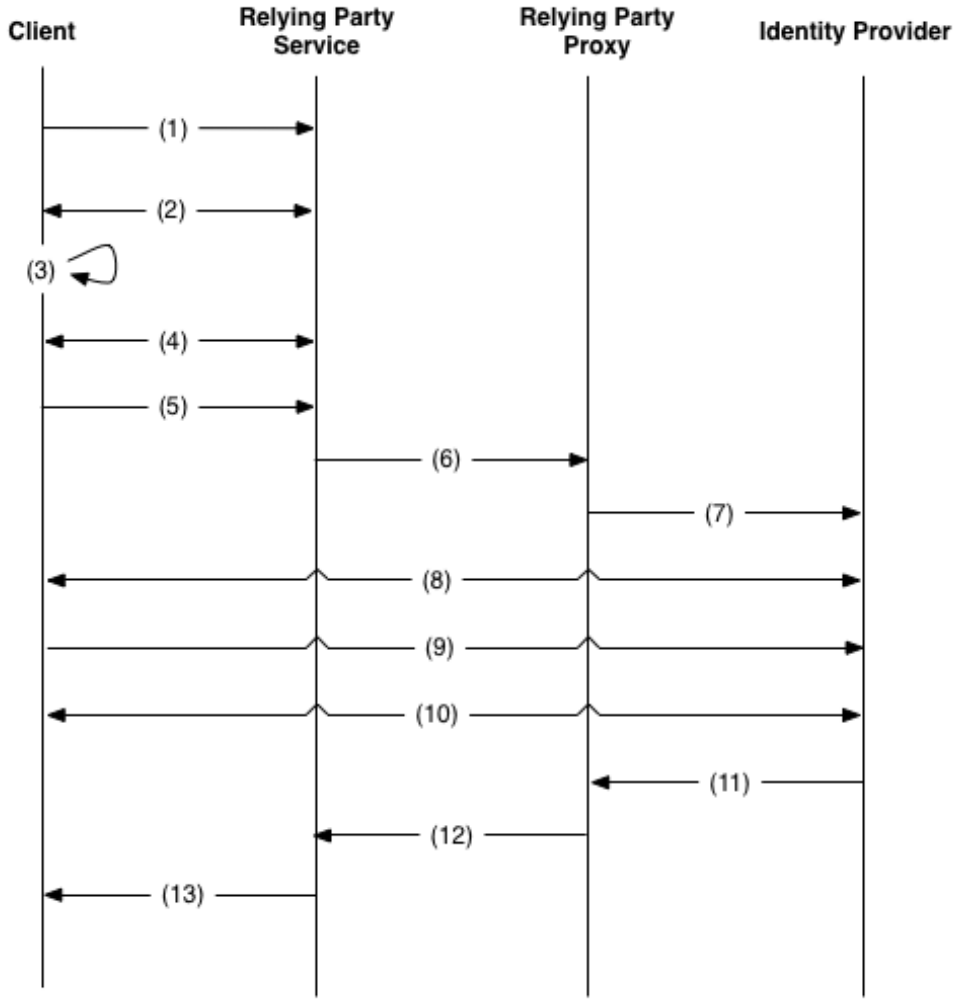

Figure 2.6: Typical Moonshot login flow (Source [Moonshot-A])

1. The application client attempts to connect to the OpenSSH server.
2. The application client and application server negotiate the use of the GSS-API for authentication, and GSS-EAP is called on the client device.
3. GSS-EAP on the client prompts for credentials to use on this particular service. This will result in an identity to use.
4. GSS-EAP is negotiated as the GSS-API mechanism to use for this authentication.
5. GSS-EAP on the application client creates an EAP request containing the anonymous version of the identity (i.e. simply @REALM). This EAP message is sent over the established GSS channel to the application server.
6. GSS-EAP on the application server opens a RadSec/RADIUS connection to its configured RP Proxy. A RADIUS Access-Request message, with the EAP message it received from the client and optional SAML authentication request encapsulated, is sent down this channel.
7. The RP Proxy receives this request and, using Trust Infrastructure, establishes a secure connection to the IdP of this realm and forwards the RADIUS request. It will also include a GSS name that it claims to be.

At this point, the RP Proxy and IdP now have a secure tunnel established between themselves, and there is a path between the client and its IdP (consisting of GSS-EAP between the application client and the application server, RadSec/RADIUS between the application server and its RP Proxy, and RadSec/RADIUS between the RP Proxy and the IdP.

8. The client and IdP now use the established secure path and choose the EAP-TTLS authentication method, which creates a secure inner tunnel between the client and IdP so that the two can communicate securely in such a way that the intermediate RP cannot see the traffic (it just passes encrypted data back and forth).
9. The client and IdP negotiate an inner EAP method (e.g. PAP), which in turn is used to verify the credentials being used. Additionally, the client sends the GSS name of the RP that it is trying to connect to.
10. The IdP authenticates the client credentials and also uses received GSS names to verify that both client and IdP talk to the same RP. Assuming all is correct, the EAP session is finalised, resulting in a set of keys shared by the client and the IdP.

At this point the Client is successfully authenticated at the IdP, but not yet authorised to use the service.

11. The IdP responds to the RP Proxy by sending it a RADIUS Access-Accept message, which contains an encapsulated EAP success message and possibly a SAML assertion.
12. The RP Proxy decides whether this authenticated user should be allowed to use the service, based on local policy and/or information provided by the IdP in the form of RADIUS or SAML attribute information. If the decision is positive, then the RP Proxy forwards the Access-Accept and related information to the application server.
13. GSS-EAP on the application server verifies the EAP keys and then consults local policies, possibly informed by provided RADIUS or SAML attribute information, as to whether this user should be allowed to use the service. In the event of a successful decision, a new session is established.

2.4.7 Major Benefits

- Web/non-web authentication: The main benefit of Moonshot is extending the Single Sign-On (SSO) benefits, known from web, to non-web applications and services. This reduces both administrative and end-user effort in providing and using different services. As an example, Moonshot could be used to authenticate and authorise SSH or email access using the home organisation credentials similar to authenticating WiFi access in eduroam. However, as of now, the existing Moonshot infrastructures and eduroam are decoupled and according to statements from the eduroam operators there are no plans to add Moonshot profiles to eduroam on a global level.
- Standards protocols: Moonshot joins many different, but widely deployed, protocols (EAP, SAML, RADIUS, etc.) into a single solution that offers both authentication and authorisation for web and non-web applications. Making use of popular APIs such as GSS-API and SSPI, Moonshot is theoretically automatically supported by a wide range of client and server applications.
- Simple IdP discovery: Moonshot provides an easy way for users to discover their Identity Provider by having the user configure his IdP in an identity selector (on every client). IdP

discovery is a challenge that other AAIs (such as eduGAIN) with an increasing number of IdPs sometimes struggle with.

2.4.8 Limitations

- Complex deployment: Setting up the Moonshot federation and trust infrastructure can be a lot of work, but, once it is up and running, adding new applications to the system is relatively low cost.
- Missing components on client: At the moment, neither the Moonshot client nor server libraries come pre-installed on any supported platform. Getting the software installed and configured is a manageable workload with regard to servers, but might prove a higher than normal entry barrier when end users are expected to install the client libraries onto their computers. This problem is somewhat lessened in smaller organisations, where system administrators can help users, or in enterprise-level organisations, where automated provisioning is available.
- Data privacy issues: As with any inter-organisation federation where potential personal data is involved, strict legal measures are required to conform to data protection laws. In most cases the NREN providing Moonshot federation takes on the role of organising the required contracts and trust with each party.
- Coverage: Even though Moonshot uses the same technologies as eduroam and eduGAIN, it does not generally rely on these infrastructures, and (at least in the pilot phase) has been separated from them.

2.4.9 How to Join

How could a research community make use of Moonshot?

Moonshot is technology to build an authentication and authorisation federation and by itself is not a ready service. Instead, an NREN or similar organisation is expected to set up a service using this technology. Jisc in the UK has set up a service called Assent and several other NRENs, including CSC, SWITCH, CARNet, CESNET, NIIFI, RedIRIS, NORDUnet and RENATER, were or are piloting Moonshot-based services.

What are the conditions to join a Moonshot service?

The conditions depend on the Moonshot federation/service. At the time of writing this document, there is only one service, the UK Assent service, where Moonshot-based services can be added.

2.5 STORK

2.5.1 Introduction

What is STORK?

STORK (Secure idenTity acrOss boRders linKed) [STORK] "is a platform which allows people to use their national electronic ID to establish new e-relations with foreign electronic services, which may be operated by public or private service providers". STORK 2.0 was the name of an EU-funded project that extended "the STORK platform by allowing legal persons (such as companies) to be represented by natural persons" and ended in September 2015. (Source [STORK-FAQs1])

STORK 2.0 "will be a step forward towards the creation of a fully operational framework and infrastructure for electronic identities and authentication in the EU" (done in STORK), focusing on strategic e-learning and academic qualifications, e-banking, public services for business and e-health areas. (Source [STORK2-About])

What does STORK do?

The STORK platform interconnects national infrastructures and allows national electronic identities to identify users towards any services that use STORK.

The platform allows people to use their national electronic ID to establish new e-relations with foreign electronic services, which may be operated by public or private Service Providers.

The objective when creating STORK was to define a framework that does not change existing national electronic identification (eID) infrastructure, but defines an interoperability layer on top of national systems that supports cross-border eID federation.

Does STORK provide anything else besides an authentication and authorisation function?

"No. STORK only interconnects national infrastructures and allows you to use national electronic identities to identify yourself towards any services that have chosen to use STORK." (Source [STORK-FAQs2]).

The STORK 2.0 project offered common specifications, standards and building blocks, such as the specification of a business attributes set or a QAA model. (Source [STORK2-About])

2.5.2 History and Current Status

When was STORK created and by whom?

The STORK 2.0 project started in 2012 as the follow-up project of STORK [STORK], implemented from 2008 to 2011. The project finished in September 2015. From that date, the project remains in maintenance status, managed within the e-SENS project [eSENS].

What is STORK's current coverage (numbers of countries, organisations and users)?

The STORK 2.0 project consortium consists of 58 participants. It directly involves 19 EU Member States / Associated countries[5]. The consortium members include national authorities, non-profit organisations, private companies and academic partners. A full list of participants is available at [STORK2]. Participation in STORK was voluntary for EU Member States.

2.5.3 Intended Audience

For whom is STORK operated primarily?

The STORK platform aims at enabling European citizens to use their national electronic identities (eIDs) in any of the 19 EU/EEA STORK Member States. The STORK project's "purpose was to gain practical experience in real applications to see where issues on cross-border electronic identification arise, and to explore how they can be solved. This experience was also helpful in the ongoing preparation of a European Regulation on electronic identification and trust services (eIDAS) [eIDAS-About; eIDAS-Regulation-1501; eIDAS-Regulation-1502]. It is expected that this Regulation will lead to a shared responsibility model, where Member States remain responsible for their electronic identity system, and where the European Commission is responsible for coordinating the efforts of setting the standards that Member States have to fulfil to maintain high levels of security and data protection." (Source [STORK-FAQs20])

In the context of research and education communities, most relevant is that the STORK 2.0 project included an eLearning & Academic Qualifications work package to "provide a set of academic services that can be used by citizens, government and companies. These cross-border academic services, involving the exchange of identity attributes, facilitated the use of academic information by users, government and private organizations, while also attempting to demonstrate the possibility of integrating other existing services concerned with the verification of certificates acquired in training courses other than those given by the Academia." (Source [STORK2-WPO]) Information about the pilot related to this work package is available from [STORK2-eLAQP].

Are there further eligible intended audiences of STORK?

Virtually any type of Service Providers may be interested in this infrastructure (government SPs, commercial SPs and SPs from research). The STORK 2.0 pilot also included projects in the areas of:

[5] Austria, Belgium, Czech Republic, Estonia, France, Greece, Iceland, Italy, Lithuania, Luxembourg, Netherlands, Portugal, Slovenia, Slovakia, Spain, Sweden, Switzerland, Turkey and the United Kingdom

- e-banking: Establishing cross-border online e-banking services using eID.
- Public Services for Business: Extending existing online Public Services for Businesses beyond national borders using eID management.
- E-health: Enabling easy and secure cross-border access to patients' data using eID authentication.

(Source [STORK2-Pilots])

For whom is STORK suited?

In the context of research and education, as mentioned above, the STORK 2.0 project included an eLearning and Academic Qualifications work package and pilot to facilitate cross-border academic services. The pilot invited "students, former students and staff from universities involved in the pilot to test the STORK 2.0 integrated services". (Source [STORK2-Pilots])

The STORK 2.0 project ended in September 2015. Nonetheless, the STORK platform is still maintained and kept alive by the STORK Member States and the eSENS project, although there is no successor project to further develop or extend it. The platform has been tested in several pilots [STORK2-Pilots] and has been production ready since Q3 2015. Therefore, for research and higher education, STORK could at most be a supplier of high-quality identity data, e.g. by integrating the STORK platform in a step-up authentication service that improves the level of assurance of users by making them authenticate via STORK.

2.5.4 Operation and Governance

Who operates STORK?

The STORK 2.0 project partners were 58 organisations from the 19 Member States [STORK2-Partners]. They included government organisations and universities as well as commercial partners.

How is it STORK controlled/governed?

The STORK platform is governed by the EU/EEA Member States, through the Member State Council where each member state has one representative.

2.5.5 Business Model

How is STORK financed?

The STORK 2.0 project ended on 30 September 2015. The costs for maintaining the existing infrastructure will be met by Member States.

How sustainable is STORK?

If the Member States continued to develop and establish the STORK platform on a large scale so that the benefits exceeded the costs, they would have a strong interest in financing the STORK model in the long term. There are some indicators as of Q4 2015 that maintenance of the STORK platform will be continued among the Member States.

2.5.6 Underlying Technology

What protocols and technologies are used by STORK?

The STORK platform uses SAML2, or rather a specific $SAML2_{stork}$ profile that extends SAML. $SAML2_{stork}$ is by default not interoperable with $SAML2_{int}$ used by eduGAIN [eduGAIN] but proof of concepts, documented in *Scholar European Electronic Identity Federation* [SEEIF], have shown that the two AAIs can, to some extent, be technically connected by a translator/proxy service.

What does the STORK architecture look like?

The STORK architecture consists of:

- Identity Provider (IdP): Authenticates users and issues SAML assertions about the user containing user attributes.
- Service Provider (SP): Consumes SAML assertions to perform access control and make user attributes available to the (web) application that it protects.
- Attribute Provider (AP): Provides additional user attributes using a previous authentication process.
- Pan European Proxy Service (PEPS): Central point of interconnection and trust for each country and protocol gateway for countries that use different technologies for their national eIDs. Also acts as an intermediary for foreign eIDs towards its domestic SPs. Each country has one PEPS.
- S-PEPS: Source PEPS.
- C-PEPS: Citizen PEPS.

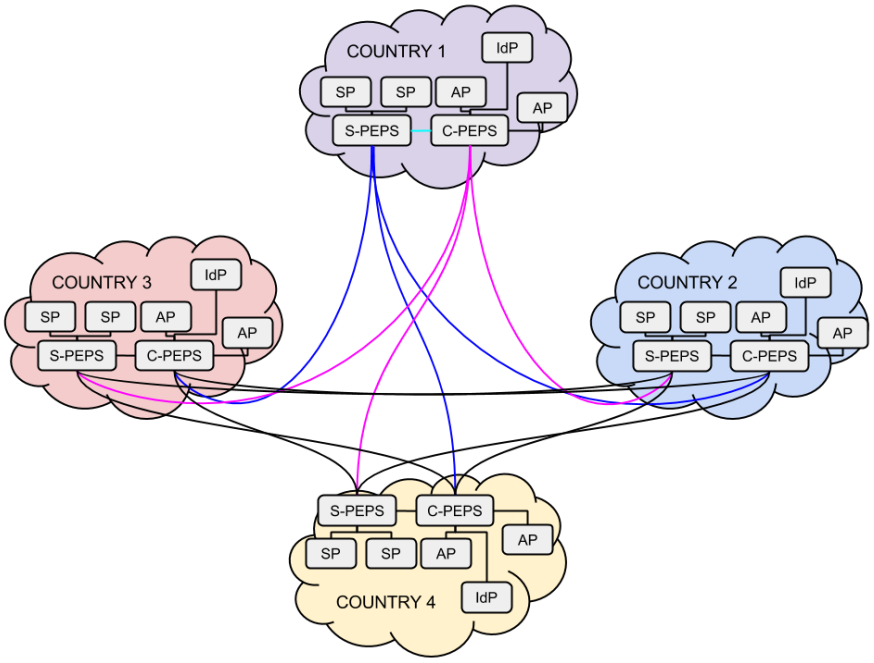

Figure 2.7: STORK architecture

What does the STORK login flow typically look like?

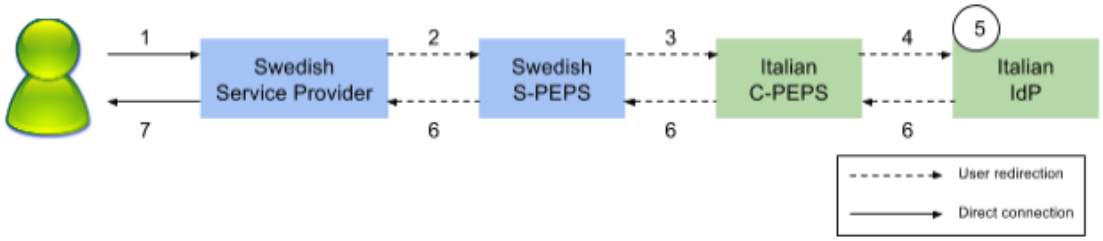

Figure 2.8: STORK login flow

1. The user asks for a service that depends on the Swedish STORK platform.
2. The user is required to authenticate himself. This triggers the STORK authentication process, so he is redirected to the Swedish S-PEPS, indicating his country of origin.
3. The Swedish S-PEPS redirects the user to his C-PEPS (Italian PEPS).
4. The Italian C-PEPS shows the user the attributes that are going to be obtained from the IdP. If the user validates the action, he is redirected to the IdP.
5. At the IdP, the user is authenticated and has to consent to share the requested attributes.
6. The user is redirected to the Service Provider through his C-PEPS (Italian PEPS) and the Swedish S-PEPS.
7. The Service Provider finally grants the user access to the specific service.

2.5.7 Major Benefits

What makes STORK attractive and why?

- Fewer data privacy issues: The STORK platform enforces user consent in line with local legislation and this is by default part of STORK.
- High level of assurance: STORK provides high-quality identities. The user data delivered by STORK is guaranteed to be the official identity data of a user. The user attributes are managed by the government/administration, therefore they are by default very reliable.
- Levels of assurance: STORK includes a concept called Attribute Quality Authentication Assurance (AQAA) levels that were used in the pilots to express the quality of different Attribute Providers (APs).
- Large user group: In theory, all citizens (including those outside higher education and research) of the Member States could use the STORK infrastructure to authenticate in any of the participating countries, provided they have a national electronic identity.
- Reuse of existing digital identity: A STORK identity is not limited to use in higher education and research but could be used for many more use cases. It could be a truly universal identity that would also allow official documents to be digitally signed.

2.5.8 Limitations

- Country coverage: Not all EU/EEA countries participated in the STORK project, partly due to varying national ID card systems and partly because participation in STORK was voluntary for Member States. With the introduction of the eIDAS Regulation (EU) No. 910/2014 [eIDAS Regulation], the results of the STORK project might be used in a Europe-wide context, as the eIDAS Regulation created a legal basis for enforcing national eIDs in other EU countries (as was the case for digital signature).
- EU citizens only: STORK identities might only be available to users of EU/EEA Member States. However, research and education projects often contain participants outside this geographic area.
- Future of STORK: The STORK 2.0 project has ended. As of Q4 2015, the STORK platform's future is unclear.
- Stork is web-based only, and cannot be accessed by non-web applications.
- "Guest" logins are not supported. All identities must be authentic.
- Being issued with and learning how to use a digital STORK identity might involve considerable effort due to the high level of assurance requirements.
- Technical difficulties: During the pilot phase some technical problems were experienced when using mobile devices on a mobile network due to the large number of attributes that were requested by an SP. This might also have been caused by the SAML redirects on the client device.

2.5.9 How to Join

How could a research community make use of STORK?

The STORK 2.0 project ended in September 2015. Service operators wanting to make use of STORK in one of the STORK-enabled Member States should contact the STORK authority in that country. For a list of partners, see [STORK-Partners].

What are the conditions to join STORK?

Joining conditions are no longer applicable because the STORK 2.0 project and its pilots finished in 2015. There is no direct successor project of STORK 2.0.

2.6 Summary Comparison Table

Table 2.1 below gives a condensed overview of the infrastructures, services and technologies described in the preceding sections. The data was gathered in October/November 2015.

	eduGAIN	EGI	EUDAT	Moonshot	STORK
Type	Service	e-Infrastructure	e-Infrastructure	Technology	(Pilot) Service and infrastructure
Main Protocols Used	SAML2	X.509	OAuth2, X.509, SAML	RADIUS, EAP, SAML2	$SAML2_{STORK}$ (extension of SAML2)
Architecture Type	Distributed	Distributed	Centralised	Distributed	Distributed
Production Level	In production since April 2011	EGI.eu has been in production (under this name) since May 2010 but it has its roots in EGEE, which started in 2004	In production since November 2015	Pilot finished; one federation (Assent, UK)	Pilot finished; future unclear
Main User Group	University and school lecturers, students, researchers, libraries	Researchers using high-throughput computing and cloud/high-performance computing community. Very important is the HEP community, for example.	Wide range of research communities [EUDAT-Communities], individual researchers and data centres	During pilot, mostly researchers from photon/neutron community	Government but also education. Citizens.

	eduGAIN	EGI	EUDAT	Moonshot	STORK
Drivers	National Research and Education Networks (NRENs) via GÉANT project	Distributed federated resource provisioning, supporting international distributed collaborations	EUDAT project and collaboration partners from different communities. Some are interested in also using B2ACCESS for user authentication of their RI services.	NRENs, mostly Jisc, UK	STORK 2.0 project and governments of STORK Member States
Estimated Number of Services/Resources	Around 1,100	About 5,000 services in EGI	Currently, EUDAT's own services (and service end points) ~50	A few dozens	19 countries involved. 58 participants (pilot services)
Estimated Number of Potential (End) Users	More than 30 million	Currently about 20,000 users have a certificate, but EGI wants to support everyone who does "relevant research"	In theory the same number as eduGAIN, plus some more due to support of other external identities	A few hundred thousand	About 300+ pilot users in STORK 2.0 project [STORK2-Pres]. All citizens with an electronic (government) identity from STORK member countries are potential users.
Financed	By GÉANT, NRENs, universities, research institutions	The EUGridPMA coordination activities are partially funded by EGI, but certification authorities are mostly operated and supported by the National Grid Initiatives and EIROs, the national funds	EU-funded via the EUDAT project; the service is intended to be sustained	Jisc (UK NREN)	Governments of Member States (STORK project till September 2015)

	eduGAIN	EGI	EUDAT	Moonshot	STORK
Support for Delegation*	Hardly	Yes, with proxy certificates	Yes: currently with OAuth2 tokens; in future also via agents using X.509 robot certificates	No	No
Participating Countries/Federations (October 2015)	35 [eduGAIN-Status]	54 [EGI-Numbers]	n/a Indirectly Germany (DFN-AAI) and countries/organisations participating in eduGAIN	1 (8 countries in pilot)	19 [STORK2]
Support for Level of Assurance (LoA)	Technically yes (eduPersonAssurance attribute), but it is not deployed at IdPs. Lacking standard.	A level of assurance is associated to every CA. In practice, all the widely used CAs have the same LoA, but IGTF is pushing for the Identifier-Only Trust Assurance (IOTA) profile.	Yes, depending on the method of authentication by which an LoA attribute is associated with the user's identity	Same as for eduGAIN	Yes (only high-quality government IDs)
Available Attributes	Recommended attributes: name, email, unique/persistent identifier, affiliation, organisation and type	name, email, institution The certificate subject is a UID within the federation of CAs.	name, email, organisational affiliation, level of assurance, UID, group membership	UID but technology also foresees integration of SAML attributes from a SAML IdP or LDAP server	Personal ID attributes (name, address, date of birth, email) and business attributes (educational and legal)
Web-Login Support	Yes	Yes	Yes	Yes	Yes

	eduGAIN	EGI	EUDAT	Moonshot	STORK
Non-Web Login Support	Hardly (SAML ECP profile, but it is not widely deployed, though it might increase with Shibboleth IdP v3)	Yes	Yes	Yes	No

Table 2.1: Summary of AA infrastructures, services and technologies

* Delegation: when one service in the name of the user can access another service (e.g. a user authenticates at a portal with AAI, then the portal, in the name of the user, accesses another service to query information with the user's credentials)

2.7 Relationship Between Existing AAIs

2.7.1 e-Infrastructure, Service or Technology

The general-purpose AAIs described in the preceding sections are either part of e-infrastructures, or they are services or technologies as shown in Figure 2.9 below. e-Infrastructures and service rely on technologies. They use a limited set of technologies that are shared by many of the described AAIs.

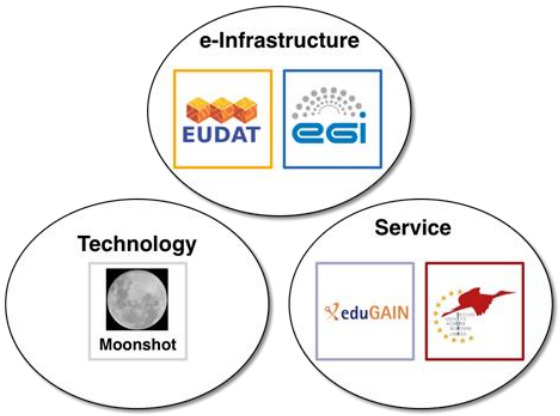

Figure 2.9: AAIs: e-infrastructure, service or technology

As Figure 2.10 below illustrates, SAML2 is one of the core technologies that is used and supported by almost all AAIs – with the caveat, however, that STORK uses a profile ($SAML2_{stork}$) that is generally not compatible out of the box with other SAML2 implementations.

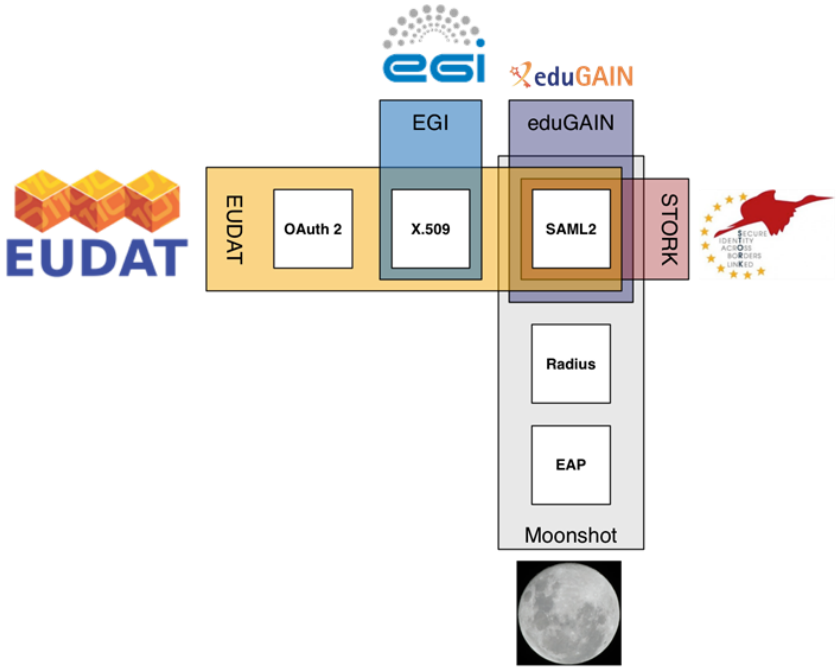

Figure 2.10: AAI core technologies

2.7.2 Major Gaps

As shown in Figure 2.10 above, the general-purpose AAIs described in the preceding sections use different technologies to perform authentication and authorisation. The technologies are SAML2, X.509 certificates, OAuth2/OpenID Connect, RADIUS and EAP. By default they are not compatible with each other. They use their own terminology, which is often not straightforward to translate to the terminology of another technology.

Even if two AAIs use the same technologies it is not guaranteed that they can easily work together because:

- AAIs were created by communities with an interest in creating and shaping their particular AAI the way it is today. This has resulted in isolated policies and workflows which make it difficult to interoperate or connect with other AAIs.
- Sometimes different profiles within a protocol hinder interoperability. For example, "supports SAML2" can mean different things because the SAML2 standard is very broad; often a product that "supports SAML" implements only a fraction of SAML and even then uses only a particular profile of that subset. For example, both eduGAIN and the STORK platform rely on SAML2 but they are not interoperable out of the box because STORK created a $SAML2_{stork}$ profile, a superset of SAML2, and uses a different architecture from eduGAIN, which uses the $SAML2_{int}$ profile, a subset of SAML2. These interoperability issues can be overcome with gateway elements, but such approaches are never perfect and usually involve some loss of trust information or flexibility.

2.7.3 Known Collaborations

In spite of the technology, terminology and policy/workflow gaps described above, there are cases where AAIs are interconnected and collaborate. Often the interconnection happens not on the level of the AAI operations but instead individual e-infrastructure providers decide to rely on different AAIs. EUDAT, for example, due to its architecture and design principles, supports external identities from eduGAIN or X.509 certificates but also from social networks. Another example is the AARC project [AARC], which is a collaboration between eduGAIN (GÉANT), EGI and EUDAT (Juelich, KIT).

An alternative approach for bringing AAIs together is to use technology/token translator services that convert (usually with some loss of information) authentication and authorisation information from one technology to another. One example here could be CERN, which relies heavily on X.509 identities (as used in EGI) but also operates as an Identity Provider in eduGAIN, which allows their users to authenticate at eduGAIN services by authenticating at the CERN IdP with their X.509 credentials. Therefore, the CERN IdP acts as technology translator.

2.7.4 Opportunities for Additional Collaborations

As was stated by *Advancing Technologies and Federating Communities* [AAA], every AAI has some advantages over another. There is no AAI that meets all requirements and could therefore be adopted universally by all research communities. Technology translators and hubs help to connect communities from a technology point of view, but it is not only the technology that has to be taken into account when connecting AAIs; policies and workflows have to be considered, too. To make this

a bit easier there are, for example, efforts to make the eduGAIN declaration/constitution technology independent and agnostic, which might lead to the incorporation of technologies other than SAML2.

EGI and eduGAIN

EGI relies on X.509 certificates, eduGAIN relies on SAML. From a technology point of view it is straightforward to let users authenticate with their X.509 certificate at a SAML Identity Provider, which then issues a SAML assertion based on the certificate. The converse, issuing X.509 certificates based on a SAML assertion, is also possible. Both approaches have been implemented and used in production. Merging these two technologies on a technology level works.

STORK and eduGAIN

STORK and eduGAIN both use SAML2 as their core technologies but use slightly different SAML2 profiles ($SAML2_{stork}$ and $SAML2_{int}$ respectively). However, as was demonstrated in *Scholar European Electronic Identity Federation* [SEEIF] and in relation to some work done in Task 1 Harmonisation of Service Activity 5 Trust and Identity Service Development of the GN4-1 project (SA5 T1), STORK and eduGAIN could be connected. The high-quality STORK identity data could, for example, be used to improve eduGAIN user data in a step-up-authentication service. The converse, using an eduGAIN identity to access a STORK service, would be of limited usefulness, if any at all. It would only be useful for those services whose target group is users belonging to the higher education community and if the level of assurance requirements of the services are lower than for a generic STORK identity.

Moonshot and SAML-based AAIs

Moonshot foresees retrieving user attributes from an LDAP directory or a SAML Identity Provider (via SAML ECP). Therefore, an obvious opportunity to connect the Moonshot technology with the SAML technology is to make a Moonshot Identity Provider retrieve the user attributes from a SAML Identity Provider, or at least to retrieve user information from the same source as the SAML Identity Provider. That way, an organisation needs to manage attributes in one place only. Ideally, a Moonshot federation would use the same policies with regard to attribute names and format as the SAML federation of which the organisation is probably already a part.

EUDAT as AAI Interconnect

EUDAT could be the glue that connects all AAIs and technologies. This should be possible (with some limitations for non-web applications) because EUDAT's B2ACCESS by design supports X.509, SAML and the upcoming OpenID Connect technologies, which are used by the other AAIs. It could also support Moonshot (as a supplier of external identities) relatively easily. Thus, creating an EUDAT identity via one of these technologies would easily be possible. The converse, using an EUDAT identity to access eduGAIN services, is also possible. EUDAT is therefore a technology translator service to some extent. One of the main challenges for EUDAT – and for similar AAIs, such as the upcoming INDIGO AAI – is to make the linking of identities and technologies seamless and easy from a user's point of view.

3 Upcoming AAIs

This section describes AAIs that are not yet in production but are currently in development. For the current iteration of the document, the only AAI included is INDIGO-DataCloud.

3.1 INDIGO-DataCloud

3.1.1 Introduction

What is INDIGO-DataCloud?

INDIGO-DataCloud (**IN**tegrating **D**istributed data **I**nfrastructures for **G**lobal Expl**O**itation) [INDIGO] is a project approved in January 2015 within the EINFRA-1-2014 call [EINFRA-1-2014] of the Horizon 2020 EU Framework Programme [Horizon2020]. It aims at developing a data/computing platform targeting scientific communities, which can be provisioned over hybrid (private or public) e-infrastructures.

As part of INDIGO-DataCloud, the INDIGO-AAI infrastructure is being developed by a collaboration of all concerned work packages.

What will INDIGO-DataCloud do?

INDIGO-DataCloud will develop and deliver software components allowing execution of applications on cloud and grid-based infrastructures, as well as on high-performance computing (HPC) clusters. The project will extend existing Platform as a Service (PaaS) solutions, allowing public and private e-infrastructures, including those provided by EGI, EUDAT, PRACE and Helix Nebula, to integrate their existing services and make them available through AAI services compliant with GÉANT's interfederation policies, thus guaranteeing transparency and trust in the provisioning of such services.

Will INDIGO-Datacloud provide anything else besides an authentication and authorisation function?

INDIGO-Datacloud will provide an integrated set of middleware components allowing the execution of applications on cloud and grid-based infrastructures as well as on HPC clusters.

3.1.2 History and Current Status

The INDIGO-DataCloud project started in April 2015 and is currently working on the development and integration of the components and services that will be part of the first INDIGO software release, currently planned for July 2016.

The latest information about progress and status is available from the project news feed [INDIGO-News].

3.1.3 Intended Audience

For whom is INDIGO-DataCloud operated primarily?

INDIGO-DataCloud is a platform aimed at supporting easy exploitation of computing and data resources for scientific communities. It is currently under development, and thus not yet deployed in production, but is targeted at the main scientific computing e-infrastructures deployed worldwide (EGI, WLHCG, OSG, etc.) supporting scientific research communities.

Are there further eligible intended audiences of INDIGO-DataCloud?

Given the goal of interoperability with EGI, EUDAT, PRACE and Helix Nebula, any of their audiences might also be an audience for the INDIGO-DataCloud.

For whom is INDIGO-DataCloud suited?

INDIGO-DataCloud will be suited for scientific communities. The goal is to have a low learning curve, for both inexperienced end users and developers. The project intends to support popular existing software suites, such as ROOT, Octave/MATLAB, Mathematica or R-Studio, in a transparent way.

3.1.4 Operation and Governance

Who is working on INDIGO-DataCloud?

The INDIGO-DataCloud Consortium is composed of 26 European partners [INDIGO-Partners].

How is INDIGO-DataCloud controlled/governed?

The project coordinator, INFN [INFN], has extensive experience in coordinating large consortiums, especially regarding middleware development and support to user communities. INFN therefore leads several key technical areas, notably the PaaS definition and development activities. The deputy coordination responsibility lies with CSIC [CSIC], a multidisciplinary institution well suited to effectively coordinating the process of requirements gathering from the user communities, providing

feedback to the development process, and ensuring an adequate strategy for knowledge management.

Where is current information about INDIGO-DataCloud available?

Information about the process at INDIGO-DataCloud can be found on the news feed at [INDIGO-News].

The AAI design document is available at [INDIGO-A].

3.1.5 Business Model

How is INDIGO-DataCloud financed?

Development of the software provided within INDIGO-DataCloud is funded within the Horizon 2020 framework. The core AAI partners INFN and KIT have a longer-term interest and funding to support the software sustainably.

3.1.6 Underlying Technology

What protocols and technologies are used by INDIGO-Data-Cloud?

The INDIGO AAI stack is based on the following authentication and authorisation technologies:

- **Authentication and Identity**: INDIGO supports SAML, X.509 and OpenID Connect (OIDC) user authentication. Identity information collected through these authentication mechanisms is exposed to INDIGO services using the OpenID Connect protocol. The main advantages of this approach are the ability to onboard users from existing interfederations (e.g. EduGAIN), high-throughput computing (HTC) infrastructures (e.g. EGI, WLHCG) and social identity providers (e.g. Google, GitHub), and the low integration friction for the relying INDIGO and external services that is provided by the OpenID Connect standard.

- **Authorisation and Delegation**: INDIGO relies on OAuth2 to implement authorisation and delegation of privileges across INDIGO services. Attribute-based authorisation is implemented at the services, leveraging identity information provided by the OpenID Connect identity layer.

What does the INDIGO-DataCloud architecture look like?

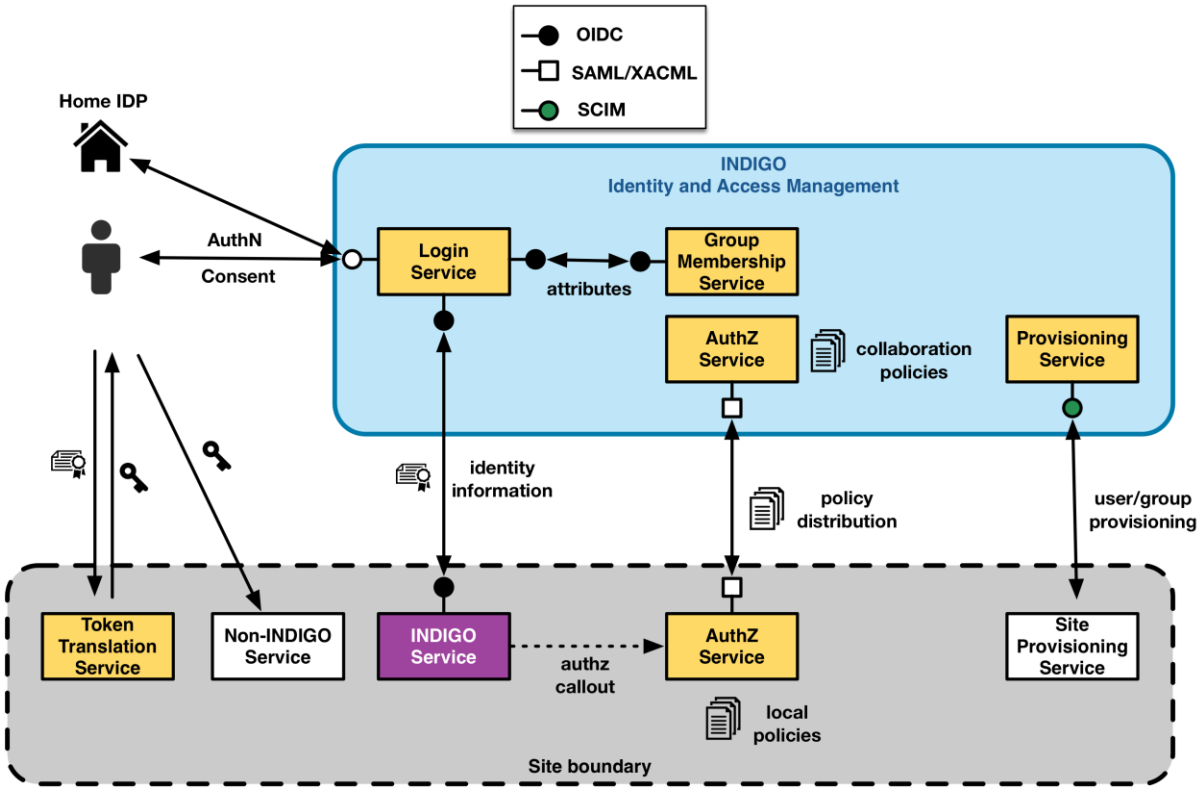

Figure 3.1: INDIGO Architecture

Figure 3.1 shows the main components of the INDIGO AAI architecture. INDIGO provides a set of services (collectively named the Identity and Access Management Service (IAM)) that deal with user authentication (the Login Service (LS)), group membership and identity attributes (the Group Membership Service), provisioning (the Provisioning Service) and the management, distribution and enforcement of authorisation policies through the authorisation service component (the Authorisation Service).

The LS translates the identity information obtained from the external authentication mechanism and provides this information, together with attribute information gathered from the Group Membership Service, to relying parties via OpenID Connect (OIDC) flows and end points. The LS currently supports SAML, OIDC and X.509 authentication.

INDIGO services act as OIDC relying parties (RPs) for the LS, and trust is established through an RP registration process following the OIDC dynamic client registration standard specification.

Authorisation is based on the aggregated claims collected from the LS via the standard OIDC methods of providing identity claims to relying parties (i.e. id_token and userinfo end points). These will include attributes expressing membership in Virtual Organisations, as well as attributes deriving from the external authentication step (e.g. released by the user home-IdP or collected from upstream attribute authorities).

Non-OIDC services are accessed through a credential translation step, implemented by the Token Translation Service (TTS).

Resource provisioning and de-provisioning can be implemented by relying parties leveraging the System for Cross-domain Identity Management (SCIM) provisioning end points provided by the IAM provisioning service.

Finally, consistent and flexible authorisation is provided by the distributed Authorisation Service, which provides policy definition, distribution, composition and enforcement functionality across the services in the infrastructure.

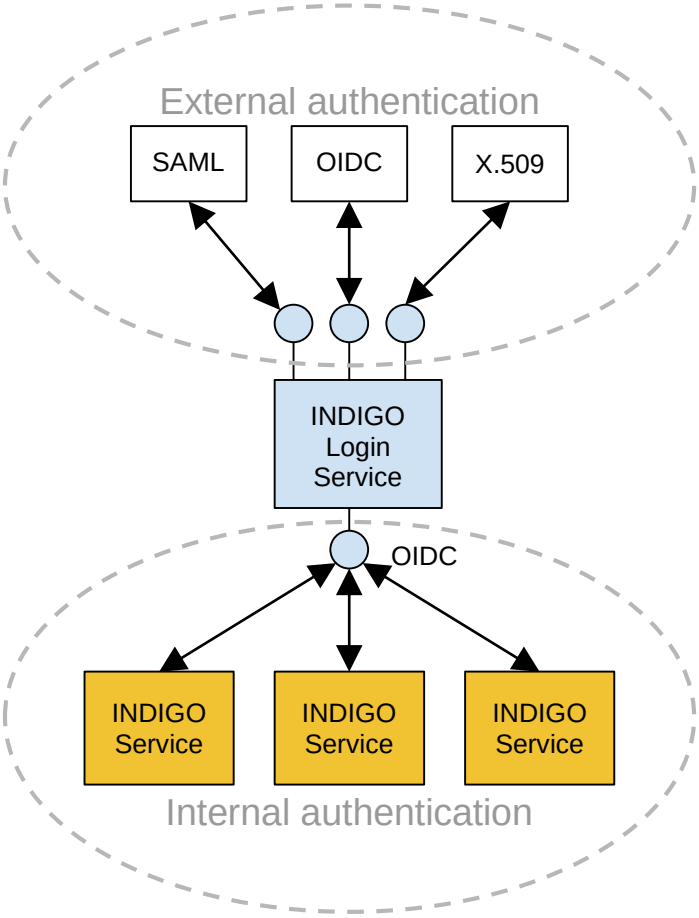

Figure 3.2: INDIGO architecture / login flow.

As shown in Figure 3.2 the INDIGO Login Service supports multiple authentication mechanisms and exposes identity information to relying parties through the OpenID Connect protocol.

What does the INDIGO-DataCloud login flow typically look like?

The following example of the INDIGO-DataCloud login process is based on the use case of users of a big experiment (BigExp). They use the BigExp science gateway (SGW) to submit and monitor computational activities on the data produced by BigExp detectors. This data is stored and then analysed on resources shared by several data centres participating in BigExp.

The authentication and authorisation flow starts with the user logging in at the SGW (Step 1). The SGW checks if the user is already authenticated and has an active session (Step 2). If not, the user is

redirected to the IAM-service (which is the combination of the Login Service and the Group Membership Service) for authentication (Step 3). This OpenID Connect flow is depicted in Figure 3.3 below (where UA stands for User Agent).

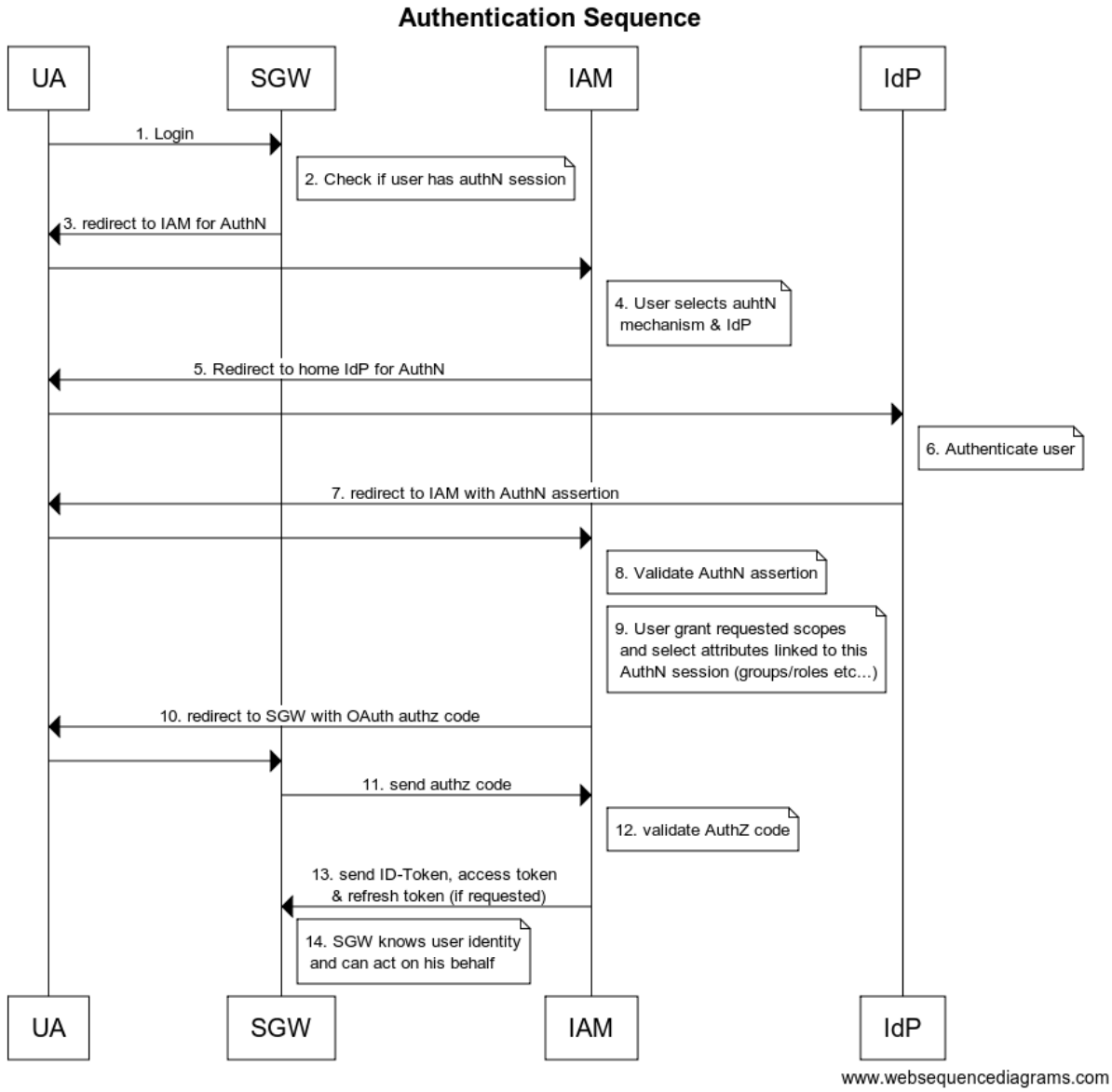

Figure 3.3: INDIGO authentication sequence

The IAM acts as an OpenID Connect provider that serves as a proxy for external authentication methods, such as those provided by the user home organisation (SAML, OpenID Connect or X.509) (steps 4 to 8 in Figure 3.3 above).

In Step 9, i.e. when granting consent to the OAuth2 scopes requested by the SGW, the user can also select the attributes he wants to include for the generated authentication session. This supports use cases in which a user acts in different roles mapped to different authorisation privileges.

In steps 10 to 13 the authorisation code obtained is exchanged with the OpenID Connect ID token, the access token and (if requested) the refresh token. The ID token is a signed JSON Web Token (JWT) providing information about the user, e.g. the user INDIGO identifier and other information about user authentication. The access token grants access to APIs and services according to the requested scopes (e.g. authorise the SGW to submit jobs and access data on behalf of the user). The access token also authorises access to the OpenID Connect userinfo end point, where additional information about the user can be retrieved.

3.1.7 Major Benefits

What makes INDIGO-DataCloud attractive and why?

- Decoupling of authorisation and authentication, so that multiple authentication mechanisms are supported and linked to the same INDIGO identity.
- Support for delegation: by leveraging and extending the OAuth delegation model, INDIGO AAI provides a delegation solution independent from the authentication mechanism used.
- Reduced integration complexity in services: choosing OpenID Connect as the identity layer facilitates integration in services. Identity provisioning, provided through standard interfaces, is another important functionality needed to enable effective integration at services and resource providers.
- Integration with legacy and non-HTTP services: the INDIGO AAI can integrate services that are not OpenID Connect-aware via a token-translation step.

3.1.8 Limitations

- The INDIGO AAI is currently in development.

3.1.9 How to Join

How could a research community make use of INDIGO-DataCloud?

The first release of the INDIGO-Datacloud AAI components will be made available in July 2016.

4 Research Community-Specific AAIs

Several large research communities have started to operate their own AAIs, which often make use of and extend some of the general-purpose AAIs described in Section 2. Examples of such research communities are:

- **DARIAH**: The aim of the Digital Research Infrastructure for the Arts and Humanities (DARIAH) [DARIAH] is to facilitate long-term access to, and use of, all European Arts and Humanities digital research data.
  DARIAH-DE, the German member of the DARIAH network, has integrated its services in eduGAIN and offers an own community-specific attribute authority that allows the attribute set of federated users to be completed/enhanced.
- **ELIXIR** [ELIXIR] is a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bio-industries and society.
  ELIXIR is about to build an ELIXIR AAI that creates an own ELIXIR identity. The system will provide external identity consolidation, step-up-authentication, levels of assurance and attribute management services. ELIXIR AAI will support linking with eduGAIN identities, but will go beyond the scope of eduGAIN as the ELIXIR community includes people from industry without an eduGAIN identity.
- **CLARIN**: The Common Language Resources and Technology Infrastructure [CLARIN] aims to provide easy and sustainable access for scholars in the humanities and social sciences to digital language data and advanced tools.
  CLARIN operates an own Service Provider federation that is partially integrated into/makes use of eduGAIN as well as national federations. This allows users of these federations and eduGAIN to access CLARIN Service Providers.
- **Umbrella** [Umbrella] is a pan-European federated identity system operated by and for the users of the European large photon and neutron facilities. It provides a unique and persistent identifier across different projects in that community. Identities can be linked with other AAI identities. As of October 2015 the Umbrella test instance allowed eduGAIN entities to be linked to Umbrella identities.

A brief description of these communities and further information on how and for what reasons they have created own AAIs can be found in *Analysis of user community and service provider requirements* [AUCSPR].

5 Conclusions and Recommendations

The existing and upcoming general-purpose AAIs described in this document are often perceived as similar at first sight, or at least as providing similar functions (authentication and authorisation), but in fact they are significantly different. This is in the first instance because they fall into different categories: they are either part of e-infrastructures (EGI, EUDAT, INDIGO-DataCloud), they are services (eduGAIN, STORK) or they are technologies (Moonshot).

What they have in common is that they rely on the same limited set of technologies (mostly SAML, X.509 and, in the future, probably OpenID Connect). However, using the same technology does not make two (or more) AAIs de facto interoperable out of the box, because sometimes the technologies are profiled for a particular AAI (e.g. eduGAIN and STORK). What also hinders interoperability is that the terminology used for similar components is often different. This requires the specialists of two technologies to first learn about the other technologies before they can eventually technically connect two AAIs. There are also several differences between the AAIs' architectures and policies that hinder interoperability.

All the AAIs described were created by a particular community that shaped the respective AAI for its own needs and purposes. The needs vary and thus so do the concepts, weaknesses and strengths behind the AAIs. In consequence, as has been stated by *Advancing Technologies and Federating Communities* [AAA], there is no single AAI that works for every research community. Inventing, and especially deploying, one universal AAI does not seem reasonable or likely within the next ten years. What can be done, however, is to interconnect the different general-purpose AAIs. One good example of this is EUDAT, which to some extent already supports all the technologies used by the other AAIs.

References

[AAA]	*Advancing Technologies and Federating Communities: A Study on Authentication and Authorisation Platforms For Scientific Resources in Europe*, SMART-Nr 2011/0056, July 2012, Licia Florio et al. https://www.terena.org/publications/files/2012-AAA-Study-report-final.pdf
[AARC]	https://aarc-project.eu/
[ABFAB]	http://datatracker.ietf.org/wg/abfab/documents/
[Assent]	https://www.jisc.ac.uk/assent
[AUCSPR]	*AARC Deliverable DJRA1.1: Analysis of user community and service provider requirements*, October 2015, Christos Kanellopoulos et al. https://aarc-project.eu/wp-content/uploads/2015/10/AARC-DJRA1.1.pdf
[B2ACCESS]	http://eudat.eu/services/b2access
[B2ACCESS-DPS]	https://b2access.eudat.eu/files/data-privacy-statement.html
[CLARIN]	http://www.clarin.eu/content/service-provider-federation
[CSIC]	http://www.csic.es/home
[DARIAH]	http://dariah.eu/
[DARIAH-DE]	https://de.dariah.eu/
[DGP]	http://eu-datagrid.web.cern.ch/eu-datagrid/
[eduGAIN]	www.edugain.org
[eduGAIN-SG]	https://technical.edugain.org/governance.php
[eduGAIN-Statistics]	https://technical.edugain.org/entities
[eduGAIN-Status]	https://technical.edugain.org/status.php
[eduGAIN-Wiki]	https://wiki.edugain.org/How_to_set_up_a_Service_Provider_for_eduGAIN
[eduPerson]	http://macedir.org/specs/eduperson/
[eduroam]	https://www.eduroam.org
[EGI]	www.egi.eu
[EGI-About]	https://www.egi.eu/about/
[EGI-EGIeu]	https://www.egi.eu/about/EGI.eu/index.html
[EGI-FAQ]	https://www.egi.eu/about/faq/
[EGI-Helpdesk]	http://helpdesk.egi.eu/
[EGI-History]	https://www.egi.eu/about/EGI.eu/history_of_EGI.html
[EGI-IF]	https://www.egi.eu/news-and-media/newsletters/Inspired_Spring_2012/infrastructure_figures.html
[EGI-Numbers]	http://www.egi.eu/infrastructure/operations/egi_in_numbers/index.html
[EGI-Ops]	http://www.egi.eu/infrastructure/operations/
[EGI-Statutes]	https://documents.egi.eu/document/1
[EGI-TaOSC]	*EGI towards an Open Science Commons*, 2015, Tiziana Ferrari
[EGI-Wiki]	https://wiki.egi.eu/wiki/Support
[eIDAS-About]	https://ec.europa.eu/digital-single-market/en/trust-services-and-eid

[eIDAS-Regulation] *Regulation (EU) No. 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC* http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG

[eIDAS-Regulation-1501] http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1441782373783&uri=OJ:JOL_2015_235_R_0001

[eIDAS-Regulation-1502] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:JOL_2015_235_R_0002

[EINFRA-1-2014] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/335-einfra-1-2014.html

[ELIXIR] https://www.elixir-europe.org/

[eSENS] http://www.esens.eu/

[ESFRI] http://ec.europa.eu/research/esfri/

[EUDAT] http://www.eudat.eu/

[EUDAT-Communities] http://eudat.eu/eudat-communities

[EUDAT-Partners] http://eudat.eu/partners

[FP7] http://cordis.europa.eu/fp7/capacities/research-infrastructures_en.html

[GÉANT] http://www.geant.org/Projects/GEANT_Project_GN4-1/Pages/Home.aspx

[Horizon2020] http://ec.europa.eu/programmes/horizon2020/

[IETF] http://www.ietf.org/

[INDIGO] https://www.indigo-datacloud.eu/

[INDIGO-A] *The INDIGO AAI Architecture - draft*, A. Ceccanti, M. Hardt, B. Wegh, P. Millar https://owncloud.indigo-datacloud.eu/index.php/s/sUTRpymjANAX0Hd

[INDIGO-News] https://www.indigo-datacloud.eu/news

[INDIGO-Partners] https://www.indigo-datacloud.eu/partners.html

[INFN] http://home.infn.it/en/

[LHC] http://public.web.cern.ch/public/en/LHC/LHC-en.html

[Moonshot] https://wiki.moonshot.ja.net/

[Moonshot-A] https://wiki.moonshot.ja.net/display/Moonshot/The+Architecture+and+Protocol+Flows+of+Moonshot

[Moonshot-SA] https://wiki.moonshot.ja.net/display/Moonshot/Source+Access

[Moonshot-Wiki] https://wiki.moonshot.ja.net/

[OASIS-SSTC] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security

[OAuth] http://oauth.net/

[OIDC] http://openid.net/connect/

[REFEDS-F] https://refeds.org/federations

[REFEDS-NUPF] https://wiki.refeds.org/display/OUT/Number+of+users+per+federation

[RFC2743] http://tools.ietf.org/html/rfc2743

[RFC2865] http://tools.ietf.org/html/rfc2865

[RFC6684] http://tools.ietf.org/html/rfc6684

[SAML] http://www.oasis-open.org/committees/security

[SAMLOverview] https://www.oasis-open.org/committees/download.php/13525/sstc-saml-exec-overview-2.0-cd-01-2col.pdf

[SCHAC]	https://wiki.refeds.org/display/STAN/SCHAC
[SEEIF]	*Scholar European Electronic Identity Federation*, 2015, Jordi Ortiz et al. https://tnc15.terena.org/core/presentation/148
[SPF]	http://www.clarin.eu/node/2965
[STORK]	https://www.eid-stork.eu
[STORK2]	https://www.eid-stork2.eu
[STORK2-About]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=398&Itemid=134
[STORK2-eLAQP]	https://www.eid-stork2.eu/pilots/elearning/index.php/en/
[STORK2-Partners]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=18&Itemid=136
[STORK2-Pilots]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=406&Itemid=112
[STORK2-Pres]	*STORK as a foundation for the eIDAS e-ID architecture* A. Lioy http://www.trustindigitallife.eu/uploads/TDW 2015/Presentation-Antonio_Lioy.pdf
[STORK2-WPO]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=399&Itemid=135
[STORK-FAQs1]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=405&Itemid=137#q1 NB If this link takes you to the top of the FAQs page instead of the specific question, in your browser address field replace the "%20-%20" segment towards the end of the URL to "#".
[STORK-FAQs2]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=405&Itemid=137#q2 NB If this link takes you to the top of the FAQs page instead of the specific question, in your browser address field replace the "%20-%20" segment towards the end of the URL to "#".
[STORK-FAQs20]	https://www.eid-stork2.eu/index.php?option=com_content&view=article&id=405&Itemid=137#q20 NB If this link takes you to the top of the FAQs page instead of the specific question, in your browser address field replace the "%20-%20" segment towards the end of the URL to "#".
[Umbrella]	https://umbrellaid.org/
[UNITY]	http://www.unity-idm.eu/
[VPIF]	*The Value Proposition for Identity Federations*, Heather Flanagan et al, REFEDS/Spherical Cow Group https://wiki.refeds.org/display/OUT/The+Value+Proposition+for+Identity+Federations

Glossary

AAA	Authentication, Authorisation, and Accounting
AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration
ABFAB	Application Bridging for Federated Access Beyond web. ABFAB is the name of the set of open standards, and the working group of the IETF that created these standards, that Moonshot is based upon.
AP	Attribute Provider
API	Application Programming Interface
AQAA	Attribute Quality Authentication Assurance levels (1-4), used in STORK
C-PEPS	Citizen PEPS
CA	Certification Authority
CDI	Collaborative Data Infrastructure
CLARIN	Common LAnguage Resources and Technology INfrastructure
CSC	IT Centre for Science
CSIC	Spanish National Research Council
CSIRT	Computer Security Incident Response Team
DARIAH	Digital Research Infrastructure for the Arts and Humanities
e-IRG	e-Infrastructure Reflection Group
EAP	Extensible Authentication Protocol
ECP	Enhanced Client or Proxy
eduGAIN	EDUcation Global Authentication INfrastructure
EEA	European Economic Area
eEC	eduGAIN Executive Committee
EGEE	Enabling Grid for E-sciencE
EGI	European Grid Infrastructure
eID	electronic IDentification / electronic IDentity
eIDAS	electronic IDentification and Trust Services
EIRO	European Intergovernmental Research Organisation
ENES	European Network for Earth System Modelling
eOT	eduGAIN Operations Team
EPOS	European Plate Observing System
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
eSG	eduGAIN Steering Group
GSS-API	The Generic Security Service Application Programming Interface (GSS-API or GSSAPI) is an API for applications to use to access security services. Moonshot is a GSS-API implementation. The GSS-API is an IETF standard [IETF], defined in RFC 2743 [RFC2743].
H	eduGAIN Hub
H&S	eduGAIN Hub and Spoke

HEP	High-Energy Physics
HLEG	High-Level Expert Group
HPC	High-Performance Computing
HTC	High-Throughput Computing
HTTP	Hypertext Transfer Protocol
IAM	Identity and Access Management
ICOS	Integrated Carbon Observation System
IdM	Identity Management
IGTF	Interoperable Global Trust Federation. A body to establish common policies and guidelines that help establish interoperable, global trust relations between providers of e-infrastructures and cyber-infrastructures, identity providers, and other qualified relying parties.
INDIGO	INtegrating Distributed data Infrastructures for Global ExplOitation
INFN	Italian National Institute for Nuclear Physics
IOTA	Identifier-Only Trust Assurance
JSON	JavaScript Object Notation
JWT	JSON Web Token
KIT	Karlsruhe Institute of Technology
LDAP	Lightweight Directory Access Protocol
LoA	Level of Assurance – degree of certainty that the user has presented a credential that refers to that user's identity
IdP	Identity Provider. A server acting in an Identity Provider role by authenticating a user and creating identity assertions about a user. Term defined in SAML 2.0 specifications, cf. [SAMLOverview].
LHC	Large Hadron Collider
LOA	Level of Assurance
LS	Login Service
LTER Europe	European Long Term Ecological Research Network
MACE	Middleware Architecture Committee for Education
MACE-Dir	MACE Directories Working Group. Looks at how to describe common identity information for campus and federated relationships.
MDS	eduGAIN Metadata Distribution Service
NGI	National Grid Initiative
NREN	National Research and Education Network
OIDC	OpenID Connect (v1.0) [OIDC] is a simple identity layer on top of the OAuth 2.0 protocol. It enables clients to verify the identity of the end user based on the authentication performed by an authorisation server, as well as to obtain basic profile information about the end user in an interoperable and REST-like manner.
OAuth	OAuth [OAuth] is an open standard security protocol for authorisation that allows you to share private resources stored on one site with another site without having to share credentials.
OS	Operating System
OSG	Open Science Grid
PaaS	Platform as a Service
PARADE	Partnership for Accessing Data in Europe
PEPS	Pan-European Proxy Service

PEPS/V-IDP	Pan-European Proxy Service / Virtual Identity Provider. Interconnected proxies that integrate the STORK common infrastructure. Typically, one per country and run by a public administration or on behalf of it.
PKIX	X.509-based Public Key Infrastructure
PRACE	Partnership for Advanced Computing in Europe
RADIUS	The Remote Authentication Dial-In User Service (RADIUS) is a protocol that provides a centralised Authentication, Authorisation, and Accounting (AAA) system. RADIUS is an IETF standard [IETF], defined in various RFCs, including RFC 2865 [RFC2865]. Moonshot uses RADIUS, and its more secure sibling RADSEC, to provide rich authentication abilities.
RADSEC	RADSEC is a variant of RADIUS that transports RADIUS datagrams over TCP and TLS, instead of UDP. RADSEC is an IETF standard [IETF], defined in RFC 6684 [RFC6684].
RHEL	Red Hat Enterprise Linux
RI	Research Infrastructure
RP	Relying Party
S-PEPS	Source PEPS
SA5 T1	GN4-1 Service Activity 5 Trust and Identity Service Development, Task 1 Harmonisation
SAML	Security Assertion Markup Language [SAML]. SAML is an XML-based open standard data format for exchanging authentication and authorisation data between parties. SAML is a product of the OASIS Security Services Technical Committee [OASIS-SSTC].
SAML2	Security Assertion Markup Language version 2
SASL	Simple Authentication and Security Layer
SCHAC	SCHema for ACademia
SCIM	System for Cross-domain Identity Management
SG	Science Gateway
SLC	Short-Lived Certificate / Short-Lived Credential
SP	Service Provider. A server acting in a Service Provider role which means consuming identity assertions. Term defined in SAML 2.0 specifications, cf. [SAMLOverview].
SSH	Secure Shell
SSP	Security Support Provider
SSPI	Security Support Provider Interface
STORK	Secure idenTity acrOss boRders linked
STORK2.0	STORK 2.0 Project
TCS	TERENA Certificate Service
TCP	Transmission Control Protocol
TF-EMC2	Task Force on European Middleware Coordination and Collaboration
TLS	Transport Layer Security
TTLS	Tunnelled Transport Layer Security
TTS	Token Translation Service
UDP	User Datagram Protocol
UID	Unique Identifier
USB	Universal Serial Bus
VIDP	Virtual Identity Provider
VO	Virtual Organisation
VOMS	Virtual Organisation Membership Service
VPH	Virtual Physiological Human

Web SSO	Web Single Sign-On
WLHCG	Worldwide LHC Grid
XML	Extensible Markup Language