19-11-2013

Enabling Users: Options for Joining eduGAIN

Last updated: 04-03-2014 (integrated feedback and suggestions from the GrIdP community)
Activity: SA5 Task 5
Document Code: GN3PLUS13-642-23

Authors: **Lukas Hämmerle (SWITCH), Wolfgang Pempe (DFN)**

Contributors: Content based on discussions with Ann Harding (SWITCH), Marina Vermezovic (AMRES), Thomas Lenggenhager (SWITCH) and the DARIAH-DE community (in particular Peter Gietz, Tibor Kalman and Martin Haase). After the initial publication various feedback and suggestions from the REFEDS, CLARIN and GrIDP communities.

Contact: edugain-integration@geant.net

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7 2007–2013) under Grant Agreement No. 605243 (GN3plus).

Table of Contents

Table of Figures

Table of Tables

1 Introduction

Within the research communities the need of federated access to services is seen as an essential success factor, especially in the Social Sciences and Humanities (SSH) sector, where users may vary widely in their technical proficiency and just need a quick and easy access to web-based electronic research tools. The experiences of research communities with grid computing showed that X.509 certificate-based infrastructures were a major hindrance for wide community acceptance of research tools. Thus federated Identity Management is seen as the only acceptable authentication and authorisation technology within the SSH community.

The FP7/ESFRI programs of the EC have led to the construction of long-term Europe-wide research infrastructures, which need to inter-federate to allow for virtual organisations with members from different countries. The interfederation service eduGAIN is an answer to such a need [eduGAIN].

Academic research projects often operate their services in different countries, with many of these services requiring authentication and authorisation and could therefore benefit from integration into eduGAIN. Enabling eduGAIN interfederation support for these services requires some know-how and efforts by the service operator. Given that the number of services operated by SSH communities is probably higher than for other research communities and given that the number of services is likely to increase even more, the question is how can research projects efficiently add their services to eduGAIN?

The following three options for adding services to eduGAIN were identified:

- Option A: Add services via an existing federation.
- Option B: Create an own federation.
- Option C: Join via a Hub or Proxy.

Each of the above options has its advantages and disadvantages and not all of them are suitable for each research group. This document investigates the above options to help decide which of them is best suited to a particular research community or case.

1.1 Federated Identity Management and eduGAIN

An identity federation usually consists of multiple organisations (e.g. universities and research institutes) that agree to use a common infrastructure for authentication and authorisation. eduGAIN is a global interfederation service that interconnects multiple identity federations, both technically and legally. It allows a user from one

identity federation to access web-based services in another identity federation. eduGAIN aims at connecting all SAML-based research and education identity federations worldwide. As of September 2013, more than half of all known academic identity federations are already connected to eduGAIN [eduGAINstatus].

The ultimate goal of eduGAIN is that a researcher from a university X in country A can access a service operated in country B by authenticating with the user account issued to by the researcher's university X. The service not only learns that this researcher is from university X but it also receives further user information. This can for example include a unique identifier, name, email address and other data, depending on what information is requested by the service and the information university X chooses to release. The identity information (especially the unique identifier) can then also be used to perform authorisation. Authorisation can rely on identity data like the researcher's organisation or affiliation but it is more likely to rely on data managed by the research project itself.

A service, like a web document storage application or a research database, in the context of SAML is protected by a (SAML) Service Provider which implements and enforces the authentication. In the context of SAML-based federations, all Service Providers of a particular federation have to be listed in that federation's metadata (XML) file, since joining a federation means accepting the federation's policies and agreements. Technically, it means registering the Service Provider with that federation's operator in order to get the SP's description included in the federation's metadata file. The same of course applies to Identity Providers. Identity Providers are those entities that authenticate users of a particular organisation: they are usually connected to an organisation's user directory.

1.2 Joining eduGAIN

The constitution of eduGAIN mandates that entities (Identity Providers and Service Providers) can only join eduGAIN via an identity federation that already participates in eduGAIN, minimising operational resources. Localised documentation and support is provided by the member federations.

Normally, services are registered and added to eduGAIN via the national federation they are operated in: this is not, however a requirement. Most federations accept any service that is related to or offers a benefit to the education and research community, although most federations have strict requirements when it comes to accepting Identity Providers.

For a single service (e.g. one protected by a Shibboleth Service Provider) that is already part of an identity federation, the changes required to prepare it for eduGAIN are as follows:

- **Opt-in/Register service for eduGAIN via the local federation.** This process varies from federation to federation. In some cases it requires the operator to log in to the federation management tool and tick a checkbox, while in others sending an email to the federation operator is sufficient. Some federations might also require the organisation at which the service is operated to sign an interfederation access declaration document before any of their services can be interfederated.
- **Configure SP to load eduGAIN/interfederation metadata.** Most federations instruct the Service Provider to load another SAML2 metadata file containing the eduGAIN/interfederation entities. This step is simply a matter of copying the file and restarting of the service.

- **Adapt Discovery Service to display eduGAIN IdPs.** This step might not be necessary as some SAML or Discovery Service implementations (i.e. SimpleSAML PHP or Shibboleth EDS) automatically include the additional eduGAIN IdPs once eduGAIN metadata is loaded.
- **Adapt attribute map and policy to accept international attributes.** The attributes provided for users accessing eduGAIN services might need to be different to those used within the local federation. An example is if the application required name and surname attributes but received only the display name from eduGAIN users.
- **Adapt access control and authorisation rules if necessary.** Generally, when a service is eduGAIN-enabled, the scope of users that might access the service is broadened considerably, which is why the access control rules need to be revised.

1.3 Current Issues

Some SSH research groups like CLARIN and CESSDA have already conducted pilot projects where some of their services were added to several federations and to eduGAIN. During their pilots they discovered several issues and problems. Some difficulties were probably related to the fact that these pilots were conducted at an early stage of eduGAIN. The current issues are described below.

1.3.1 Lack of Identity Providers

When an identity federation joins eduGAIN this does not mean in all cases that all of its Identity Providers (IdP) and Service Provider (SP) also participate in eduGAIN. In fact, most federations implement an opt-in model that leaves each IdP and SP the choice to become interfederated or not. The opt-in model was implemented mostly because it makes no sense for all IdPs and SPs of a federation to be part of eduGAIN as they might only be used within a single organisation or a single federation. Some additional efforts are required to enable an IdP or SP for eduGAIN. These efforts may include policy, configuration and application adaptations and may take some time to implement. The opt-in process guarantees that only those organisations that are exposed to eduGAIN which have completed these steps.

From the research community's perspective it is mostly the Identity Providers (IdPs) that are of interest to their services because the IdPs allow the research project participants to use their organisation's identity to access the research project's services. However, most communities have participants whose organisations have yet to start or complete the opt-in process. Therefore, their users are not yet able to use eduGAIN to authenticate. This forces the research communities to provide alternatives for "homeless" users, and become a so-called "Homeless Identity Provider". An alternative to operating such an Identity Provider oneself would be to make use of social media Identity Providers (i.e. Facebook or Google). Most of them don't support SAML natively but gateways can translate between their proprietary protocols and SAML. There are, however, data privacy and other reasons that may speak against this approach. Additionally, such gateways cannot support certain SAML profiles (e.g. SAML ECP), which have become increasingly important for some applications of SAML. This then could hinder the community from using these advanced SAML features in the future.

It is expected that with time, this issue will become less of a problem as more and more Identity Providers opt-in, motivated by their federation operators or by an increasing number of attractive services available via

eduGAIN. However, it can be said that eduGAIN, like any other federation in its infancy, initially faces a "chicken-and-egg" situation.

1.3.2 Lack of Attributes

Another problem for services making use of eduGAIN is that they often need more attributes than the Identity Providers (often operated by universities) are willing or able to release about their users. This causes problems when users want to access a service and authentication fails due missing attributes. Problems most certainly arise if the eduPersonPrincipalName (ePPN), the eduPersonTargetedID (ePTID) and/or the email address are not released, as these are key identification attributes. Collaboration in research involves sharing resources and infrastructures with reliably identified users, so for many services, anonymous usage is not an option.

There may be technical and legal reasons why Identity Providers are sometimes reluctant to release all the requested information in the form of attributes. Generally, the technical problems concerning attribute release can be addressed with better tools (e.g. to capture the user's consent for attribute release after login), less privacy/security sensitive attributes (eduPersonTargetedID or the new eduPersonUniqueID) and documentation. To address the legal and policy issues that may hinder attribute release at the organisational level, several identity federations started assigning entity categories to certain services in their federations. Entity categories allow grouping those services together that meet certain requirements or follow certain procedures. This allows Identity Provider operators to create simple rules for the attribute release for such services. In the US, the InCommon federation uses a "Research & Scholarship" [RS] entity category. In Sweden, the SWAMID federation uses a "Research & Education" [RE] entity category. Both entity categories basically are assigned by the federation operators to services that support the research and education community. From a data privacy point of view, the release of user information is easier to justify for such services. The two entity categories above-mentioned have so far been used only within their national federations (InCommon and SWAMID).

A more generic approach to facilitate attribute release was initiated by GÉANT with the GÉANT Data Protection Code of Conduct (CoC) [CoC]). This is a declaration stating that the operator of a Service Provider obeys basic data privacy principles in compliance with Directive 95/46/EC – the Data Protection Directive. Based on the CoC, Identity Providers can then release data about their users based on attribute release rules without risking legal issues. There is currently also on-going work within the GÉANT 3 plus project to create an international version of the GÉANT Data Protection Code of Conduct. This international version specifically targets services that are not in the scope of the CoC for EU/EAA countries (or countries with similar data protection laws). The international CoC therefore targets mostly non-European countries (e.g. USA, Canada, Brazil, Japan, …).

1.3.3 Lack of Level of Assurances

Research communities would often like to know more about how the identity vetting of an eduGAIN users was carried out but there is currently no agreed attribute that could be used to reliably express such information. Also, it is likely to take years for organisations to harmonise their identity-vetting processes and to provide an agreed-on attribute for all of their users. Solutions for this issue can only be implemented on the level of a particular service. This implies that currently level of assurances and improved identity vetting have to be performed by the service itself, which of course is not ideal.

Potential solutions would outsource this work to another service - “Authentication-as-a-Service”. SURFnet, the Netherlands NREN, plans to introduce such a service that allows users to go through a standardised identity vetting process and authentication using a two-level method. The process of registration with the attribute authority, because it requires specific roles and entitlements which can be checked during the registration process, provides an additional opportunity to confirm the user's identity. This results in the Service Provider receiving user attributes from a user’s Identity Provider as well as from the third-party attribute authority.

1.3.4 Lack of support for non-browser applications

Most SAML-based identity federations today support only the web SSO profiles, which limits the application to web-based applications. Some research communities however also have use-cases for non-web applications like SSH access or applications[1] running on mobile phones. SAML includes profiles such as the SAML Enhanced Client or Proxy Profile (SAML ECP), which could serve as a solution in eduGAIN. Unfortunately, many Identity Providers have yet to deploy this profile, so it is not practicable for use by the research infrastructure. There are several workarounds for accessing non-web resources that involve an initial web authentication, but these are frequently user-unfriendly and troublesome to maintain.

2 Options for Joining eduGAIN

The following sections explain the different options that allow Service Provider (SPs) and IdP(s) to be added to eduGAIN. It is assumed that research communities are mostly interested in reusing the identities administrated by the universities and research institutes for which their project participants are working. Therefore, in the context of research communities, the focus for these research projects is to primarily add SPs, which protect the actual services, to eduGAIN. Large research communities are likely to want to operate at least one IdP that contains identities for users that are affiliated with a university or a research institute that is yet to participate in eduGAIN.

[1] For example INFN Catania has developed several mobile apps to allow mobile users to access parts of their Science Gateways via an API that uses SAML for authentication: https://play.google.com/store/search?q=INFN

2.1 Option A – Adding Services via an Existing Federation

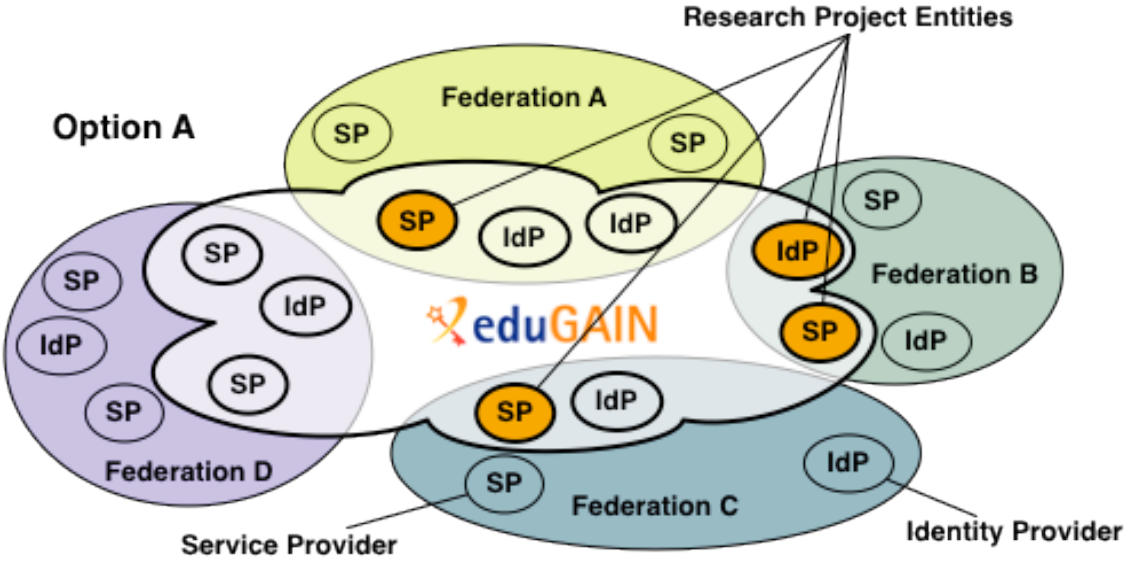

Figure 2.1: Adding Services via an Existing Federation

Adding all SPs to eduGAIN via one or more existing federations is the most straightforward option, although the registration procedures to add entities to an existing federation vary from federation to federation. The same applies to the steps necessary to enable a service for eduGAIN. In the case of a research community with services operating in multiple countries, the registration of the entities probably would be done in the respective federations.

Advantages	Comments / Discussion
Familiar infrastructure, documentation and guides, policies, legal framework, processes of existing identity federations.	When the services of a research community are registered in the country they operate in, the administrators registering the services might already be familiar with the processes of registering SPs. They will get support and assistance from their local federation operators in the local language. Most federation operators are also likely to have know-how in the area of SAML and federated identity management because many of them have been operating federations for years. Federations operated by NRENs are also likely to persist as the identity federations have become important for the NRENs when it comes to offer new services to their community.
Identities not tied to project services only.	Services registered with the respective home federations are also available for users from local IdPs that for some reason are unwilling or unable to

Advantages	Comments / Discussion
	interfederate.
From the user's point of view a very transparent solution.	Accessing a service in eduGAIN is no different from accessing a service within the local federation.
Technically straightforward.	The procedure as described in 1.2. The procedure is the same as for any other service in a particular federation that wants to become eduGAIN-enabled. All existing identity federations are able to support this model as it is "business as usual" for them.

Disadvantages	Comments / Discussion
Potentially many different federations/processes to deal with when registering SP/IdP, cf. also [REFEDSbarr].	This is primarily an issue if a central service unit of the research community carries out the registration and deployment of all SPs. Also, it is only relevant if different organisations in different countries take responsibility for these SPs even though a central service unit manages them. If the research project operators at their respective home organisations register the SPs and/or Homeless IdPs and if their home organisation is responsible for the services, this should be no issue because the operators of these services need only deal with a single federation. Federations care about who is (legally) responsible for a SP/service. An alternative approach with a single organisation to take responsibility for all SPs by a particular research group regardless of their country of operation, could register all SPs in a single federation, leading to consistent registration procedures.
Integration of the Homeless IdPs is still needed, and thus also possibly one user multiple identities.	Except for very small research groups it is always the case that the research community includes users that don't have an identity at an eduGAIN-participating organisation. It is likely that most research groups will have to operate a home for the "homeless" users. Ideally this homeless IdP would then join eduGAIN. If for some policy reason that is not possible, it would still be an option to bilaterally add this IdP's metadata to all

Disadvantages	Comments / Discussion
	SPs of that research group.

Table 2.1: Advantages and Disadvantages of Adding Services via an Existing Federation

Examples

- There are already quite a few research applications that have joined eduGAIN, among them Science Gateways developed in the context of several projects and initiatives: agINFRA [agINFRA-SG], DCH-RP [DCHRP-eCSG], DECIDE [DECIDE-SG], EarthServer [EarthServer-SG], eI4Africa [AfricaGrid-SG], EUMEDGRID [EUMEDGRID-SG], GARR [GARR-SG], GISELA [GISELA-SG], and IGI [IGI-Portal].

2.2 Option B – Creating a New Federation

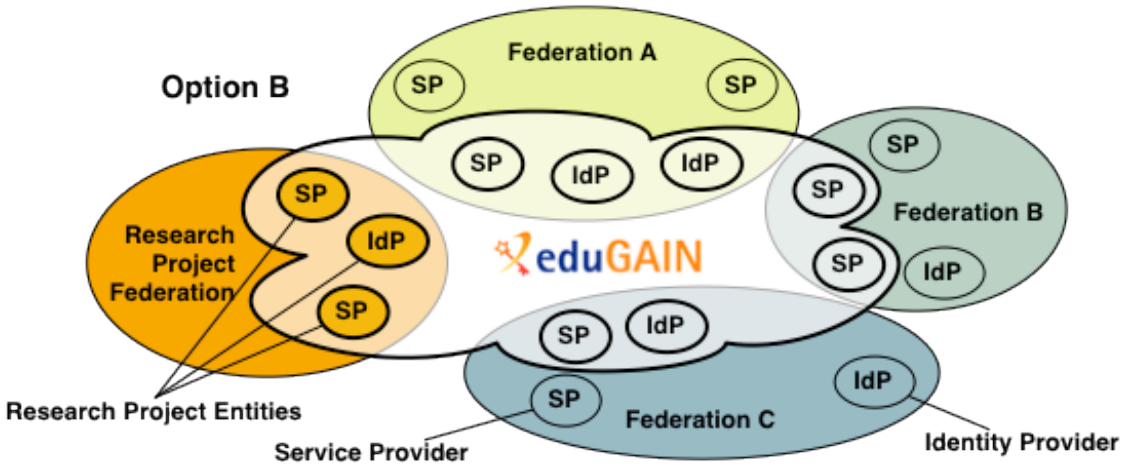

Figure 2.2: Creating a New Federation

A research community with many services may choose to create an own federation with all SPs and Homeless IdPs of all participating (national) partners and join eduGAIN as a whole federation. Each of the community's entities is registered with this federation according to a uniform set of rules.

This is the only one of the three options that is independent of existing identity federations. For this option, a research community is likely to have to deal only with the eduGAIN operations team. This option might also be especially attractive for research communities that are of a long-lived nature, maybe because they are organized as a permanent legal body in form of an European Research Infrastructure Consortium [ERIC].

2.2.1 Operating a Federation

Operating an identity federation normally involves technical, legal and policy aspects. Most existing identity federations operate at least the following services:

- An SP and IdP registration service.
- A central Discovery Service used in case services don't have their own Discovery Service.
- A metadata aggregation tool to process, publish and consume federation as well as eduGAIN metadata.
- An Identity Provider for users not (yet) affiliated to a federated organisation.

On the legal and policy side of things, most production federations have a legal framework that defines the rights, liabilities and duties of participating Service and Identity Providers. For eduGAIN it is also necessary to have a metadata registration practice statement, which describes how entities are registered.

Federations also have to offer various documentation to their community. This includes at minimum how to deploy and configure a Service Provider and how to register it with the federation. As federated identity management is often non-trivial, all federations also operate a help desk to provide technical support. Depending on the size of the federation, these help desks require a significant amount of manpower resource.

Advantages	Comments / Discussion
Potentially greater influence on IdPs to release attributes if SSH entities join up to create a federation.	It is unlikely that IdP administrators notice the registration authority of entities. Who registered a particular entity is visible in eduGAIN metadata but it is mostly irrelevant to IdP administrators.
Consistent registration of SPs within a single federation.	By creating an own federation and adding all SPs operated in multiple countries to that federation according to own instructions and deployment guides makes the installation and configuration more consistent across the research comunity.
Representation in eSG brings influence on eduGAIN operations.	By creating an own federation and joining an eduGAIN member federation, the research community is granted representation in the eduGAIN Steering Group (eSG), which controls the operation of eduGAIN and accepts new member federations. As of now, each federation has the right to assign one representative.
From the user's point of view a very transparent solution.	Information is displayed about the service the user is accessing and only those attributes are released which are requested by the service.
Technically straightforward.	Federations already have deployment guides for this scenario as it is no different than registering any service

Advantages	Comments / Discussion
	within a particular federation.

Disadvantages	Comments / Discussion
Overhead to manage a federation (policies, metadata management, own deployment guides, etc.). This requires a sustainable *operating unit* and more or less permanent *legal advice*.	Operating an own identity federation comes at certain costs and requires some persistence. Because research projects typically last only a few years, the overhead for creating and decommissioning a federation seems rather high. There are, however, also some large projects that have their own long-lived legal bodies in form of an [ERIC]. Such projects will last longer than just four years.
One user – multiple identities.	If a research community operates its own Identity Provider containing identities for all users of that community, it is very likely that a growing number of users of that community will have multiple identities: one managed by their home organisation (e.g. university) and one managed by the research community. This situation might be confusing if the research community starts allowing both identities to access its services. Unless some linking mechanism is implemented this may cause increased support requests. This might also be an issue for the other options. If users have multiple identities, it is important to have a migration strategy that allows them to eventually use only a single identity, preferably the one managed by their home institution. This strategy will have to cover an approach for account linking.
Overhead of managing more identities than (theoretically) necessary.	Services are not available for users from Home Organisations / IdPs which for some reason are not willing or able to interfederate. Those users have to be additionally registered with the community's Homeless IdP.

Table 2.2: Advantages and Disadvantages of Creating a New Federation

Examples

- The current SAML-based identity federations in eduGAIN are all operated by National Research and Education Networks (NREN). However, the European CLARIN (Common Language Resources and Technology Infrastructure) [CLARIN] research community has created a so-called Service Provider Federation [SPF] for its services. The CLARIN Service Providers are connected to a small number of national identity federations. A few of their services are already part of eduGAIN but not all of them. Therefore, the CLARIN SPF is not a good example for Option B but it is an example for a federation that is managed by a research community.

- Another example of an identity federation operated by a research community is GrIDP [GrIDP-IdPs], which is kind of a "catch-all" federation that has been operated for almost two years by the Italian National Institute of Nuclear Physics (INFN), in collaboration with GARR (the Italian NREN) and Consorzio COMETA. It counts several partners from various continents belonging both to academia and industry [GrIDP-Partners]. GrIDP also includes several IdPs that are in no federations yet as well as a "home-for-the-homeless" [IDPOpen] and a SAML-to-social media bridge Identity Provider [IDPSocial].

2.3 Option C – Joining via a Proxy

Operating a proxy that allows eduGAIN users to access services of a research community might have the advantage that only one Service Provider has to join eduGAIN because all services can be hidden behind the proxy. There are sub-options of how to implement a proxy, here named C.1 and C.2.

2.3.1 Option C.1 – (SAML) IdP Proxy/Hub

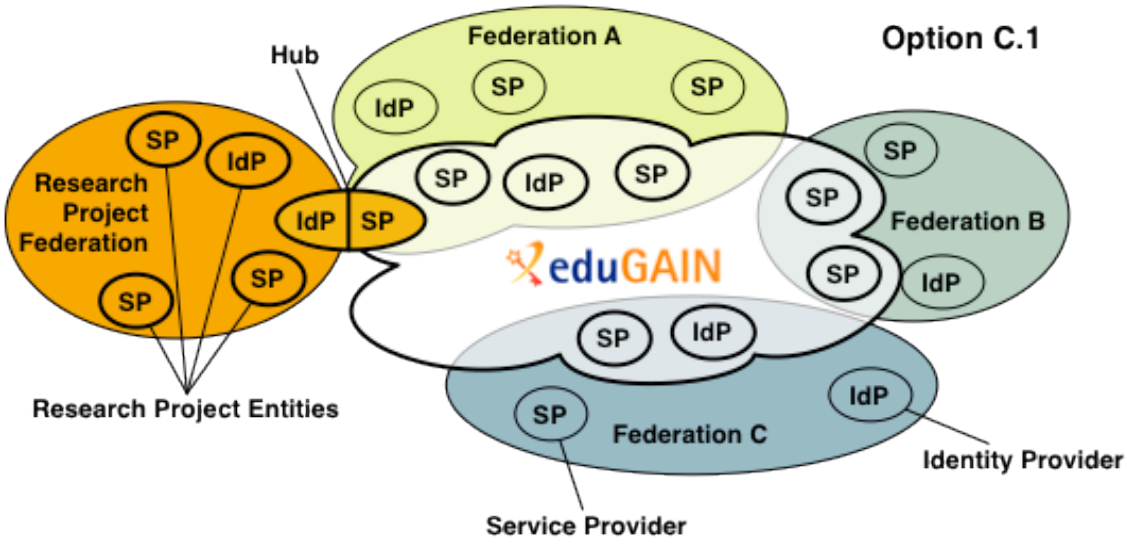

Figure 2.3: Joining via a Proxy – Option C.1

With this option[2] a research community will build create a hub that transforms eduGAIN SAML2 assertions to SAML2 assertions used within the research community. In the simplest scenario, the hub consists of an SP facing eduGAIN, and an IdP facing the research project SPs. Optionally, a user directory on the hub is used to store, transform and extend user data. The IdP's SSO login handler would have to be protected by the Service Provider exposed to eduGAIN. The hub would also have to provide a Discovery Service that could also include any number of Identity Providers (for the homeless users) operated by the research community.

Advantages	Comments / Discussion
User data can be extended, transformed, augmented.	As all assertions containing data flow through the hub, it is relatively easy for the hub to modify this data. This could be useful to introduce project internal level of level of assurance or to add group/affiliation attributes that then can be used by the services behind the proxy.
Bridging communities becomes easier.	The hub could be extended to support multiple protocols and authentication mechanisms bridging different research communities and infrastructures (e.g. SAML federations and X.509-based grid communities). The advantage would be that such changes have to be implemented only at the hub itself but maybe not at the services behind it.
eduPersonTargetedID [ePTID] would be sufficient for user mapping.	The Hub would need at minimum a unique identifier like the eduPersonTargetedID for a user. Getting this attribute should in general be unproblematic, as it does not convey more information about the user than a random string and the Identity Provider where the user authenticated. The user then could personally add attributes on the proxy, or the proxy itself could add attributes based on the user's affiliations within the research community.

Disadvantages	Comments / Discussion
Requires development work to implement the bridging/proxying.	The effort to develop and maintain such a proxy should not be underestimated. Currently, there is no out-of-the-box solution that implements this proxy solution easily. The services behind the proxy have to interoperate with the proxy itself, which technically also requires a mini-federation with a minimum set of agreements and policies. Of course these services also need to be

[2] This approach is described in detail at: https://spaces.internet2.edu/display/GS/SAMLIdPProxy

Disadvantages	Comments / Discussion
	specifically configured to interoperate with the proxy.
Hub is a single point of failure.	As all assertions from eduGAIN to the services have to go through the proxy, the proxy becomes a single point of failure. The proxy itself must be part of a high availability solution that minimises downtime.
The proxy hides all services behind it. Because they often have different attribute requirements, the proxy itself has to request the superset of all attributes required.	When users access a service behind the proxy, the Identity Provider at which they authenticate only knows the proxy but not the service behind it. Therefore, the user does not get information on the actual service being accessed. Requesting more attributes than actually needed is problematic from a data protection point of view.

Table 2.3: Advantages and Disadvantages of Joining via a Proxy – Option C.1

Examples

- The large photon and neutron research community (CRISP [CRISP] and PaNdata [PaNdata] projects) intend to build a hub like described above in form of the Umbrella infrastructure.

Bi-directional Proxy

In the simple unidirectional scenario described above, a single SP is registered in an existing federation that is an eduGAIN member. This allows eduGAIN users to access the services of that community. Assuming that the community also operates an own Identity Provider (maybe on the hub itself), this scenario is easily extended so that these research communities' users could also be enabled to access eduGAIN. This bidirectional hub architecture would imply that the hub is registered as an Identity Provider with an existing federation and with eduGAIN.

While most academic identity federations are relative tolerant of accepting most services/SPs in their federations, they often have stricter rules when it comes to accepting new organisations and Identity Providers. This is because federations generally want to limit the type of identities to research and education organisations. However, research communities are "virtual" organisations consisting of project participants from different physical organisations. The research projects thus often don't have a legal body allowing them to sign a federation service agreement as a typical organisation like a university. Exceptions are those large projects that have their own [ERIC]. It is likely that not all federations currently would be able to accept an IdP operated by a research community and if it does, it might not be for free as some federation will charge organisations for running an IdP. This applies especially where the organisation running the IdP does not pay for other services (e.g. a network) that the federation operator provides. One potential solution would be if one of the organisations that participate in the research project decided to vouch for the whole research community by

running that research project's Identity Provider. That organisation would then become responsible and liable instead of the federation.

2.3.2 Option C.2 – (Web) Proxy

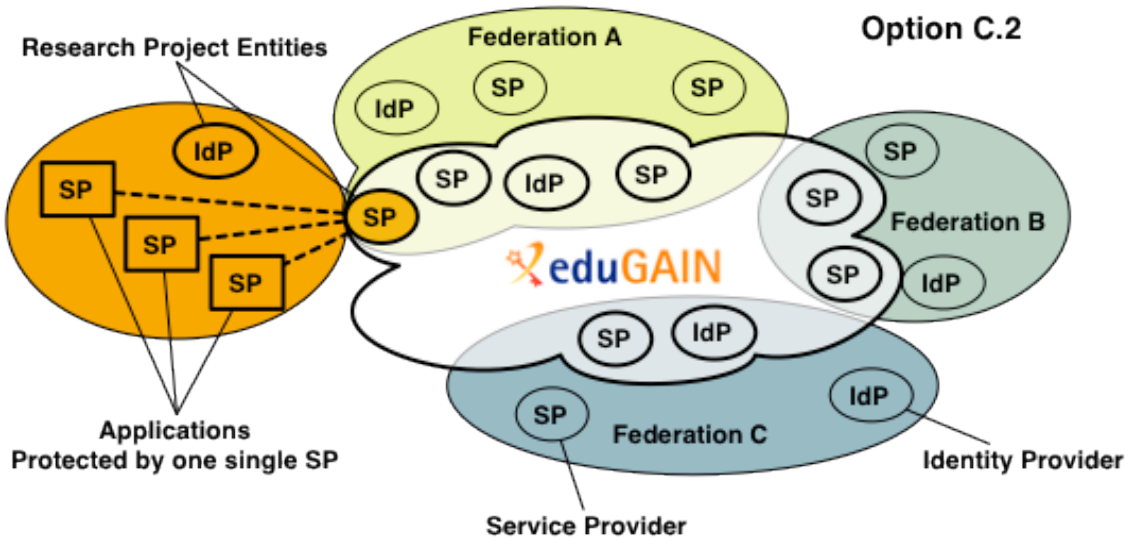

Figure 2.4: Joining via a Proxy – Option C.2

An alternative to operating a fully-fledged proxy consisting of an IdP/SP component would be to operate a standard web server with one physically installed SP that is configured for multiple virtual SPs[3]. Together they could serve as reverse proxy that can protect multiple applications behind it. It must be ensured that all traffic to the applications flows via the proxy and that applications behind the proxy reject HTTP headers containing Shibboleth attributes from hosts other than the proxy. The proxy can host multiple Service Providers and they can be registered to one or many different federations. One SP registration could be used to serve multiple applications, but this is opaque to the user because for example the applications behind the proxy might have different needs for attributes, which is likely to conflict with data protection principles.

This approach (Option C.2) can also be used in combination with Options A, B and C.

Advantages	Comments / Discussion
Only one Service Provider would have to be operated to protect multiple applications.	As all services are behind the proxy, it is essentially sufficient to operate a single physical Service Provider that is configured to host multiple virtual/logical Service Providers for the different applications it protects. Operating only a single Service Provider might also have the advantage that fewer people need to be familiar with SAML and federated identity management.

[3] The general concept of this approach is described in detail for Shibboleth at: https://wiki.shibboleth.net/confluence/display/SHIB2/SPReverseProxy

Advantages	Comments / Discussion
	To improve availability, it might be advisable to operate multiple redundant Service Providers.
The SP need only be registered in a single federation.	As the proxy is operated only in one single country, it probably will have to be registered only in a single federation even though the actual services behind the proxy might be operated in various countries.
SP can still be configured for different applications with different attribute needs. Therefore, the proxy is transparent for the user.	Depending on the SAML implementation, each virtual/logical Service Provider can be configured individually for each service it protects. From a user's point of view, there are no drawbacks regarding data protection and transparency.

Disadvantages	Comments / Discussion
Proxy becomes a single point of failure.	As all traffic to the services flow through the proxy, it should be protected by redundancy (i.e. multiple instances, shared database for session management). An outage of the proxy will otherwise cause service disruptions for all services behind the proxy.
All network traffic has to flow through proxy.	While this generally is not an issue on a technical level, of course it increases the risk that network problems affect the operation of the service.
Increased complexity and harder to debug.	The increased complexity with services distributed on different web servers (behind the proxy) managed by different administrators makes it difficult to maintain this solution. In the case of problems, debugging might be more difficult.

Table 2.4: Advantages and Disadvantages of Joining via a Proxy – Option C.2

Examples

- There are many universities in various federations that use this approach within a single federation. The difference between this scenario and an interfederation scenario should however be relatively small, as the basic principles are the same.

2.4 Summary Table

Table 2.5 summarises the above points for the four options considered. Note that the background colours of cells denote advantages (green), neutral (orange) or disadvantages (red).

	Option A: Add services via an Existing Federation	Option B: Create Own Federation	Option C.1: (SAML) IdP Proxy/Hub	Option C.2: (Web) Proxy
Technical Overhead	Reuse of infrastructure, documentation, guides, help/support desk and processes of existing federations	Own metadata management (register entities, create/sign/publish metadata file) must be deployed and maintained.	Must implement and maintain the bridging/proxy. Own deployment instructions and metadata managements is needed for the SPs behind the bridge.	Web Proxy (specially configured Apache and SP) must be deployed. Applications behind the proxy must accept only connections from the proxy, otherwise, HTTP Header spoofing becomes easy.
Administrative Overhead	None	Setting up a federation might include creating own agreements, policies, own deployment guides, metadata registration statements, etc. To join eduGAIN, at least a federation policy and a metadata practice statement must be available. Federation must be accepted by the eduGAIN Steering Group.	Similar to option B but can be less formal as the proxy would join eduGAIN via an existing federation.	None
Entity Registration	Each SP needs to be registered. If	Consistent registration of SPs	At minimum one SP needs to be	In a basic scenario one SP needs to

Overhead	there is a large number of SPs, potentially they may have to be registered in different federations which means many processes to deal with. If the research project operators at their respective home organisations register the SPs and/or Homeless IdPs and if their home organisation is responsible for the services, this should not be an issue because the operators of these services only deal with a single federation.	within only one federation, regardless of where the service is operated. But the SP has to be somehow registered.	registered within a single federation. If identities of that community must access other eduGAIN services, an IdP also must be registered.	be registered. If multiple SPs are used in order to reflect different attribute requirements by applications, they must be registered separately, potentially in one or multiple federations.
Overhead to register own IdP in eduGAIN	Registration with an existing federation depends on the federation's membership rules and might require fee payment.	Easy to register an IdP with an own federation. Criteria according to which IdPs can join federation are created by the federation operator.	Registration with an existing federation depends on the federation's membership rules and might require fee payment.	Registration with an existing federation depends on federation's membership rules and might require fee payment.
Maintenance overhead	No additional maintenance besides that of operating the individual SPs and IdPs.	Requires additional maintenance to manage and operate a federation. Support has to be provided to new SPs and IdPs. They have to be registered. Conformance to policies has to be	Requires additional maintenance of the Proxy/Hub code and configuration. Hub must somehow manage which SPs and IdPs behind the hub are accepted by this hub. SPs and IdPs behind	Requires maintenance of the Proxy (basically an Apache web server plus a Shibboleth SP). Both will have a non-trivial configuration and setup.

		checked, etc.	the hub must be supported.	
Resilience to failure	Depends on application. No single point of failure usually.	Depends on application. No single point of failure usually.	Proxy is a single point of failure. Should be operated redundantly and with high-availability setup.	Proxy is a single point of failure. Should be operated redundantly and with high-availability setup.
Transparency and Data protection from user's point of view	Good. Normally, no big difference between accessing a service in local federation or via eduGAIN.	Good. Normally, no big difference between accessing a service in local federation or via eduGAIN. If users also get an account for accessing services of the particular research community, this might be confusing as users might use two accounts.	Not very transparent and ideal from a data privacy point of view because the Hub's SP must always request the maximum set of attributes that are behind the proxy.	Good if each application with different sets of attributes is registered individually even though they are protected by the same proxy. It would however also be possible to register multiple applications with the same attribute requirements together as one logical SP. In this case, transparency would suffer.
Other Aspects		Greater influence on the operation of eduGAIN because all participating federations have a representative in the eduGAIN Steering Group. Might not be suited for short-term research projects.	Allows transforming, extending and augmenting user attributes before sending them to services behind the proxy. The proxy could also serve as protocol translator, allowing bridges to other communities by supporting protocols other than SAML.	Compared to the other solutions only one physical installation of an SP is operated. This approach can also be mixed with Option A, a part of Option B or Option C.1.

Table 2.5: Summary of the four Options Considered

Recommendations

Options B and C include additional overheads both on the technical and organisation level. While creating an own federation (B) has some advantages, when it comes to deploying the Service Provider consistently and gaining a vote in the eduGAIN Steering Group, the overhead of creating and operating an own federation seems not to justify these advantages. Even more so as the argument of consistent registration only applies if a large research community provides a specialised team to registers services in the respective local federation. This is unlikely and might not be needed as some federations also accept SP registrations from other countries as long as they are Research and Education-related. What's more, creating an own federation is unlikely to increase the chances of receiving more attributes from the Identity Providers operated at universities, for example. Creating an own federation currently seems only to make sense for large long-lived research communities or several affiliated research research communities and even then the advantages are marginal. In both cases, it should be ensured that the governing body of a federation persists for more than a few years.

Adding service via a hub (Option C.1) has some advantages when it comes to process and enrich user data. It also allows bridging different communities easier because this can be centralised at the hub. It should be remembered that creating and operating a hub is technically non-trivial. This is usually only done by large identity federations like SURFconext (NL) or WAYF (DK) that have the know-how and manpower to operate such a hub. In addition, hubs and proxies are usually non-transparent and can have disadvatages from the user-friendliness and data privacy point-of-view.

Operating a web proxy (Option C.2) is an option that also could be part of Option A as it has little impact on the way that services are registered. This approach might however be a convenient way to centralise the SAML know-how and the operation of the Service Provider in a research community. While the actual services behind the web proxy can be operated anywhere, the web proxy itself (typically consisting of a standard Apache web server and a Shibboleth SP) can be configured to protect multiple applications with different attribute requirements. As is the case for C.1, C.2 introduces a single point-of-failure that has to be carefully taken into account.

Appendix A Requirements for Federations

The requirements for a federation that wants to join eduGAIN are described in [eduGAINconstitution], section 3.1. In particular, the federation has to serve primarily the research and education sector and it has to follow the eduGAIN SAML 2 metadata profile [eduGAINmdprofile]. The eduGAIN Steering Group (consisting of one representative of each member federations) has to approve new federations.

Usually a federation has:

- A name, a logo, a web page
- A policy stating:
  - who is accepted based on what criteria
  - processes for handling complaints and incidents
  - what technical standards and profiles are used
  - how entities are registered (metadata registration practice statement)
- Deployment guides and instructions how to install and configure SPs and IdPs
- Support contact and support staff

For a comprehensive checklist for joining eduGAIN, see [eduGAINjoining].

Glossary

CESSDA	Council of European Social Science Data Archives [CESSDA].
CLARIN	Common Language Resources and Technology Infrastructure, which aims to provide access to digital language data for HSS scholars [CLARIN].
CRISP	Cluster of Research Infrastructures for Synergies in Physics [CRISP]
ePPN	eduPersonPrincipalName, cf. [eduPerson] attribute schema.
ePTID	eduPersonTargetedID, cf. [eduPerson] attribute schema.
ESFRI	European Strategy Forum on Research Infrastructures, [ESFRI].
Federation	Identity federation. An association of organisations that come together to exchange information as appropriate about their users and resources to enable collaborations and transactions.
FP7	Seventh Framework Programme [FP7].
IdP	Identity Provider. A server acting in an Identity Provider role as defined in SAML 2.0 specifications, cf. [SAMLOverview].
PaNdata	Photon and Neutron Data Infrastructure [PaNdata]
SAML	Security Assertion Markup Language [SAML]
SP	Service Provider. A server acting in a Service Provider role as defined in SAML 2.0 specifications, cf. [SAMLOverview].

References

[agINFRA-SG]	http://aginfra-sg.ct.infn.it
[CESSDA]	http://www.cessda.org
[CLARIN]	http://www.clarin.eu
[CRISP]	http://www.crisp-fp7.eu
[CoC]	https://refeds.terena.org/index.php/Data_protection_coc
[DCHRP-eCSG]	http://ecsg.dch-rp.eu
[DECIDE-SG]	http://applications.eu-decide.eu
[EarthServer-SG]	http://earthserver-sg.consorzio-cometa.it
[el4Africa-SG]	http://sgw.africa-grid.org
[eduGAIN]	http://www.edugain.org
[eduGAINconstitution]	http://www.geant.net/service/eduGAIN/resources/Documents/GN3-10-326%20eduGAIN_constitution%20v2.0.pdf
[eduGAINjoining]	http://www.edugain.org/technical/joining_checklist.php
[eduGAINmdprofile]	http://www.geant.net/service/eduGAIN/resources/Documents/eduGAIN_metadata_profile_v3.doc (final draft)
[eduGAINstatus]	http://www.edugain.org/technical/status.php
[eduPerson]	http://middleware.internet2.edu/eduperson/ eduPerson(200806)
[ESFRI]	http://ec.europa.eu/research/esfri/
[EUMEDGRID-SG]	http://applications.eumedgrid.eu/science-gateway
[ERIC]	http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric
[FP7]	http://cordis.europa.eu/fp7/capacities/research-infrastructures_en.html
[GARR-SG]	http://sgw.garr.it
[GISELA-SG]	http://gisela-gw.ct.infn.it
[GrIDP]	http://gridp.garr.it
[GrIDP-IdPs]	http://gridp.garr.it/identity-providers.html
[GrIDP-Partners]	http://gridp.garr.it/partners.html
[IDPOpen]	http://idpopen.garr.it
[IDPSocial]	http://idpsocial.garr.it
[IGI-Portal]	https://portal.italiangrid.it
[PaNdata]	http://www.pan-data.eu
[REFEDSbarr]	https://refeds.terena.org/index.php/Barriers_for_Service_Providers
[RS]	https://refeds.terena.org/index.php/Entity_Categories/R%26S
[RE]	https://portal.nordu.net/display/SWAMID/Entity+Categories
[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
[SPF]	http://www.clarin.eu/node/2965