Introduction

Federated authentication has become the dominant way of handling authentication for many institutional and research services alike. Typically, the institution (the Identity Provider or IdP) provides the user with an identity to be used for login into the Service Provider (SP). Both the SP and IdP are members of national R&E identity federations as they exists now in more then 63 countries. National federations provide a trust framework (the federation policy), and sometimes also technical infrastructure to support the federation. With the use of the eduGAIN interfederation, national federations can now facilitate authentications across national borders.

Unfortunately, several scenarios exist where the use of eduGAIN or the national R&E federation does not suffice for all authentication needs for the services. This is especially true for research collaborations, where participants from other sectors like healthcare, government or commercial entities may want to login. On campus, ‘guest accounts’ are for example needed in case an employee of a company becomes a student again, be it temporarily.

Over the years many solutions have been adopted to try to resolve this issue.
One very common solution is to create an account for the user in a local LDAP or AD. While this technically works, its associated cost for setup, maintenance and support is rather high, between 200 to 1000 euros per account. Also, the management of these accounts comes with GDPR concerns and very often these accounts do not get deleted.

Another possible solution is to allow a user to login via a so called ‘social’ identity provider like Google or Facebook, typically next to an existing federated authentication e.g. via eduGAIN. Benefit of such a solution for the service is that management of the identity, including the credentials, is outside of the service. Setback of such a solution is that this requires additional custom implementations that need to be developed and maintained. For the user involved, it may be that the guest identity provider implemented by the service may force the user to use a personal account to login, whereas the purpose of the login is often work related. And finally, several of the companies that offer the capability to be used as an authentication provider have business motives that are in sharp contrast with values associated with the academic community.

Nevertheless, many research communities as well as national identity federation operators have chosen to implement or operate a solution to facilitate such logins into services in their community or federation.

A guest identity proxy for eduGAIN

In an attempt to reduce the need for deploying such solutions for campuses, federations and research collaborations alike, the GEANT 4.2 project started to investigate the feasibility to deploy a centrally hosted proxy solution to allow such guest identity providers to be injected into eduGAIN. This way, all services in eduGAIN would be able to use these guest identities in a similar way as they are currently using the existing eduGAIN entities, reducing the additional integration cost to basically zero. It turned out that a technical solution was rather easy to deploy, but does come with a number of operational concerns. Next to this, the introduction of GDPR made many entities much more aware of the need to properly arrange for privacy and data protection.

The GEANT 4.3 Trust and Identity Incubator has continued work in this area and with this document it investigates the GDRP considerations for running a guest IdP proxy for eduGAIN where it takes into account the use of identities from various specific sources.

Proposed setup and features

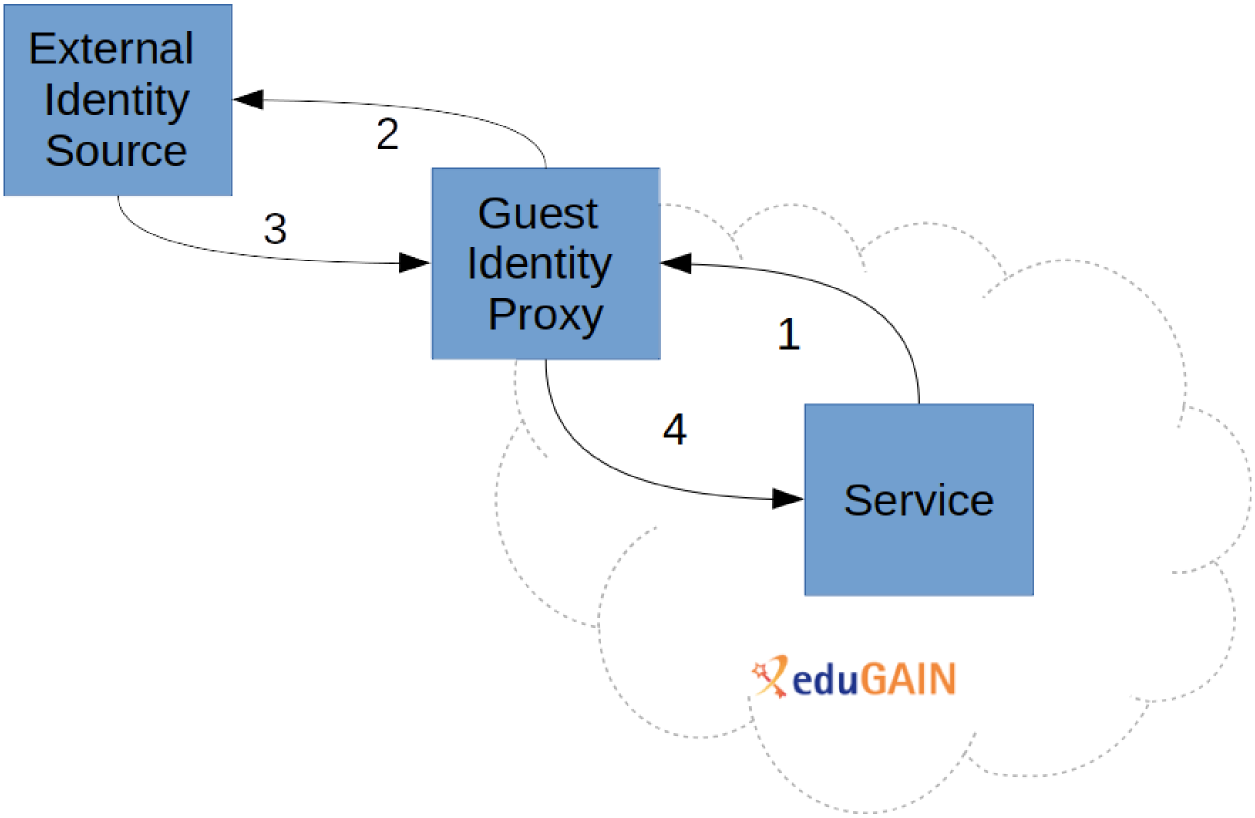


Figure 1

In the proposed setup, a Guest Identity Proxy (GIP) aggregates multiple external identity sources. In a typical flow the authentication (1) is started at a service who queries the proxy. Depending upon setup, it the GIP may present the user with a discovery page to allow the user to select a preferred identity source. The GIP then starts an authentication request towards the selected source (2). Upon successful authentication, a response is returned, including an identifier and possibly some user profile information (3). This data is interpreted by the GIP and then passed on to the requesting service (4)

It is assumed that:

- Both the GIP as well as the service are part of eduGAIN
- GIP provides the *R&S attribute bundle*[1] passing on no less, but also no more than:
  - A persistent identifier
  - A name of the user
  - An email address
- The persistent identifier is created by and scoped to the GIP
- It is not possible for the Service to determine the originating identity based on the personal data provided by the GIP
- The authentication through GIP is 'live' and no caching of user data is required at the GIP

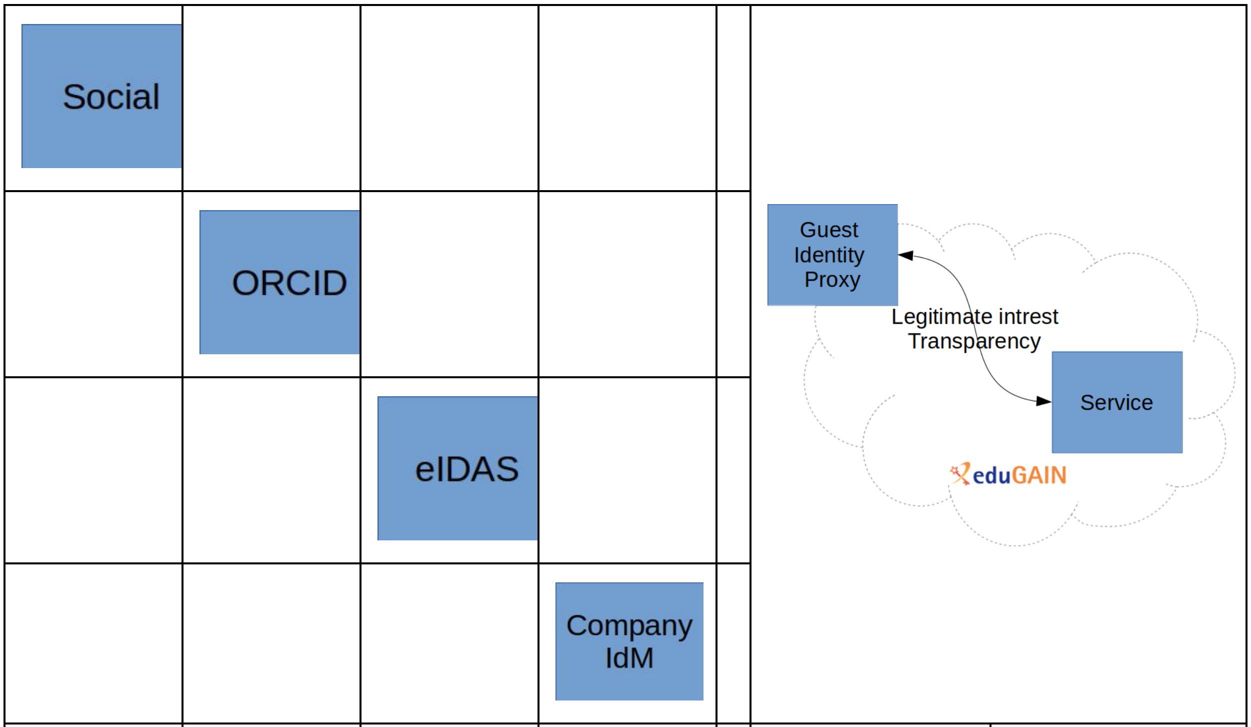
External Identity Sources

We are assuming the following entities will be providing identities as a External Identity Source, and will describe the consequences of using these respectively in the context of GDPR:

- 'Social' identity, including Google and Facebook: Many social platforms offer free to use APIs which can be leveraged to authenticate a user at such platform.
- ORCID: ORICD offers a free to use APIs (the public API) and for members of ORCID also a members API
- eIDAS: eIDAS provides an ability to use the national identity systems of EU nation states to authenticate users.
- Company Identity: Many companies collaborate with the R&E sector. Many of them have internal IdM systems. Some may also already provides authentication capabilities for their employees towards other (cloud) services.

In this document we will discuss the potential roles of the entities involved, be it data controller or data processor and the consequences that has for the way we can engage with such entities in teh proposed setup

[1] https://refeds.org/category/research-and-scholarship



					Joint controller	Joint Controller
Joint controller					Joint controller	
	Joint controller				Joint controller	
		Controller			Processor	
			Controller		Processor	

ORCID GDPR considerations

Legal basis for attribute release

In Figure 1 above the external identity source must release (and no more) the R&S attribute set towards the guest identity proxy (GIP), and under the GDPR rules a lawful basis must exist for this release. The GIP (which is a part of eduGAIN) can then pass this attribute information onto the SP and the justification for this release on the basis of legitimate interest under GDPR is already well established and described in https://wiki.refeds.org/display/ENT/Guidance+on+justification+for+attribute+release+for+RandS

In the context of ORCID since GEANT has a membership agreement in place with ORCID and since the GIP is owned and operated by GEANT the agreement grants a licence to use the API credentials to access the member API to read/deposit and use the ORCID record data subject to the privacy settings that are set by the user.

When an ORCID ID is registered the user is the legal owner of their ORCID record. At this point the user is also prompted to set the privacy (visibility settings) for their ORCID record so that trusted parties who access it via the API or ORCID.org website are only able to access authorised parts of their ORCID record. The visibility settings give three levels of privacy which are:

- Everyone;
  Information can be viewed by anyone
- Trusted parties;
  These can be either trusted organisations, who you have granted permission to access and interact with your ORCID record, or trusted individuals who have been granted access to update your ORCID record. In the context of attribute release only trusted organisations are relevant. .
- Only me.
  Information can only be viewed by the user

Hence the user must add GEANT as a Trusted organisation and give their consent (i.e set the visibility) for the appropriate attributes that are part of the R&S set - persistent identifier, name, email address. A unique persistent identifier can be generated inside the GIP. The attributes ‘name’ and ORCID Id are available to everybody by default when a user registers, however the default setting for ‘email address’ is private to the user so this would need to be set appropriately on registration.GEANT will need to provide the user with notice that their data is being disclosed to a third-party.

The setting of attributes at registration presents an issue for the workflow shown in Figure 1. In general it would be natural to ask for attribute release consent at the point of releasing the attributes to the GIP. However, in the case where the user has not at registration consented to release the necessary R&S attribute set no mechanism exists other than directing the user back through the ORCID login to change the visibility settings associated with the attributes. This should only be an issue in the case of email address, but would not be an optimal user flow. This should be investigated to see how best it can be mitigated - by instructions to the user, or by an additional workflow. There is currently no means of doing this via the member API.

Operational Considerations

In order for ORCID to be a viable IdP of last resort it must provide an appropriate level of service at the scale necessary for those using it as an IdP. Where this is not possible/available it may be possible for GEANT to mitigate some of the issues.

- **email scope OIDC API**
  We do not support the OIDC email scope, or returning email addresses in userinfo data. This decision was made because only 2% of accounts have a public email address and we did not want to give an unrealistic impression that requesting email would result in getting an email

- **use of state parameter**
  Yes, we support the state parameter. We support all functionality required to pass the OIDC conformance tests for implicit and code based flows.

- **email consent**
  We are working on enabling per-client email release for ORCID members, where an attribute release request will be presented to users during the OAuth workflow. This release process will be triggered by the inclusion of the email scope in the request. Currently this is undergoing privacy review, which we expect to complete by the end of the year. If approved (which is likely), I expect this functionality to be available at some point in 2020

- **Are there some targets for availability and reliability for the Member API and Registry?**
  We do not set targets, but uptime is generally above 99.90%. We publish our recent uptime on Pingdom.

- **Are there any known capacity limits on the number of requests that can be processed in a given time?**
  We support 40 concurrent API requests per-ip, per member.

- **Is there an optimal location for that would ensure the fastest response times?How do we monitor that status of these services and how are we informed of any scheduled downtime?**
  We use cloudflare for our website, which is located worldwide. Our member API servers are located in Rackspace’s Dallas/Fort Worth datacentre.

Scheduled downtime is announced in the Member newsletter, and on the ORCID ORCID API Users Google Group.

We publish system status on Pingdom and service status changes are announced on our Twitter feed and the ORCID API Users Google Group.

- **What are the service hours for contacting their support desk and what is the expected availability?**
  During the business week (Monday-Friday) tickets are responded to within 48 hours (typically within 24 hours). On Saturday and Sunday only emergency tickets will be addressed.

- **How are incidents prioritized and what are the associated response and resolution times?**
  Support tickets received at support@orcid.org with the word ‘urgent’ in the subject line are treated as high priority for triage.
  Highest priority areas:
  - Anything impacting site security or registrant privacy
  - Registry functionality not working for all users that is preventing registration or sign-in.

  We do not define guaranteed response times, but we generally expect critical issues to be resolved or mitigated within 24 hours.

- **What is the escalation process when incidents have not been addressed as agreed?**
  Members may escalate to their assigned Engagement Team contact.

- **Are ORCID's operational, security and response processes sufficient to claim Sirtfi (https://refeds.org/sirtfi)?**
  There are some areas where we are not compliant, but we will be working on those in 2020. The assertions required by the Sirtfi framework are consistent with the processes and safeguards we are implementing at ORCID. We should be able to self certify once the missing areas are addressed, but that will be as a service provider within the federated identity system. So, it will not be directly applicable to our relationship with

eduTeams, where we are acting as an OIDC provider. However, I think it would be reasonable to expect an equivalent level of service for OIDC.

- **What processes are in place to support disaster recovery?**
  We automatically backup the database twice daily, encrypt the dump, and push it to AWS eu-west-1 region (Ireland) so that, in the event of a disaster at our main datacenter, we can use the database backup to restore the system. We regularly test that this process is working using a temporary offline server.