ORCID IdP as a last Resort

Objectives

Many Institutions and research collaborations who provide services have participant users who are not members of an academic research federation and consequently do not not have academic credentials that can be used to access those services. Consequently, many collaborations or research federations operate IdPs or proxies to allow such users to authenticate using an external ‘guest’ identity. This imposes extra costs on these organisations, potentially limits users to the set IdPs supported by the service they wish to access, and results in much duplication of effort across the various organisations. As a consequence, ‘Guest Identity’ is high on the wishlist for research communities, as reflected e.g. in the Fim4R papers. However it should be noted that also many national federations have a need for guest identity. This was previously expressed e.g. by SURF and ACOnet.

Near the end of the GN4-2 project, the GN42 eduTEAMS project set up a pilot infrastructure, called IDhub, with the aim to investigate the possibility for offering a generic infrastructure for guest identity. Goals for this service were:

- Provide the services with a consistent (SAML 2) based interface for connecting to these guest identities, preferably throughеduGAIN;
- Provide a persistent identifier and if possible a default set of attributes, e.g. conforming to the R&S bundle;
- Provide end users with choice in the external IdP they want to use, preferably combining social IdPs, company IdP and government identity.

After comparingavailable options, it was decided to initially use 5 external social IdPs: ORCID, Linkedin, Github, Google and Facebook. The latter two were however later-on discarded because of both policy and operation issues. Github and Linkedin were introduced as Guest IdPs in eduGAIN as part of a pilot,technically ORCID was connected as well, the GN42 project ended before this was brought into the pilot.
In the GN43 project we have decided to reinvestigate ORCID. The most important consideration was that we wanted to introduce ORCID with full consent from and in collaboration with ORCID. We intendto use the existing technical solution based on the IDHub component which was initially developed within the GN4-2 eduTEAMS project, however, further work is required to scale this to serve this need. It is also in scopeif a generic guest IdP service should be positioned as an ‘eduTEAMS’ service.

This document discusses the requirements and conditions as well as the positioning of a Guest IdP using ORCID.*f*

Scope

The ORCID service can be used in a number of ways which are all potential relevant to the GEANT services in general:

1. Read & Write access to the full ORCID API.
   This is typically used by member institutions and publishers. While we do want to use the members API, we have no want to use the full set of capabilities of the API, nor do we want to write into a researchers profile in the context of this activity.
2. Add an ORCID ID to a users profile
   In the context of the eduTEAMS Membership management services (coMANAGE, HEXAA, PERUN) a research community may want to add the users ORCID to the profile. This way the ORCID ID can be provided to backend services as part of the authentication by the eduTEAMs proxy. This use case is out of scope for this activity.
3. Provide ORCID for Authentication locally
   The eduTEAMS service currently offers research communities the ability to add external identity providers on a per community basis. ORCID could be one such IdP. It is expected however offering ORCID in this way is no longer needed if scenario 4 becomes available.
4. Provide ORCID for Authentication through a generic IdPolr service
   In this scenario a centralized service “IDhub” is offered where ORCID is on of the possible external IDs to be used

The activity in the Trust and Identity Incubator only investigates scenario 4, which from now on for further reference will be mentioned as IDhub for the remainder of this document.

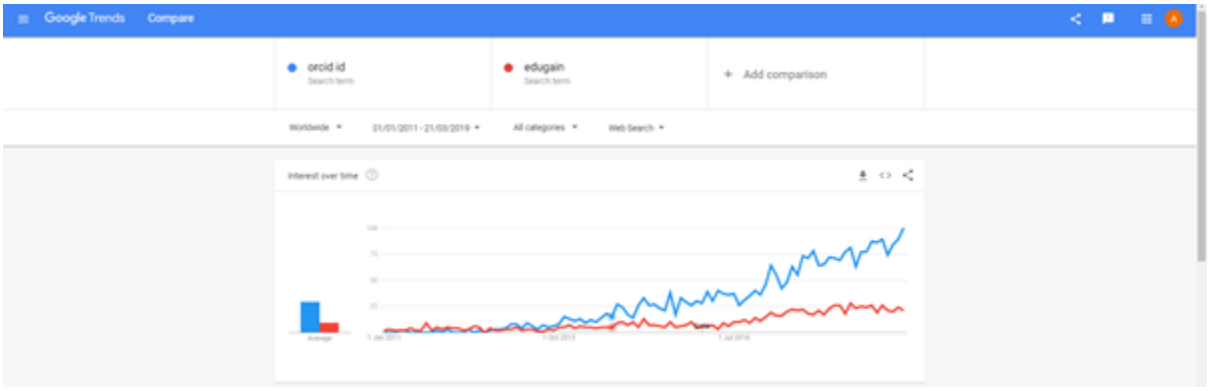
As a side note, the technology used for deploying scenario 3 and 4 is exactly the same (SaToSa), which is also the component that serves as the central proxy in both eduETAMS and InAcademia.

ORCID Market Overview

Persistent identifiers (PIDs) offer scientists and researchers benefits by providing unique keys for people, places and things, which enables accurate mapping of information between these systems and supports the research process by facilitating search, discovery, recognition and collaboration.The existence of such PIDs offer significant time-savings by reducing the time spent in administrative tasks such as registering for services. According to research published in nature in 2016 researchers spend 10% of their time in such administrative tasks.[1]

In addition to PIDs for people PIDs can also exist for organisations, funders, publishers etc. and the trustworthy connection of these PIDs together can both increase their value and offer efficiencies to all parts of the trust network.ORCID provides both personal and organisational PIDs and operates an infrastructure acting as a trust network to connect them together.

ORCID celebrated its sixth birthday in October 2018 and in that time it has grown to over 1000 member organisations in 45 countries, with over 6 million ORCID's registered globally. ORCID has over 5000 new registrations/day, and the pace of ORCID's uptake has been increasing, with over a quarter of those members joining in 2017. Revenues have increased by 19% y-on-y allowing the organisation to anticipate breakeven in 2019 and to develop reserves to repay loans (currently the cost-base is approximately $3.98M, with membership fees yielding around $4M). ORCID has a broad user base with 47% of members being in Europe, 28% in the U.S and N.A, and 21% in APAC.



ORCID is established as an independent not-for-profit 501(c)3 tax-exempt organization registered in the United States with a membership and subscription fee structure. Its business model is such that individual membership is free of charge, but institutional membership is subject to tiered pricing ranging from $5,000 - $180,000, depending on the type and size of the organisation and the level and nature of the ORCID integration that is proposed. This tiered membership is linked in to the capabilities of the API that is available - with institutional members being able to add data to individual ORCID records (with consent), and premium institutional members enjoying prioritised support, more regular analytics etc.

The model encourages researchers to register for an ID as a way of building a network that allows their contributions to be correctly identified, discoverable and recognized. There is therefore a strong incentive to register (as witnessed by its rapid growth) and the growth in such registrations is an important component in building the relevance of ORCID for institutions (who are fee paying). The current focus for institutional member categories are funders, publishers and research organisations, with concomitant application integration focus on publishing system, research information management systems, Repositories, and funding systems. Currently there are 1036 member organisations who have in some form integrated with ORCID (this compares with 2430 SPs registered in eduGAIN).

The ORCID service provider allows SSO using either an ORCID ID an institutional ID (via the supported access federations SURFconext into eduGAIN, and social Ids such as Google and Facebook.
The ORCID API supports both OAuth 2.0 and OIDC authentication mechanisms, for allowing access to the user profile data.

Because of the nature of the ORCID service, ORCID has a critical need to keep its core identifier (the ORCID ID) both stable and persistent for as long as the ORCID service exists. In addition ORCID operates in the same sector as GEANT with mostly the same stakeholders. Finally ORCID has shown in the past to be both technically competent, privacy aware as well as engaging with the Trust and Identity community in R&E. All these factors set ORCID aapart as compared to most of the other guest identifiers.

SWOT

Analysing the effect of offering ORCID as an IdP of last resort within the IDHub and potential impact on other GEANT services such as eduGAIN.

Strengths	**Weaknesses**
GEANT reputation and services in the Trust and Identity domain GEANT skill and experience in federated identity and specifically eduTEAMS Both eduTEAMS as well as InAcademia use the same software (SaToSa) to deliver their service, hence there is a good deal of expertise both functionally as well as operationally. GEANT has a unique position as a coordinating partner in a number of activities in Research and Academia eduGAIN has a broad user base of researchers, educators and students	The pilot that was set up as part of GN42 eduTEAMs IDHub is an unproven service and currently lacks credibility Increased support burden ORCID data control exceptions - relating to business usage of data and other uses Not clear how costs of running the service would scale or if there is scope for revenue generation to cover such costs Calculation of cost savings at Research communities and Federations is probably difficult.
Opportunities	**Threats**
Provides a clear benefit to the community by enabling authentication via (ubiquitous) ORCID ID Provides a trustworthy ‘guest’ credential when compared with other options Google, Facebook etc. Reduces duplicate effort by institutions and collaborations to support (ORCID as) a guest ID solution Further leverage ORCID as the IdP of choice for academia Promote eduGAIN as the central organising entity of academic IdPs and therefore increase usage Provides the basis for a proxy that could be widely shared with the community, even if not	May encourage researchers and NRENS to use ORCID as IdP of first choice, obviating the need for eduGAIN May confuse the eduTEAMS brand for organisations only seeking to use it as an authentication proxy vs. a virtual organisation platform Perception of close linkage with ORCID unless a similar capability offered for Google, FB etc. may be counter-productive if ORCID runs into problems. May further encourage authentication with ORCID directly if the eduTEAMS IDHub proxy proves an inferior user experience Competitive threats from EGI who already claim to have this capability in their AAI proxy

offered as a service by GEANT. Collaboration with ORCID will create mutual understanding and interdependency and may let us influence their service roadmap	

Risks and Mitigations

RISK
The GN42 eduTEAMS IDHub pilot is as yet an unproven service which can act as a central proxy to allow authentication by a variety of identity providers including ORCID. Scaling IdHub (if feasible) would allow the component to support ORCID IDs within a federated eduGAIN setting. However, it should be understood that although the component was a part of eduTEAMS, the future offering should probably not be. To do so might cause confusion of the eduTEAMS positioning and purpose.

MITIGATION
Consider how IDHub could be developed as a separate service from eduTEAMS and how this might affect the current eduTEAMS offering. Decoupling of IDhub from eduTEAMS brand allows eduTEAMS to focus more on supporting just Collaborative Organisations, while at the same time allowing broader use cases for IDhub, e.g. including support for use within national identity federations.

RISK
Providing support for a single persistent academic identifier (although other non-academic identity providers such as Google and Facebook are also supported) could be seen as an endorsement of ORCID as the de-facto centralised persistent academic identity provider.

MITIGATION
Consider supporting other persistent identifiers make it clear that the provision of a broad choice of guest academic identifiers is the strategy.

RISK
ORCID uptake is currently growing much faster than eduGAIN and the ORCID strategy is to expand the offering beyond academia as well as to continue to add new classes of integration.

ORCID is expanding efforts in outreach and marketing communications and such a campaign could help to raise the profile of ORCID further viz-a-Viz eduGAIN.

MITIGATION
Consider what further activity could take place to promote eduGAIN and position is relative to ORCID so that the two mechanisms can coexist and address their own specific needs

RISK
The GN42 eduTEAMS IDhub pilot service is new an unproven and the user experience may prove inferior to that offered by ORCID (especially as ORCID intend as a part of their strategic roadmap to improve the user experience) encouraging users to use ORCID directly rather than as a guest identity via IDHub.

MITIGATION
Improve the IDHub proxy experience if needed.

RISK
For a variety of reasons OAUTH 2.0 and OIDC are a better fit for many SPs than SAML. ORCID supports both OIDC authentication and OAUTH2 authorisation, whereas eduGAIN is currently only SAML based. This may limit the scope and type of services that are available via eduGAIN and make ORCID a more attractive authentication solution.

MITIGATION
It is hard to weight this risk. While OIDC and OAuth2 are easier to implement, the ecosystem of the service will likely also require interaction with eduGAIN (SAML) IdPs. If that is the case, then native ORCID integration is an addition on top of also needed SAML federation integration. In such scenario a SAML2 based ORCD proxy might actually be a much better fit

Assumptions

This section lists the principal assumptions concerning ORCID, IDHub and GEANT which are made to inform the strategy and recommendations.

- ORCID and GEANT are prepared to work cooperatively in order to best serve the needs of their Research and Education customers;
- ORCID is primarily focused on the needs of researchers rather than educators or students;
- The eduTEAMS IDHub is capable of scaling to become a proxy authentication solution and its useability will be broadly equivalent to that of ORCID;
- eduGAIN wishes to promote the use of its interfederation framework as widely as possible within Research and Education;
- ORCID wishes to promote the use of ORCID persistent identifiers as widely as possible with Research and Education and beyond;

- The cost of offering ORCID support in the IdHub is supportable and does not impact the cost of offering eduTEAMS service as a whole, or require additional fees to be charged to sustain it;

Discussion and Strategy Options

Discussion

Generally speaking, ORCID is a very good candidate to act as an external ID provider for a guest IdP solution. ORCID is primarily focused on supporting researchers and as such does not address the broader research and education user group of students that is covered by IdPs already in eduGAIN. But at the same time research is just the scenario where the requirement for guest identity is most dominant.

In terms of the service providers that would be able to authenticate research users via eduGAIN or ORCID there is some overlap and for such services it may perhaps be simpler to sign-on with an ORCID ID directly, or with an institutional ID via ORCID. Indeed the stated strategy of ORCID is to promote to researchers good reasons to use ORCID and evangelise support of persistent identifiers. Plans to improve the ORCID user experience and to make the authentication process more straightforward may also add further reasons for using the ORCID SSO. Future plans to expand beyond Academia may lead to a broader range of services being offered making the ORCID ID a more pervasive first sign-on choice.
At the same time, many research services value the fact that a user was authentication form a home institution as that provides a certain level of assurance that ORCID cannot offer.

For researchers with an ORCID ID (and this is an increasing number) there would likely be a desire to have a single SSO IdP to access all services and a tendency to choose the ORCID ID over the institutional ID for the following reasons:

1. The ORCID ID has life long persictanceand so better supports the migration of researchers between institutions that a single institutional ID;
2. Publications and grant finance are central needs for a researcher and as such there would be a tendency to use an ID associated with a system that supports such needs (where is it possible to access the needed services), than another form of ID. However, this can also be accomplished by using the ORCID ID as an attribute, e.g. in eduPersonORCID, and does not require authentication trough ORCID.
3. The sign-on process may be more straightforward and less prone to error and delay in a centralised rather than federated approach (when the ORCID infrastructure is used). Yet at the same time ORCID itself relies on external IdP for its authentication, as it currently supports both login through eduGAIN as well as through some Social IdPs.
4. Where the services are available via both the eduGAIN interfederation service and ORCID infrastructure a researcher may find it easier to use ORCID.

Were ORCID sign-on to be available the IDHub proxy and is being used by a researcher this could be for two main reasons;

1. The researcher does not have an institutional ID (they are a guest or not a member of an identity federation);
2. The researcher has an institutional ID and an ORCID ID but for the reasons given above prefers to sign on with their ORCID ID.

In the first case the IDHub proxy is providing a useful service to allow the researcher to access a service or resource that they would not otherwise be able to. Since they already have an ORCID ID and have no other option to access the resource or service it does not encourage the use of ORCID, although the researcher could choose to use the ORCID ID directly on the ORCID infrastructure if the service or resource was available there.

In the second case if the researcher chooses to sign into a service via IDHub using their ORCID ID rather than by their institutional ID this would potentially weaken the use of institutional identities, and consequently identify federations and eduGAIN. It is up to the service provider to decide if incoming identities from ORCID are treated in the same way as institutional accounts. If they do, apparently login via eduGAIN does not add any additional value over ORCID logins for the service and this is therefore unlikely to add additional risk to eduGAIN.

Although eduGAIN serves a broader base than ORCID (student, educators and researchers), support for ORCID within the IDHub is likely to encourage researchers to advocate, and SPs to implement, support for ORCID within their services which is also in sync with ORCIDs stated strategy to expand and promote the offering. This would have a consequential negative effect on the use of institutional identities, thus affecting identity federations and the eduGAIN interfederation service.

Strategy Options

Assuming eduGAIN, GEANT and ORCID want to work together to best serve the needs of their users notwithstanding any impact that would have on their respective offerings the following approach could be adopted.

- IDHub and eduGAIN should be able to provide a similar SSO experience to ORCID ID and support the same broad range of mechanisms (SAML, OIDC);
- IDHub should provide support for ORCID ID and other guest identities;
- eduGAIN should promote itself and the IDHub capability in a complementary fashion to ORCID;

Researchers and especially SP will ultimately decide how they want their identity provision to be offered and eduGAIN, IdHub and ORCID will need to adjust over time to focus on the areas

where they bring most benefit. For example, eduGAIN could focus more on educators and students and ORCID on researchers. The need to an ORCID IDHub proxy could diminish over time as ORCID expands its scope and integrations, or ORCID could remain focused on its core constituency and the need for the IDHub proxy ORCID integration could remain.

References

Young scientists under pressure - what the data show.
https://www.nature.com/news/young-scientists-under-pressure-what-the-data-show-1.20871