D2.7

Service Provider Federation Full Extension

Document information

Title	Service Provider Federation Full Extension
ID	CLARINPLUS-D2.7
Author(s)	Dieter Van Uytvanck, Willem Elbers, André Moreira
Responsible WP leader	Dieter Van Uytvanck
Contractual Delivery Date	2017-04-31
Actual Delivery Date	2017-04-31
Distribution	Public
Document status in workplan	Deliverable

Project information

Project name	CLARIN-PLUS
Project number	676529
Call	H2020-INFRADEV-1-2015-1
Duration	2015-09-01 – 2017-08-31
Website	www.clarin.eu
Contact address	contact-clarinplus@clarin.eu

Table of contents

1 Executive Summary

The CLARIN Service Provider Federation (SPF) is the organisational vehicle by which all available CLARIN Service Providers are connected to national identity federations. This enables cross-border single sign-on for academics. This deliverable describes how it was extended from 6 to 18 countries, what issues were encountered and which solutions were applied, and how it can be further enlarged in the future.

2 Introduction

Several resources in the CLARIN infrastructure are not openly accessible, mostly due to copyright (e.g. contemporary newspaper corpora), privacy concerns (e.g. video recordings of doctor-patient interaction) or the intentionally restricted scope (e.g. tools that are only intended for academic use). To enable the most open web-based access to these resources, CLARIN has been relying on a SAML-based[1] single sign-on system.

In this document we describe the chosen setup, and how this task in CLARIN-PLUS (2.1.1 – Extending the SPF coverage) has lead to a broader potential user base. It also includes a strategy for future extensions. Finally we list some options for future streamlining of the organizational setup of our single sign-on system.

[1] See https://en.wikipedia.org/wiki/Security_Assertion_Markup_Language

3 The Service Provider Federation

3.1 The Identity Federation concept

Identity Federations allow users to login to remote services with their own academic credentials (usually a username and a password). During this process there is communication between the remote *Service Provider* and the local *Identity Provider* about the authentication of the user: the Identity Provider ensures that the user really is who (s)he claims to be by checking the validity of the credentials. The identity provider then sends an OK-signal to the service provider, who trusts the identity provider statement and gives the user access to the requested resource. Often, next to the OK-signal, some additional information about the user is provided: name, an identifier, email address, etc. These personal information snippets are usually called *attributes*.

The overall process of logging in and sending the trust assertions is named *authentication*, while the part where a Service Provider relies on the trust assertions and attributes to decide whether a user is allowed to access a resource is referred to as *authorisation*. Combined they form the *Authentication and Authorisation Infrastructure* (AAI).

The Identity Federation scheme has several advantages:

- Security. As the password is never sent to a remote party, it cannot be intercepted.
- Scalability. The local identity provider (e.g. a university) has the best insight in the status of user accounts: e.g. when someone leaves the organization the account can be deactivated as part of the local procedures.
- Single-sign on. When using multiple service providers after each other, the login state at the local identity provider can be shared.

Originally this system was mainly used within a single country – hence the concept of *national identity federations*. Soon enough international use cases arose, among which those from CLARIN: providing researchers access to language resources and tools hosted at a centre in another country[2]. The *interfederation* concept addressed these demands: cross-connecting Service Providers and Identity Providers from different countries.

3.2 How it started

During the CLARIN preparatory phase (2008-2011) the state of interfederation was rather immature. eduGAIN, an initiative from GÉANT, was running ahead in this field but suffered from the fact that most national federations made participation in eduGAIN optional. In effect, only few Identity Providers could be used to login to Service Providers in another country.

The proportion of Identity Providers per national federation that have access to Service Providers from another country via eduGAIN is often referred to as *opt-in ratio*. This ratio indicates how many identity providers can access any service provider via eduGAIN – so the opt-in is valid for the whole group of all Service Providers[3]. In the table

[2] See https://www.clarin.eu/content/easy-access-protected-resources for an overview of the currently available services via the CLARIN authentication and authorization infrastructure.
[3] This differs from the still existing national federation policies where the opt-in is only valid for a *single* Service Provider. See section 3.5.3.2 for details.

below[4] we present the opt-in ratio for several federations, in December 2013 and in April 2017.

	2013	2017
Belgium	0%	50%
Czech Republic	14%	81.7%
Italy	0%	100%
Switzerland	15%	68.2%

Table 1: a sample of the percentage of Identity Providers per country that can login to a Service Provider in another country via eduGAIN.

CLARIN addressed the low opt-in ratio [CLARINPP-D2R-3b] by setting up a construction where one organisation (at that time the Max Planck Institute for Psycholinguistics, later on CLARIN ERIC) joins several national federations, representing a group of Service Providers. This so-called Service Provider Federation was in fact a pragmatic workaround for the then disfunctional interfederation scenario.

For more detailed information about both the legal and the technical implementation of the SPF we refer to https://www.clarin.eu/spf

Conceptually, the SPF works as illustrated in Figure 1.

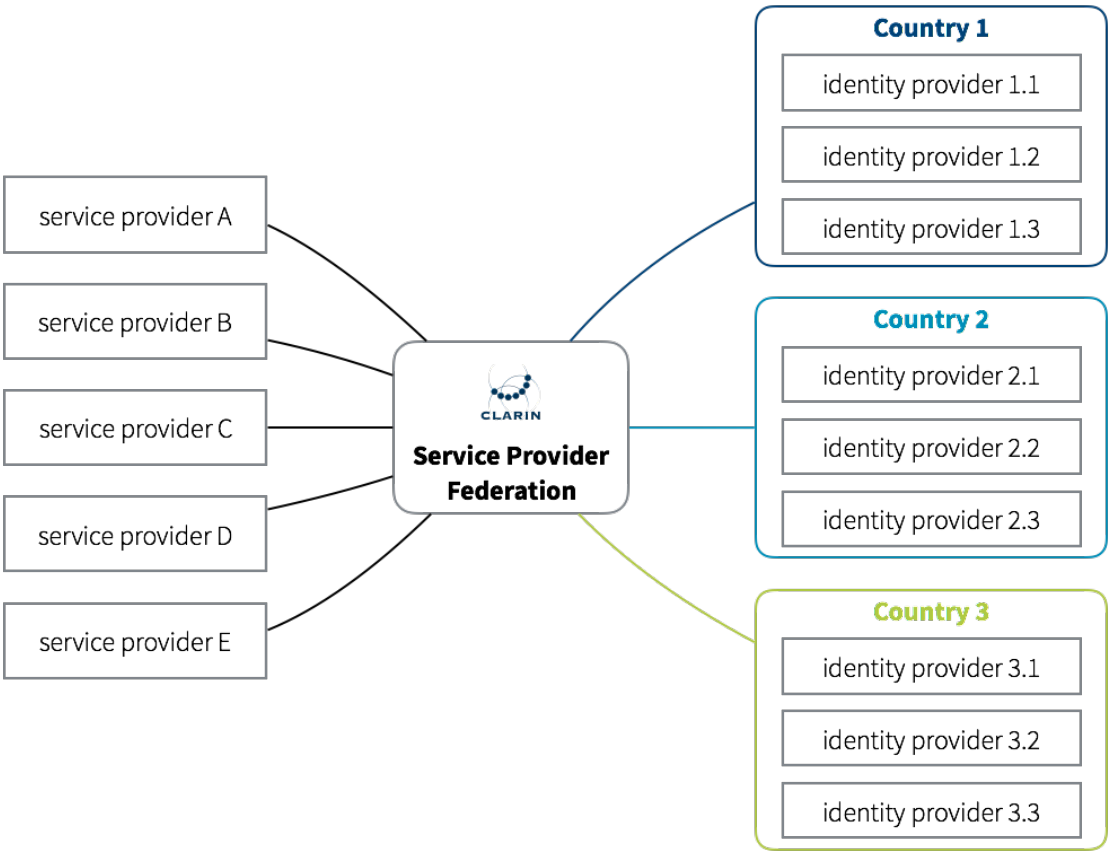

Figure 1: a conceptual overview of the Service Provider Federation.

[4] Sources: https://www.clarin.eu/node/3869 (2013) and https://technical.edugain.org/isFederatedCheck/Federations/ (2017)

Since the founding of the SPF in 2010, it has proven to be a reliable solution. Although it requires some administrative and technical overhead, to manage the membership of the individual identity federations, the broad outreach to the full academic user base of a country has even contributed to CLARIN ERIC's value proposition [CLARINPLUS-D5.4], since for many countries there is no alternative to achieve this. In other words: if a new country joins CLARIN ERIC as a member, it will be connected to the SPF, enabling its research community to use all CLARIN services via easy single sign-on.

3.3 The extension and the current state

Over time, the situation with the opt-in in eduGAIN has improved for certain countries. For those cases, CLARIN has made an agreement with the German DFN-AAI federation to publish its SAML metadata to eduGAIN. This takes away the necessity to join the national federations: those federations connected to eduGAIN will directly retrieve the information about the CLARIN service providers from the German federation.

In practice, the SPF now has 2 ways of publishing the CLARIN Service Providers to the national federations:

- the classic way, through joining the federation, in case the opt-in ratio is too low
- via eduGAIN, when the opt-in ratio is high enough

Table 2 presents a summary of the current state of the national federations in the Service Provider Federation. It illustrates that the total number of federations has been rising significantly since CLARIN-PLUS has addressed this task: from 6 to 18 over the period 2015 - mid 2017.

Country	Federation	Status	Date of inclusion	Number of Identity Providers
Austria	ACOnet	Connected (via national federation).	2015	44
Belgium	Belnet	Connected (via national federation).	Before CLARIN-PLUS	38
Czech Republic	eduID.cz	Connected (via national federation).	Before CLARIN-PLUS	92
Denmark, Iceland	WAYF	Connected (via eduGAIN).	Before CLARIN-PLUS	61
Estonia	TAAT	Connected (via eduGAIN).	2016	3
Finland	Haka	Connected (via national federation).	Before CLARIN-PLUS	47
France	Fédération Éducation-Recherche	Connected (via eduGAIN).	2017	276
Germany	DFN-AAI	Connected (via national federation).	Before CLARIN-PLUS	234
Greece	GRNET Federation	Connected (via eduGAIN).	2017	32
Hungary	eduID.hu	Connection planned since 2016.	(planned for 2017)	(27)
Italy	IDEM GARR	Connected (via eduGAIN).	2015	73
Latvia	LAIFE	Connected (via eduGAIN).	2016	18
Lithuania	LITNET fedi	Connected (via eduGAIN).	2017	8
The Netherlands	SURFconext	Connected (via national federation).	Before CLARIN-PLUS	31
Norway	FEIDE	Connected (via eduGAIN).	(connected, but opt-in issue, see next section)	
Poland	PIONIER.Id	Connected (via eduGAIN).	2016	12
Portugal	RCTSaai	Connection planned since 2015.	(planned for 2017)	(61)
Slovenia	ArnesAAI	Connected (via eduGAIN).	2015	13
Sweden	SWAMID	Connected (via eduGAIN).	2015	49
United Kingdom	UK Federation	Connected (via eduGAIN).	2015	546

Table 2: overview of the current state of the SPF for all CLARIN ERIC members. In total users from 1577 organisations can use single sign-on to login to CLARIN services.

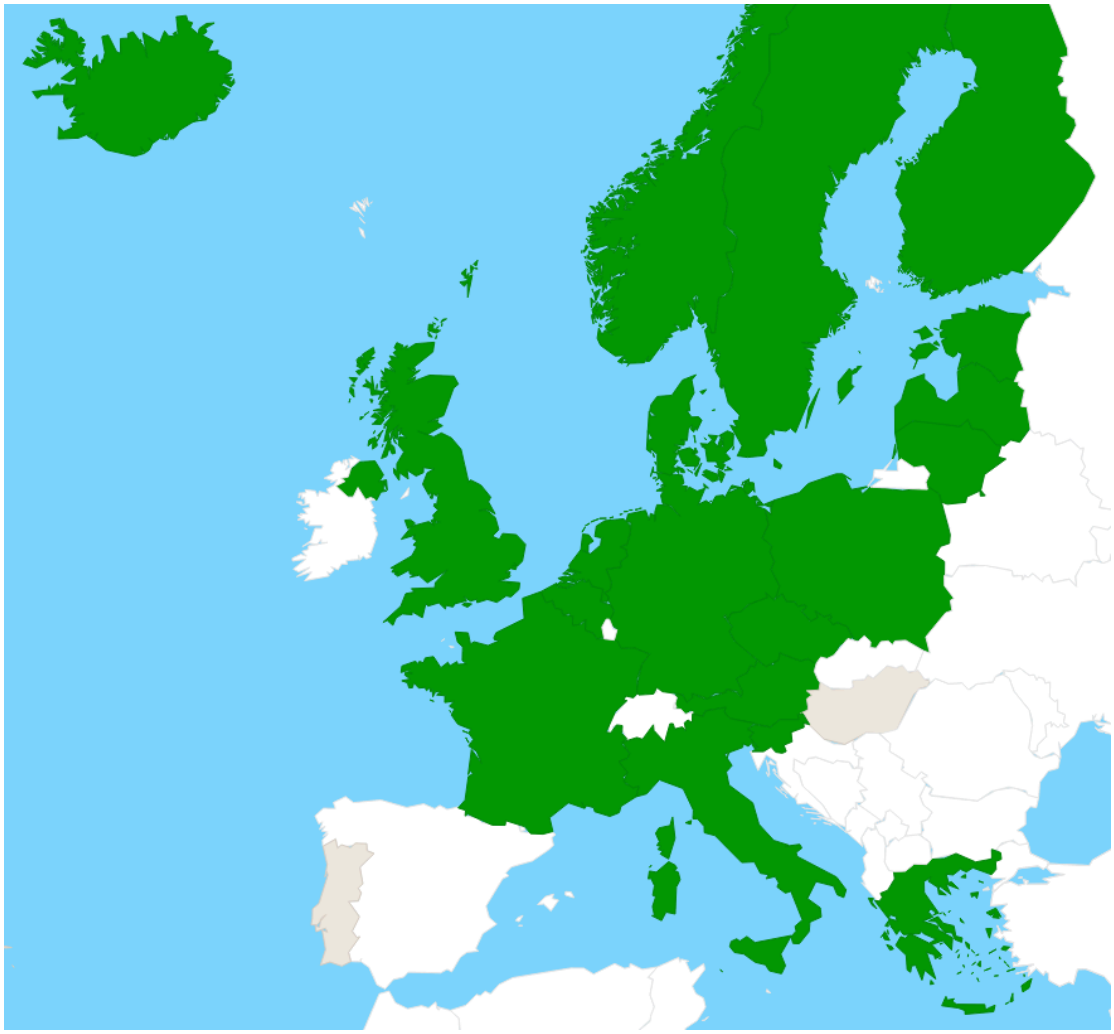

Figure 2: SPF coverage of European countries in April 2017. Green countries are included, grey ones are planned to be included shortly.

In addition to the national federations mentioned above, it should also be noted that Carnegie Mellon University (USA) has become an official CLARIN ERIC third party[5] in March 2017. Therefore it was also included into the SPF, illustrating one of the advantages of joining CLARIN ERIC on the institutional level: all users form that institution get easy access via single sign-on to the CLARIN services. Other potential third parties have indicated that SPF membership plays an important role as a factor of why to join CLARIN ERIC.

3.4 Additional services

In addition to the core functionality of Service Provider Federation – the formal construction that allows CLARIN ERIC to connect to the various national federations – there are also various additional services that are of importance to the CLARIN centres that are hosting a Service Provider. They all can be considered to deliver added value to the SPF through economy of scale.

[5] See https://www.clarin.eu/news/cmu-talkbank-sign-third-party-agreement-clarin

3.4.1 Single point of connection

After setting up a Service Provider, various federations require different steps to actually connect it to the national Identity Providers. These steps (e.g. SAML metadata submission) and related requirements (e.g. about SAML metadata) vary widely between federations.

Originally each SP administrator needed to understand these idiosyncrasies. It was concluded that having a central specialist at the side of CLARIN ERIC that would take the necessary steps for each federation is overall more efficient. Since then, local administrators only need to submit their SAML metadata to CLARIN, and from there on a specialist takes over all necessary steps of distribution to the national federations and to eduGAIN, including diagnosing problems should they occur.

3.4.2 Training, documentation and support

Even when the practical connection to the federations is made easier, setting up a Service Provider is a rather specialised technical task. Moreover, there are also few specialists in this field, so it is important to provide training[6], documentation[7] and support to the local system administrators.

CLARIN has been doing this through organisation of training sessions, the creation of documentation and by supporting the local technical staff.

3.4.3 CLARIN Identity Provider

Notwithstanding the steady growth of national federations, there are still many potential users that have no access to an institutional Identity Provider, either because there is no national federation (e.g. Bulgaria) or because they have no academic account (e.g. citizen scientists). In such cases there is a central fall-back identity provider[8] that is run by CLARIN ERIC, where account requests are individually verified on a best-effort base. This takes away the need for CLARIN centres to setup such a user store themselves. Moreover, even "homeless" users can use one account to access any CLARIN service.

A closely related task in CLARIN-PLUS focused among others on creating a more stable and maintainable version of the CLARIN Identity Provider [CLARINPLUS-D2.2]. As of April 2017 this improved version has been available in production.

3.4.4 CLARIN Discovery Service

CLARIN ERIC provides a central *discovery service*, a web application that is shown during the single sign-on procedure where the user can select the home institution – as illustrated in Figure 3. Using it is optional, but doing so means less work on the side of a Service Provider administrator. A lot of effort has been put into making it highly reliable (with a redundant setup) and user friendly (e.g. by optimizing the loading speed).

[6] See e.g. https://www.clarin.eu/event/2016/centre-meeting
[7] See https://cdn.rawgit.com/clarin-eric/SPF-tutorial/master/Shib_SP_tutorial.html and https://www.clarin.eu/content/creating-and-testing-shibboleth-sp
[8] See https://www.clarin.eu/content/clarin-identity-provider

Figure 3: screenshot of the CLARIN Discovery Service.

3.5 Issues and solutions

3.5.1 End of the Kalmar Union

The Kalmar Union is an interfederation initiative[9] among several Nordic countries that start in 2008. CLARIN was connected to the Kalmar Union via the Finnish HAKA federation. As such it could reach the Identity Providers from Norway, Denmark, Iceland and Sweden. Recently the news was spread that the Kalmar Union would be retired by the end of April 2017. It will be abandoned in favour of the eduGAIN interfederation. Since CLARIN is also connected to eduGAIN, and the Nordic countries all have a high opt-in rate, this will pose no problem.

3.5.2 Versioning of SAML metadata in various federations

National federations relay on various ingestion mechanisms for the SAML metadata. This can potentially lead to different versions that occur and worst case to conflicting versions. To make these differences clear, a tool[10] was developed that allows keeping track of all versions, which has proven to be very useful.

On the longer term the hope is that more and more national federations will use a centralized metadata distribution system (like eduGAIN's one), eliminating the different versions that are around now.

3.5.3 National opt-in policies

In some federations, each Identity Provider has to explicitly approve the connection to a specific Service Provider. This has serious implications for:

[9] See https://www.kalmar2.org
[10] See https://centres.clarin.eu/spf

- the scalability, since with M Service Providers and N Identity providers, M * N approval (opt-in) requests need to be handled
- the transparency, since it is unclear for a user which opt-in requests have been made and what the state of these is

In the following subsections below we illustrate the national opt-in issue for 2 federations where this has proven to be a serious issue.

3.5.3.1 SURFConext's opt-in policy

In the Dutch SurfConext federation an opt-in policy applies. Luckily SurfConext has agreed with CLARIN to only require a one-time opt-in from each Identity Provider to the whole block of CLARIN Service Providers. Though a little tedious, CLARIN-NL has organized this very well and as a result almost all Dutch universities and research organisations (currently 31) have access to CLARIN services, as illustrated in Figure 4.

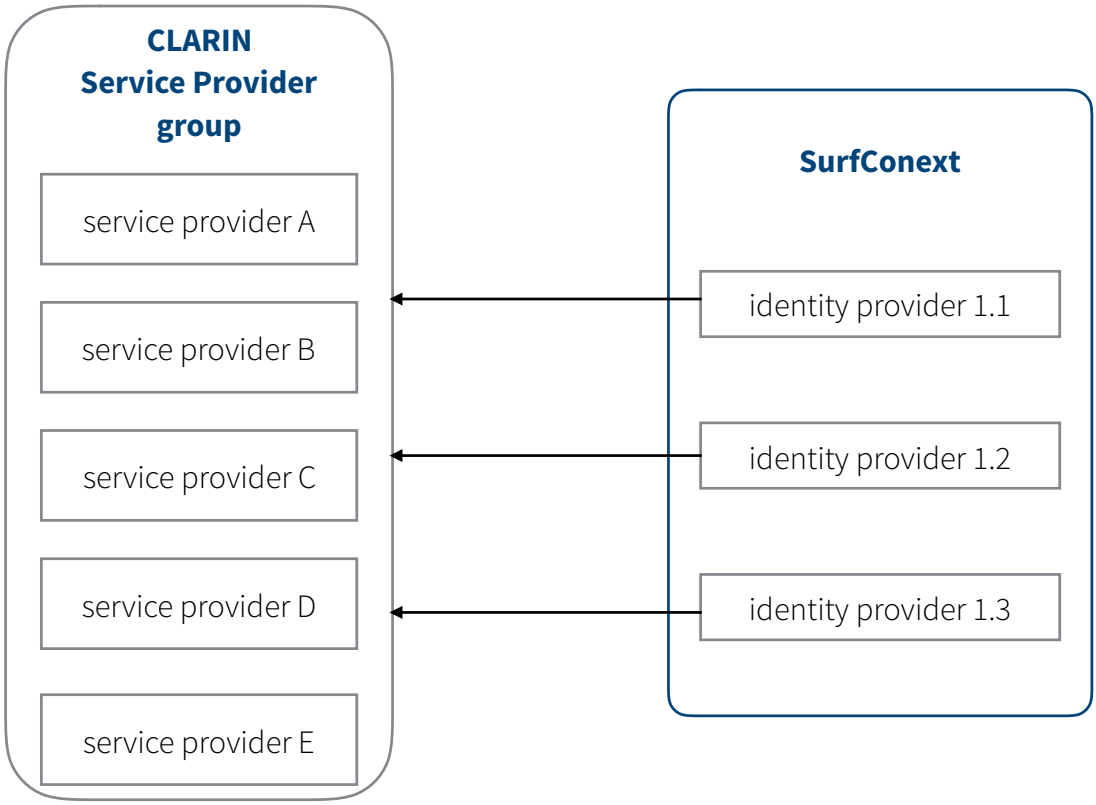

Figure 4: each identity provider agrees once to connect to the group of CLARIN Service Provider. Every arrow stands for a single opt-in operation.

3.5.3.2 FEIDE's opt-in policy

The Norwegian FEIDE federation has a strict opt-in policy: no connections are allowed from any Identity Provider to any Service provider, unless the Identity Provider has requested so. In practice this system does not scale at all[11], leading to a non-functional setup, as illustrated in Figure 5.

[11] There are currently around 30 CLARIN Service Providers and 484 Identity Providers in FEIDE. Connecting all of them would require 484 * 30 = 14520 opt-in requests. Even when only looking at a reasonable subset of all existing Norwegian Identity Providers (say 25) still 750 opt-in

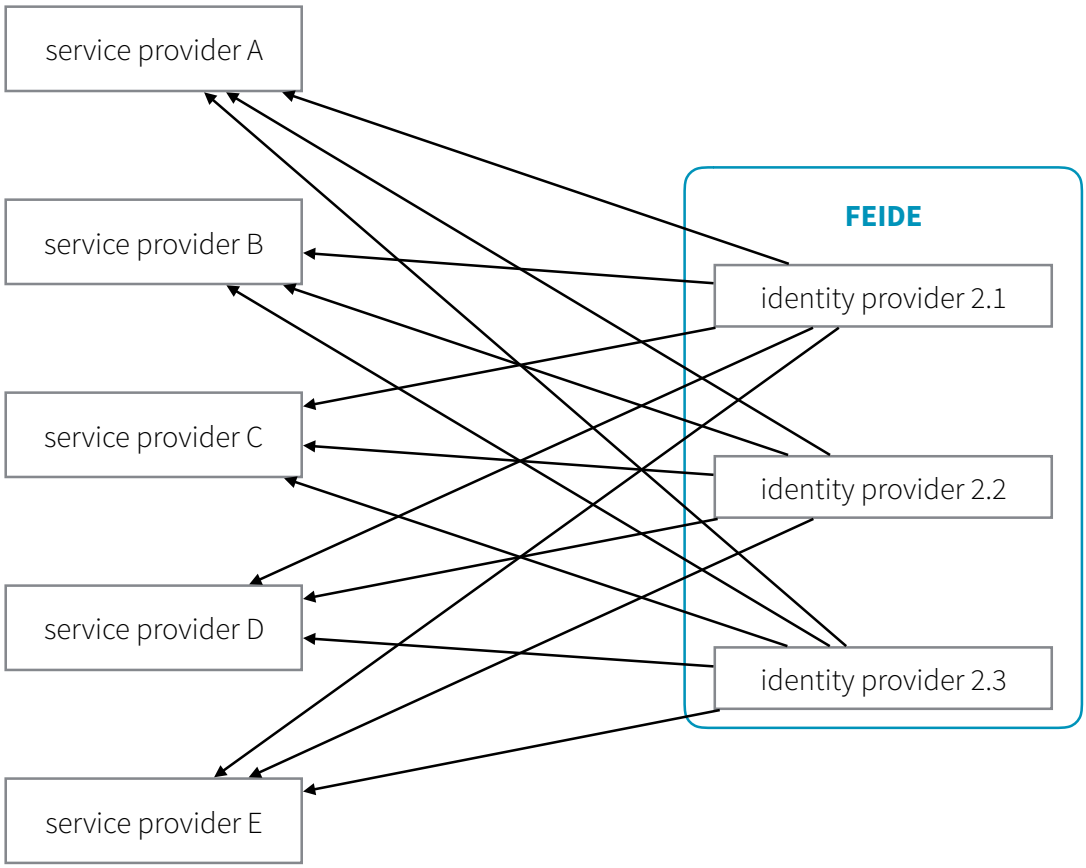

Figure 5: each single identity provider needs to approve the connection to each single CLARIN Service Provider. Every arrow stands for a single opt-in operation.

This is a long-standing problem in the FEIDE federation. The Norwegian CLARINO consortium is discussing this and looking for solutions together with the Norwegian Research Council and FEIDE.

A promising pilot that could potentially address the opt-in issue is Dataporten[12], which includes a dashboard that allows the Service Provider administrator to enable logins from eduGAIN Identity Providers. CLARINO plans to test this pilot service. If it works as expected, CLARIN ERIC could also connect to it.

3.5.4 Lacking attribute release

Some Identity Providers do not release enough (if any) attributes during the authentication procedure. Reasons cited for this behaviour are often lack of trust and privacy concerns.

requests need to be dealt with. From the experience with other federations this is not realistically achievable.

[12] See https://www.uninett.no/en/service-platform-dataporten

Although there is no easy fix for this problem, the following strategies are in place to ensure that the requested attributes are delivered by the Identity Providers:

- CLARIN is taking part in several initiatives to pledge it respects researcher's privacy and is complying with all relevant legislation. Most relevant in this respect is the Data Protection Code of Conduct[13].
- Through the Research and Scholarship[14] entity category CLARIN indicates it uses the personal attributes solely for research purposes.
- Thanks to work done in CLARIN-PLUS [CLARINPLUS-D2.2] statistics are kept about which Identity Providers are not releasing attributes. These are contacted with a polite request to provide the attributes.

The gathered statistics so far – see Table 3 – indicate that:

- the attribute release issue is not as big as it is sometimes pictured
- the problem varies widely among the national federations, with Germany being the most problematic case (7 out of 17 for UFAL, 6 out of 8 for HZSK, 1 out of 3 for CLARIN.SI)

	number of Identity Providers used	**number of Identity Providers not releasing personal attributes**	**percentage of Identity Providers not releasing personal attributes**
UFAL	165	17	10.30
HZSK	44	8	18.19
Korp (CSC)	35	0	0
CLARIN.SI	27	3	11.11

Table 3: the proportion of Identity Providers not releasing any personal attributes[15]. Overview for the four most popular services providers that are measured.

3.6 The future

With the growing number of CLARIN centres and CLARIN ERIC member countries we anticipate a growth of the Service Provider Federation too. To ensure the scalability and stability, the following factors will be kept in mind.

3.6.1 Technical architecture of the SPF

With over 30 Service Providers and 7 distribution federations (6 national federations and eduGAIN), the distribution of the Service Provider SAML metadata over these channels is by far the most cumbersome part of administering the Service Provider Federation.

Other research infrastructures and e-infrastructures (e.g. ELIXIR and EUDAT) have addressed this by setting up an infrastructure-wide proxy Service Provider. Instead of 30 entities, only 1 needs then to be registered with the 7 distribution federations. Changes to individual Service Providers can be arranged "behind the scenes".

[13] See https://www.clarin.eu/content/how-can-i-comply-data-protection-code-conduct
[14] See https://refeds.org/category/research-and-scholarship
[15] Source: https://lindat.mff.cuni.cz/services/aaggreg (state of 24 April 2017)

While this umbrella approach simplifies the overall administration, it introduces a new central software component, thus introducing a potential single point of failure and the need for additional software maintenance. This balance needs to be evaluated in more details before deciding on the way forward.

Two obvious candidates for the umbrella setup are UnityIDM[16] (as used by EUDAT and by the CLARIN Identity Provider) and OpenConext[17] (as piloted by the Dutch CLARIAH consortium).

3.6.2 Further extensions

As we did in the past, future extensions of the SPF will be planned taking into account the membership of CLARIN ERIC (for the prioritization) and the opt-in ratio for eduGAIN (for the choice between the national federation or eduGAIN). Table 4 lists the remaining candidates from the SPF agreement and their current status[18].

Country	Federation name	eduGAIN opt-in	number of Identity Providers
Hungary	eduID.hu	40.7%	27
Ireland	Edugate	70.5%	44
Portugal	RCTSaai	3.3%	61
Spain	SIR	100%	135
Switzerland	SWITCHaai	68.2%	66
USA	InCommon	95.2%	435

Table 4: Foreseen SPF extension candidates. Non-exhaustive list.

For a full overview of all existing federations in the world (in 53 countries) we refer to the REFEDS website[19].

The situation with regards to new potential CLARIN ERIC members is documented in [CLARINPLUS-D5.1].

3.6.3 Engagement with the national federation community

Maintaining good personal connections to GÉANT, the national federations and colleague research infrastructures is key for a successful collaboration. As obvious as it might sound, this really can help in remediating problems and planning future steps.

From that perspective, the following events can be instrumental to share experiences with the relevant groups:

- FIM4R workshops
- eduGAIN townhall meetings
- the TNC conference and satellite events

Attending at least some of these on a regular basis will help to understand the evolutions in the field of authentication and authorisation infrastructure.

[16] See http://unity-idm.eu/
[17] See https://openconext.org/
[18] Note that it does not include all European federations, since some (e.g. the Croatian one) had an opt-in rate of 100% when the updated SPF agreement was written. These federations did not need to be included.
[19] See https://met.refeds.org/

4 Conclusion

The Service Provider Federation has been a successful platform to establish effective interfederation single sign-on to CLARIN service providers. Over time it evolved to easily include national identity federations that have high opt-in ratios for eduGAIN. All in all this enabled the SPF to grow from 6 to 18 federations. This enlargement of the potential user base has also strengthened CLARIN ERIC's value proposition.

At the same time there are still issues that need to be handled, especially related to national opt-in policies and attribute release. The experience gained from and monitoring tools developed within CLARIN-PLUS will help to identify and address such stumble blocks.

References

[CLARINPLUS-D2.2] Mišutka, J. (2016). *D2.2 Robust SPF 1: workflow and monitoring* (CLARIN-PLUS Deliverables). Retrieved from https://office.clarin.eu/v/CE-2016-0809-CLARINPLUS-D2_2.pdf

[CLARINPLUS-D5.1] Maegaard, B., & Olsen, S. (2016). *D5.1 Analysis of the situation in non-CLARIN countries* (CLARIN-PLUS Deliverables). Retrieved from https://office.clarin.eu/v/CE-2016-0846-CLARINPLUS-D5_1.pdf

[CLARINPLUS-D5.4] Maegaard, B., Van Uytvanck, D. & Krauwer, S. (2016). *D5.4 CLARIN Value Proposition* (CLARIN-PLUS Deliverables). Retrieved from https://office.clarin.eu/v/CE-2016-0847-CLARINPLUS-D5_4.pdf

[CLARINPP-D2R-3b] Broeder, D., Van Uytvanck, D., Wittenburg, P., Bruun, S., Jansen, S., Kupietz, M., … Zimmer, K. (2010). *D2R-3b Language Resource and Technology Federation Building - v2* (CLARIN Preparatory Phase Deliverables). Retrieved from http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-32