



25-04-2017

Deliverable DNA3.3: Recommendations for Research and e- Infrastructures to Build Sustainable Services

Deliverable DNA3.3

Contractual Date: 28-02-2017
Actual Date: 25-04-2017
Grant Agreement No.: 653965
Work Package: NA3
Task Item: 3
Lead Partner: DAASI International
Document Code: AARC-DNA3.3
Authors: L. Florio (GÉANT), P. Gietz (DAASI International), H. Hütter (DAASI International), M. Haase (DAASI International), D. Hübner (DAASI International), P. van Dijk (SURFnet), J. Jensen (STFC), D.L. Groep (Nikhef), Christos Kanellopoulos (GÉANT)

© GÉANT on behalf of the AARC project.

The research leading to these results has received funding from the European Community's Horizon2020 Programme under Grant Agreement No. 653965 (AARC).

Abstract

This document provides guidelines to enable sustainability for services in the research and education community and presents a template to assess existing services. Furthermore, it provides policy recommendations for service providers and federations and identifies strategies and risks of enabling guest identities for services.



Table of Contents

Executive Summary	1
1 Introduction	3
2 Considerations for Building Sustainable Services	4
2.1 Use Case and User Base	4
2.2 Operator	5
2.3 Sponsor / Funding Model	6
2.4 Governance, Policies and Processes	6
2.5 Service Implementation	7
2.6 Analysis of First Responses to the Template	8
3 Policy Recommendations	10
3.1 Research and e-Infrastructures Service Providers, IGTF and Related Activities	10
3.2 National Research and Education Identity Federation operators, REFEDS and eduGAIN	13
4 Strategy and Risks of Using Guest Identities	17
4.1 Enabling Guest Identity Access to Services	17
5 Conclusions	20
Appendix A First Responses on the Template from Selected AARC Pilots	22
A.1 AARC RCauth.eu	22
A.2 DARIAH Guest IdP	27
A.3 Social IDs (to SAML) pilot	30
A.4 WATTS pilot	33
Appendix B Managing Risks of Using Guest Identities	36
References	38
Glossary	39



Executive Summary

In a world increasingly reliant on digital content and online services, long-term sustainability is a critical aspect. Software and services are created and operated within the context of a project and its funding cycle. Once a project ends, it becomes difficult to ensure the necessary support to run services. Research and e-infrastructures in particular need to look at the longer-term sustainability beyond project funding and define a scalable way to ensure that necessary resources are available for those services that are relevant to their community.

From May 2015 to end-April 2017, the Authentication and Authorisation for Research and Collaboration (AARC) project has worked with research and infrastructures, libraries and NRENs to deliver an integrated cross-discipline authentication and authorisation framework and to test the integration of some of the framework components in production environments.

This document summarises the project's:

- Recommendations for research and e-infrastructure operators, as well as for service operators operating within these infrastructures, to build sustainable services that follow well-defined policies and practices.
- Recommendations to federation operators to better meet research and e-infrastructures requirements and to facilitate the adoption of federated access.
- Considerations on the usage of guest identity solutions for those relying on them.

This document has been organised in three sections:

- Section 1 of the document starts with providing the current policy landscape in the R&E sector. One of the main challenges for the research and education community as a whole is to ensure that successful services can be operated and supported beyond the funding cycles. A cost-recovery analysis, defined in the early stage of the service delivery, influences how the service is deployed. This section offers guidelines and templates that aim to ease international scientific collaborations and e-infrastructures to operate services in a sustainable way. These guidelines are based on the experience of the pilots carried out in the AARC project.
- Section 2 presents AARC recommendations to:
 - Research and e-infrastructure service providers operating within research and e-infrastructures to follow standardised approaches.
 - Federation operators to streamline policies and best practices to make the adoption of federated access technologies easier for international research communities and e-infrastructures. The proposal is based on the requirements gathered within research and e-infrastructure communities represented in the AARC project. This approach was initiated by the FIM4R community in 2012, and has proven to be very effective.



- Section 3 focuses on strategies and risks associated with enabling guest identities. In order to make service in general accessible to all targeted user groups, it is important to allow users without an identity at a home organisation that is part of eduGAIN or another federation to access the service. Broadening the access to services also has implications with respect to sustainability. This section provides recommendations on how to implement guest identities and points out the major aspects service providers should consider.



1 Introduction

One of the goals of the Authentication and Authorisation for Research and Collaboration (AARC) project is to improve the adoption of federated access in the research and e-infrastructures. A scalable way to achieve this goal is to identify clear requirements from the research and e-infrastructures communities, explore existing solutions or test new approaches within the AARC project. Further to that, to propose technical guidelines and policy best practice for research and e-infrastructures and to make recommendations for federation operators to better support community requirements.

In a world increasingly reliant on digital content and online services, long-term sustainability is a critical aspect. Software and services are created and operated within the context of a project and its funding cycle. Once a project ends, it becomes difficult to ensure the necessary support to run services. In particular, research and e-infrastructures need to look at the longer-term sustainability beyond project funding and define a scalable way to ensure that necessary resources are available for those service that are relevant to their community.

The chosen operational model for a service has a huge impact on its implementation, as well as on the costs and efforts necessary to sustain the operation of the service. Decisions on how to maintain the code and how to recover the operational costs should happen as early as possible in the service development process. There are different factors that have an impact on selecting an approach.

When identifying target user groups, service providers should think about users without a home organisation or with a home organisation that does not belong to a federation. Service providers can enable guest identities to access their service and enable a larger user base, which can help achieve greater sustainability.



2 Considerations for Building Sustainable Services

This section provides guidelines and templates for use in international scientific research collaborations and e-infrastructures to follow best practice and implement AARC frameworks to operate services in the most sustainable way. Examples Filled in examples as initial responses to the templates from AARC, are provided in Appendix A.

One of the main challenges for the research and education community as a whole is to ensure that successful services can be operated and supported beyond the funding cycles. A cost-recovery analysis has to be defined in the early stage of the service delivery. Five critical aspects have been identified and are presented in this section for each of aspect there a number of questions that are proposed, which form a reference template:

- Use case and user base
- Operator
- Sponsor/funding model
- Governance, policies and process
- Service implementation.

2.1 Use Case and User Base

Considerations in this area focus on the service itself and the target user group. In light of scarce resources, it is not worth embarking on new development unless there is a demonstrated need for it. The requirements that lead to the services should determine if federated access is a requirement. In today's world, however, federated access is an important aspect, as users are accustomed to leverage existing credentials to access services beyond the context of their research.

Participation in eduGAIN, the interfederation services, is hierarchical: institutions (or home organisations) and services join a national federation, which, in turn, participate in eduGAIN. Different federations have slightly different processes and policies. In general, although there is no geographical restriction, services tend to join the R&E federation in their own country, but nothing prevents a service from joining a federation located in a different country.

Some services are expected to deal with sensitive data, such as medical services. These requirements have implications not only on the service implementation but also on its policies.

Use-cases and Users
<i>What is the service about?</i> <i>What problem does it solve?</i> <i>Which user group is it aimed at?</i> <i>What are the typical use-cases?</i>
<i>Estimated user base</i>
<i>Are there already similar services? If so, what does this new service add?</i>
<i>Is the service to be developed or procured?</i>
<i>Is federated access a requirement?</i>
<i>Is there sensitive data that the service needs to support?</i>

2.2 Operator

In many cases, services are piloted as part of funded projects, but by the time the projects end, there is already a user base that is significant enough to merit continuation of the service. It is not uncommon that service operations are moved to a different operator after the pilot phase and/or that multiple instances of the service are operated by different parties. Questions to address at this stage include: *What are the operational requirements? Has the pilot highlighted some weakness? What type of support does the service need? Is the service procured and provided by a third party or is it developed in house? How will the service be promoted?* A candidate service operator with the necessary characteristics should be identified as soon as possible. The choice of the operator may have implications on the overall costs, as well as on the service delivery model.

Operator Choice
<i>List the potential operators</i> <i>Which operator is best suited for the service?</i>
<i>Who is going to support and train the users?</i>
<i>Who should be responsible for the promotion of the service? (optional)</i>
<i>Costs estimation:</i>



What are the operator's expected costs (in terms of cost of hardware/software and effort):

- *For bootstrapping the service*
- *for the annual operations*

When possible, find a key parameter that drives the costs (e.g. number of users, storage, hardware, procured service, software etc.) and state the constraints of your implementation

2.3 Sponsor / Funding Model

There are different options to fund services. Typically, in the research and education community cost recovery models are never implemented by directly charging the end-users. This is due to several reasons, one of which is the lack of a real mechanism in place for this to be efficiently implemented. There are discussions on how to enable communities to cover service costs for their users, but for the moment, no single procedure can be recommended. Typical questions in this area include: *How are the service costs recovered? What features are really needed? For how long can the initial funding last?*

Sponsorship and Funding

What are the plans for long-term cost recovery?

In general, the following models are possible:

- *Each organisation and research community pays for the service usage.*
- *Established e-infrastructure or research infrastructures run the service; normally they have already a framework in place to recover costs.*
- *Third-party funding.*

What are potential risks in service operation? Who bears these risks - operator or sponsor?

2.4 Governance, Policies and Processes

When deploying a service, different aspects should be considered that have an implication on the policy, for instance, if SLAs are needed, requirements on monitoring, accounting and data storage, etc.

The AARC project has defined a policy template in [\[DNA3.5\]](#) to provide recommendations and template policies to resource providers and user communities that establish and operate infrastructure components. These recommendations are intended to facilitate the ability to collect, transfer, provide access to, and/or publish data related to the accounting, monitoring, logging, or any kind of processing of personal user data needed for the operation of the services provided by the resource providers.

Governance, Policies and Processes
<p><i>Are there specific policy aspects that should be taken into consideration? What are the security requirements?</i></p>
<p><i>Is there sensitive data stored by the service?</i></p>
<p><i>What are the availability requirements?</i></p>
<p><i>Are Service Level Agreements (SLA) necessary or expected?</i></p>
<p><i>What are the monitoring and accounting requirements?</i></p>
<p><i>What are the documentation requirements (such as user documentation, tutorials, administrator documentation and installation documentation)?</i></p>

2.5 Service Implementation

This area focuses on the actual implementation of a service. Its purpose is to evaluate the technical feasibility of architectural decisions. These decisions might have an influence on all other aspects of the service previously mentioned and require careful consideration.

Service Implementation
<p><i>What is the current architecture? Are there dependencies with external tools/software/licences? What are conceivable deployment and operational scenarios?</i></p>
<p><i>How many elements compose the service? (provide schematic if possible, see example)</i></p>
<p><i>Technical requirements (VMs, storage, network...)</i></p>
<p><i>Estimation of sustainability of software being used?</i></p>



2.6 Analysis of First Responses to the Template

The considerations proposed above were tested with three different pilots conducted in AARC, namely, RC Auth.eu, Social IDs pilot and WaTTS pilot. Additionally, DARIAH-DE Guest IdP was also matched against the same criteria. All of these pilots and their associated services have very diverse levels of maturity regarding their operational model. Completed tables can be found in Appendix A.

The AARC CI-Logon-like pilot is an AARC token translation service to leverage federated authentication, to generate eScience certificates. These certificates, in combination with VOMS attributes, are then used to access non-web resources offered by several research- and e-infrastructures. The RCAuth.eu is the underlying certificate authority (CA) that issues certificates to end-entities based on a successful federated authentication. A sustainability study was conducted in [\[SUSTAIN\]](#) that resulted in different deployment scenarios and a recommendation to operate the service, at least in a pan-European environment, either as a jointly procured single service, or a distributed service offered collectively by a consortium of e-Infrastructures. Both the “Use Case and User Base” and the “Service Implementation” area, are well defined for this pilot, and are in an advanced state of development. The “Governance, Policies and Processes” area is worked out, as far as it is possible, from both a general and an abstract point of view. The operational model is not completely finalised, but there is a clear indication that to build a pan-European service, the RCAuth.eu should be operated by experienced parties and offer redundancy capabilities.

The Social IDs pilot leverages social IDs (such as Google and Facebook) as a way to enable access to ‘guest’ users [\[SOCIALID\]](#). The pilot is meant to support individual users that are not affiliated with any of the traditional home organisations, as well as those users whose identity providers (IdPs) are not part of any of the eduGAIN participating federations. The key factors in enabling such guest identity services are to be able to support multiple technologies and flexible policies in a scalable and trustworthy manner. The pilot is based on an **SP-IdP proxy** architecture, see also [AARC Blueprint Architecture \[BLUEPRINT\]](#), through which users are able to authenticate with the credentials provided by the IdP of their home organisation (via eduGAIN), as well as using social Identity providers, or other selected external identity providers.

Specifically, the proxy has built-in support for SAML, OpenID Connect and OAuth2 providers, and enables user logins through Facebook, Google, LinkedIn, and ORCID. The proxy is then responsible for enriching the identity information that comes from these external IdPs with additional attributes.

WaTTS is a Token Translation Service developed by Karlsruhe Institute of Technology (KIT), in the context of the INDIGO Data Cloud project [\[INDIGO\]](#). WaTTS was developed to address the users’ needs to access services that cannot directly utilise federated access and require that the users use security tokens, such as SSH keys, X.509 certificates, S3 access tokens etc. In this AARC pilot, WaTTS is integrated with the EGI CheckIn service, so that users can access WaTTS using their EGI accounts, while authenticating, either at their home organisations or using their social IDs [\[EGI\]](#).

With WaTTS, users are able to manage the SSH access to a number of trusted VMs from a single point in a secure and user-friendly manner. In this pilot, WaTTS is used to manage their SSH public keys and provision them on demand to an authorised set of VMs. Although in this case, WaTTS is integrated with the EGI CheckIn service, the solution is not limited to EGI, and can be used at any community/infrastructure/service where there is a need to “bridge” between different technologies, and can also be run as a standalone “plug-and-play”



solution. The only requirement is that the community/infrastructure/service supports integration of OIDC services.

The DARIAH-DE Guest IdP is a so called “last-resort IdP” for users that want to access DARIAH-services but are not affiliated with an institution operating an IdP within eduGAIN, or whose home IdPs release no or too few required attributes to enable access to DARIAH services [[DARIAH](#)].

Users that can prove their affiliation to the target user group (of researchers / scholars in the field of Digital Humanities), can get a dedicated DARIAH-DE account to access the DARIAH-DE services. Additionally, the DARIAH-DE Guest IdP handles permissions for all known identities, both for dedicated DARIAH-DE accounts, as well as federated identities (*Users and User Base*). The service has been running in production for several years (*Service Implementation*) and is currently in the transition from being operated within a project to being operated by an organisation, as DARIAH-DE itself is in the process to become a legal entity. The cost-recovery aspect for the guest IdP is managed through the DARIAH e-Infrastructure Service Unit (DeISU). At present, the operations of the guest IdPs are funded via the DARIAH EC-funded project resources, and going forward, it will be financed by a the German federal government and its 16 states (*Sponsor*). New requirements emerge over time, which require adjustments to the service operational model accordingly.

The responses to the pilots shows a very common phenomenon. In early project phases the focus is on “*Users and User Base*” and “*Service Implementation*”. Only at later project stages do the other areas get addressed. This stands in contrast to the impact the other areas have on the overall service operation, especially if considered over the whole service lifetime. Only by addressing all areas early, and developing the service operational model alongside the service, can the risks of exceeding operational costs or even shut down of services be contained.



3 Policy Recommendations

There are different, but related, places where policies are defined for the benefit of the global Research and Education sector, namely:

- Research and e-infrastructure service providers, Interoperable Global Trust Federation (IGTF), FIM4R and related activities.
- National Research and Education Federation operators (REFEDS) and eduGAIN.

3.1 Research and e-Infrastructures Service Providers, IGTF and Related Activities

This section offers a number of recommendations for service providers operating within research and e-infrastructures and for the research infrastructures (RIs) and the e-infrastructures (EIs) themselves to ensure they implement their technical and policy framework in a future proof way.

Some of the policies for research and e-infrastructures have been traditionally addressed within the Interoperable Global Trust Federation (IGTF). The IGTF defines common policies and guidelines that help establish interoperable, global trust relations between providers of e-infrastructures, cyberinfrastructures, identity providers, and other qualified relying parties.

Core work for the IGTF was to establish a set of identity credential providers (traditionally x.509 certificates issuers) that could be trusted by the research organisations and e-infrastructures. More recently, the IGTF has produced a technology-agnostic assurance level that represents the IGTF consensus on trustworthy authentication from the relying party's point of view, while still achievable from the identity providers' view, covering a variety of scenarios.

Since 2015, the AARC project is working together with several RIs and EIs to address their need to use federated access more widely, however, some of their requirements that characterise international research collaborations go beyond present-day federated access capabilities. AARC has worked on both the technical and policy sides. The technical work focused on the definition of a blueprint architecture; the latest version to be released in April 2017. AARC is also finalising production-ready architectural building blocks, best practices and is contributing to the definition of policy frameworks to enable research collaborations and e-infrastructures to build interoperable authentication and authorisation infrastructures (AAls) and integrate them into their production environment.

Many research and e-infrastructures have to support users that do not belong to any federations. AARC has also tackled this aspect and produced specific recommendations in other documents [\[RECOMMEND\]](#).



A set of recommendations for research and e-infrastructures (EIs/RIs) to ensure they build services that are accessible by as many users as possible, in the most cost-effective way, is listed below:

- Users should be able to access services in EIs/RIs using the credentials they have in their home organisations.
 - **Description** – Mechanisms should be in place to enable users that already have credentials in their home organisation (which is connected to a national identity federation) to access EIs/RIs' services, without the need of any additional set of credentials. This implies that EIs/RIs services are also accessible via eduGAIN, either directly or via proxies.
 - **Benefits** – EIs/RIs do not have to re-implement costly identity vetting processes, which are already in place at the home organisations of the users. Users do not have to maintain multiple accounts and can use the same account they have at their home organisation in order to participate in international research activities. AARC offers the CI-Logon-like pilot (to leverage federated access to generate eScience certificates).
- EIs/RIs should adopt the AARC Blueprint Architecture when implementing federated access across a number of internal services [BLUEPRINT].
 - **Description** – EIs/RIs often provide a large number of services to their users. By adopting the AARC Blueprint Architecture, EIs/RIs can maintain one integration point with eduGAIN and the national identity federations, the IdP-SP proxy, through which users can access all the EI/RI services, without having each internal service provider joining eduGAIN separately.
 - **Benefits** – EI/RIs can take full advantage of eduGAIN and the national academic federations, while still being able to have full control of their administrative and technical domains. Leveraging eduGAIN for federated access can be a costly, time consuming and error-prone endeavour if each service had to implement its own policies, technical stack for federated access, and join eduGAIN.
- Service providers that participate in eduGAIN should support the GÉANT CoCo, whenever possible.
 - **Description** – The GÉANT Data Protection Code of Conduct (CoCo) is a data privacy policy which is in line with current data privacy legislation. A service provider can easily link to this policy within its metadata. EIs/RIs that have adopted the AARC Blueprint Architecture need to support CoCo centrally on the IdP-SP proxy.
 - **Benefits** – IdPs will be more likely to provide the user attributes needed by the research and e-infrastructures, e.g. a permanent ID to link with other possible accounts, and an email address to communicate with the user.
- Service providers that participate in eduGAIN should comply with the Security Incident Response Trust Framework for Federated Identity (SIRTFI).
 - **Description** – SIRTFI aims to enable the coordination of incident response across federated organisations. SIRTFI compliance can be expressed in SP and IdP metadata. EIs/RIs that have adopted the AARC Blueprint Architecture, need to declare SIRTFI compliance centrally, on the IdP-SP proxy.



- **Benefits** – Support of SIRTFI increases the overall security and thus trust between federation partners.
- Service providers that operate in the research and education sector that participate in eduGAIN should apply for R&S entity category.
 - **Description** – Research and Scholarship (R&S) Category is intended for Service Providers that are operated for the purpose of supporting research and scholarship interaction, collaboration or management. Compliance to R&S has to be applied for, and can then be expressed in SP metadata.
 - **Benefits** – By asserting to be member of the global research community, again it might be easier for an SP to retrieve needed personal attributes from IdPs.
- EIs/RIs that have adopted the AARC Blueprint Architecture should implement the SNCTFI policy framework [[SNCTFI](#)].
 - **Description** – The Scalable Negotiator for a Community Trust Framework in Federated Infrastructures (SNCTFI) proposes a policy framework that allows determination of the 'quality' of SP-IdP proxies and the community of SPs behind the Proxy
 - **Benefits** – by implementing the SNCTFI policy framework, assertions made at the proxy level (e.g. CoCo SIRTFI or R&S compliance) can be transferred to the SPs behind the proxy.
- Service providers within RIs/EIs that require specific level of assurance, should monitor the development in the REFEDS Assurance WG, where such a framework is being discussed.
 - **Description** – In this REFEDS work, several dimensions of assurance levels (ID uniqueness, identity vetting, authentication methods and data currency) are being harmonised in at least two profiles.
 - **Benefits** – By implementing such profiles, rather than evaluating the single dimensions, services can take finer-grained access control decisions (e.g. on the level of traceability required). By choosing to require one of the few pre-defined levels, services can concisely communicate their requirements to IdPs.
- EIs/RIs should be prepared to manage users relying on social IdPs.
 - **Description** – Some community target groups have users that cannot rely on federated credentials via eduGAIN, either because they are not affiliated to any organisation at all (citizen researcher), or their organisation is not federated, or is not properly federated. This is also discussed in more detail in Section 4 of this document.
 - **Benefits** – Relying on social providers is in the case described above a simpler and less costly alternative to managing and maintaining a dedicated guest IdP.
- EIs/RIs should follow REFEDS discovery guidelines.
 - **Description** – These guidelines detail in simple steps how to implement federated login in a way which protects branding, improves user satisfaction, and increases successful logins.



- **Benefits** – A good login interface improves the ability of users to access resources. It is affordable and easy to maintain.

The implementation of a number of these recommendations will be much easier, if the proxy-approach recommended in the AARC Blueprint architecture is being followed.

3.2 National Research and Education Identity Federation operators, REFEDS and eduGAIN

National identity federations to date are built using the standardised SAML technology; most of them have a mesh or hub and spoke architecture, although hybrid approaches are on the rise, as shown by the REFEDS 2016 survey [[SURVEY](#)]. Aside from the technical aspects, each federation has a policy in place that defines the behaviours of the federations' participants (both service providers and identity providers, also called home organisations or IdPs).

Research and education Identity federation operators cluster in REFEDS, the international forum to articulate their needs. Most of the REFEDS participants' requirements are about policies and best practices. Over the last ten years, REFEDS has worked with federation operators to harmonise national policies and deliver agreed common practices. This is particularly useful when national federations interact with each other and when they participate in eduGAIN, the global inter-federation service [[eduGAIN](#)].

eduGAIN imposes some lightweight requirements to participating federations. Some of the service providers in the international collaborations have asked for eduGAIN to be more restrictive and to increase the entry point requirements. This approach is, however, not possible, as eduGAIN caters for diverse user groups, some of them with less-stringent requirements. The consensus is, therefore, to build additional 'layers' on top of eduGAIN to meet various provide access to a greater range of groups/users.

The existence of REFEDS makes it easy to gather the necessary insight on current and planned policies. The AARC project liaises closely with REFEDS on these topics, although AARC focuses on the point of view of the research and e-infrastructures providers.

Whilst the technical implementation is well-understood and outside of the scope of this document, this document only addresses aspects that have implications for service providers in the context of international research collaborations. It is worth noting that a service run by a research community using federated identity management is bound to the federation operator policies and practices, as well as having to address the policy requirements of that specific collaboration.

Identity federations, as all institutions, research and e-infrastructures, and service providers, have to comply with data protection. The current EU data protection framework (Data Protection Directive) will soon be replaced by the General Data Protection Regulation (GDPR) which will enter into force in May 2018. This is a particular complex space, due to the fact that the implementation of the GDPR will change some aspects of the current directive and its adoption in the member states is still being discussed.



Federated identity management is recognised by regulators as a privacy-enhancing tool. However, the three-sided relationship between user, identity provider and service provider does not obviously fit the models provided by European data protection law. Neither the identity provider nor the service provider satisfies the legal definition of a data processor. As a result, no contract may be drawn up between the identity provider and service provider when the user chooses (and agrees a contract) with an identity provider. Instead, it is better to view the arrangement in terms of an individual instructing two parties – each an independent data controller – to transfer personal data. This would normally fall within legal provisions for Consent (Data Protection Directive [D95/46/EC] Article 7(a)). However, if a user needs to access a particular service as part of his/her research, then it is not clear that the consent can be freely-given, as the law requires. To avoid the complexities of applying different legal regimes to different requests, Research and Education federations in Europe have generally considered that both identity providers and service providers process personal data in their **Legitimate Interest (Article 7(f))**, providing the service that an individual has requested from them. This allows each to focus on the relationship with their user, rather than having to collude to try to establish the appropriate legal regime for each individual request.

Whether using Legitimate Interests for national or international transfers, EC rules require users to be informed of the release of information and the interests that it serves (Article 14 of the Data Protection Directive).¹ Federated services and identity providers already use a number of different mechanisms to provide information to their users, so any additional information requirement should not be onerous. Federation operators have developed recommendations for some aspects of user interfaces in [FED-BP] – as requirements under the new Regulation became clear there may be an opportunity for further work to develop standards in this area.

The proposed recommendations are meant to improve the overall user experience and facilitate the adoption of federated access.

- Federation operators should agree to support eduPersonUniqueID.
 - **Description** – This is a long-lived, non-re-assignable identifier, suitable for use as a unique external key by applications. Values of this attribute must be assigned in such a manner that no two values created by distinct identity systems could collide. Many services operated by research and e-infrastructures require a unique, non-reassigned and persistent identifier. Whilst non-reusable persistent identifiers such as SAML2 NameIDs with persistent format, eduPersonTargetedID and eduPersonUniqueID are basically available in all federations (some federations even mandate them), it is not clear which one of the three will be released by which federation.
 - **Benefits** – It would be useful if federation operators could all agree to support eduPerson unique ID.
- Federation operators should be caution in filtering eduGAIN metadata.
 - **Description** – eduGAIN publishes the bundle of metadata downstream for the benefit of federation operators. Most of the federations consume this bundle either as such or filtering out some entities. Filtering to support security and other interoperability issues is recognised as important, however, some federations may apply filtering for other things. In the case in which filters are applied or federations create a curated eduGAIN stream, the user experience may be negatively impacted.

¹ See also <https://community.jisc.ac.uk/blogs/regulatory-developments/article/federated-access-management-and-gdpr>



- **Benefits** – Knowing that all services are available to all federations participating in eduGAIN in the same way improves the user experience and expectations.
- Federation operators should promote adoption of R&S entity category.
 - **Description** – The main challenge for service providers is about getting attributes from the home organisations' (IdPs. REFEDS has standardised an approach called entity category, which enables services to be grouped into categories; identity providers then release a defined number of attributes for that category. Federation operators are responsible for tagging services in a category as well as for promoting the approach to participating IdPs. Even if this approach has been used for many years and even if IdPs are tagged as supporting entity category and SPs are tagged into a category, there are cases in which attributes are not correctly released. It is clearly not scalable for service providers to debug the problem and have bilateral conversations with the IdPs they need to interact with. Hub and spoke federations have the possibility to influence this process better than mesh federations, as all communications go via the hub, which can then complement missing attributes.
 - **Benefit** – Such an approach, if widely adopted, would ensure that service providers get the attribute they need, in a more scalable way.
- Ensure that home organisations do not reassign user identifiers.
 - **Description** – Due to the complex issues of managing legacy systems within organisations, re-use of identifiers (such as eduPersonsPrincipalName or ePPN in short) can occur within federations. Because ePPNs are widely used, reassigning them can lead to problems when the new owner of the user identifier will receive the account of an existing user at an SP, or worse, at many SPs behind some national hubs.
 - **Benefits** – the SPs receiving an ePPN do not need to have an algorithm in place that invalidates user identifiers.
- Promote and support participation to SIRTFI to handle incident response.
 - **Description** – With the wider usage of digital identity, security has now grown to also encompass security incidents in identity federations and eduGAIN. This work is being addressed by the Security Incident Response Trust Framework for Federated Identity (SIRTF) Working Group, hosted by REFEDS and also sponsored by the AARC project. In January 2016, version 1.0 of SIRTFI was published via REFEDS following community consultation practice. The challenge is to persuade IdPs and other type of services to see the benefit of SIRTFI and comply with it.
 - **Benefits** – A joint approach to handle incidents can only work properly with as many federations as possible participating in it.
- Use eduGAIN to create a support help desk.
 - **Description** – There are many questions federation operators might have when joining eduGAIN in the first place and also when they are already part of the interfederation.



- **Benefits** – Federation operators would know who to address with any kind of questions. eduGAIN could consolidate requests, manage them and delegate them to experts in a structured and automated way.



4 Strategy and Risks of Using Guest Identities

In this document “guest identities” refer to identities provided to the users by entities other than the home organisation of the user. Guest IdPs are needed to provide access to:

- Nomadic users (those without a “home” organisation, such as “long-tail” researchers),
- Citizen scientists
- Users belonging to an institution that does not operate an Identity Provider (IdP), or one which operates an IdP that is not part of eduGAIN.

Clearly, the use of guest IdPs (whether provided by a third party or whether self-managed) has an implication on the service operations model and on the costs.

4.1 Enabling Guest Identity Access to Services

There are several ways in which infrastructures may offer guest identities.

One of the obvious options would be for RIs and EIs to deploy their own guest IdPs. This is, however, not a recommended option, as operating guest IdPs comes with recurring costs associated with maintaining the technical components, as well as appointing dedicated people to take care of the curation of identities, and to ensure that proper policies and procedures are followed for the guest IdP to be trusted.

In many cases, they might need to find a legal entity which is able and willing to take over the operational and legal obligations that accompany joining a federation. While long-term (e.g. ESFRI) projects are more likely able to deal with these issues, this may be more challenging for (non- or loosely organised) research communities and smaller/short-term projects.

For the reasons above, we propose to EIs/RIs to assess the following options for supporting guest identity services:

- Social media – a list of identity providers, such as Facebook, Google, LinkedIn – that aim to provide user-specific identities, but allow anyone to sign up.
- Government (eGOV, eIDAS) and banking (although at the moment these are not widely deployed for international collaborations).
- Commercially provided identities - for example, trusted third parties that can be contracted to offer this service.

Each of these options have advantages and disadvantages, and they do not have to be mutually exclusive. Even the option of a community-operated guest identity provider service can sometimes be justified. For further details on the pros and cons of each of the above approaches, please see [\[MJRA1.2\]](#).

We have identified the following principles for services within EIs/RIs when deciding how to support guest identities:

- Users may have credentials issued by a community-managed guest IdP. It may be possible to leverage these credentials to support guest users.
- Users may have an ORCID identifier, created to support research and publications. Services may consider using ORCID as a possible guest IdP (it provides an identifier attribute).
- Different guest IdPs should be supported to ensure that more users can access the service.
- At the moment it is not possible to widely rely on eGov IDs (or eIDAS) as the level of deployment is very different among countries.
- It is not possible to rely only on social IdP, as all users are comfortable to use their personal social IDs to access their work-related services. Furthermore, for some services the usage of social IDs is not possible, as they do not meet the service requirements.
- Use common, interoperable, (preferably open) standard protocols (and do test them beforehand).

The problem of determining the source of guest identities considered for a service is shown below in [Figure 4.1](#).

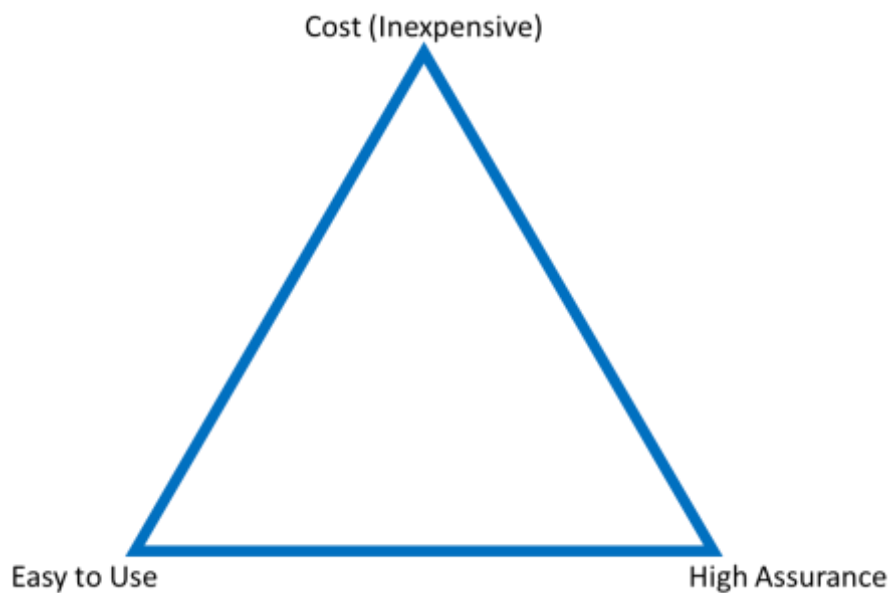


Figure 4.1: System trade-off triangle



This diagram shows a trade-off triangle where one can have, at most, two of the following features: an affordable system, high assurance, and user friendliness. Systems that provide high assurance, yet are reasonably user friendly are also expensive (e.g. OTP tokens); and cheaper solutions based on, for example, personal X.509 certificates, are considered less user friendly. Guest identities tend toward the “easy to use” corner; some, such as government IDs, are generally expensive, but provide high assurance, whereas community identities are typically cheaper to run (e.g. do not require expensive HSMS²), but have lower assurance.

In summary, there is no standard approach to the adoption of guest IdPs. Beginning with the target groups listed in the AARC proposal, “guest” users and institutions without an (inter-)federated IdP, various operational and cost models are applicable, depending on the individual conditions. In any case, well-established institutional partners, contractual frameworks and long-term funding are important factors for every sustainability model.

² While there may be communities that set up a “cheap and cheerful” IdP, particularly very small communities, most community *projects* will take the IdP seriously (cf. [IGTF-CSG]), and allocate effort and secure infrastructure to run a “proper” IdP.

5 Conclusions

The research and education sector has a long history in developing and deploying services. User requirements, increasing complexity and cost of infrastructures, demand for highly distributed infrastructures, availability of commercial services have changed the landscape. Embarking on the development of a new service requires resources and a plan, not only during the development phase, but also to ensure the service, if successful, can continue in the longer-term.

To that end the set of guidelines presented in this document for service providers in EIs and RIs highlights aspects that should be investigated by all relevant parties when developing and deploying new services. The guidelines should be considered as such, as it is not possible to provide a real template that can fit all cases.

The exercise made with the pilots/services above shows a very common phenomenon. In early project phases the focus is on the “User and User Base” and “Service Implementation”. Only at later project stages are the other service areas addressed. By addressing all areas early and developing the service operational model alongside the service, the risks of exceeding operational costs or even shutting down services can be contained.

The following policy recommendations were provided, taking existing frameworks and complexities into account.

- Users should be able to access services in EIs/RIs using the credentials they have in their home organisations.
- EIs/RIs should adopt the AARC Blueprint Architecture when implementing federated access across a number of internal services.
- Service providers that participate in eduGAIN should support the GÉANT CoCo, whenever possible.
- Service providers that participate in eduGAIN should comply with the Security Incident Response Trust Framework for Federated Identity (SIRTFI).
- Service providers that operate in the research and education sector that participate in eduGAIN should apply for R&S entity category.
- EIs/RIs that have adopted the AARC Blueprint Architecture should implement the SNCTFI policy framework.
- Service providers within RIs/EIs that require specific level of assurance, should monitor the development in the REFEDS Assurance WG, where such a framework is being discussed.
- EIs/RIs should be prepared to manage users relying on social IdPs.
- EIs/RIs should follow REFEDS discovery guidelines.
- Federation operators should agree to support edPpersonUniqueID.
- Federation operators should be caution in filtering eduGAIN metadata.
- Federation operators should promote adoption of R&S entity category.



- Ensure that home organisations do not reassign user identifiers.
- Promote and support participation to SIRTFI to handle incident response.
- Use eduGAIN to create a support help desk.

It is also understood that changing current procedures is a challenging task that requires time. These recommendations will be promoted as standalone documents among RIs, EIs and federation operators.



Appendix A First Responses on the Template from Selected AARC Pilots

The tables below show how the proposed guidelines to deploy sustainable services are applied. Not all questions are applicable to all pilots or services.

A.1 AARC RCauth.eu

Besides the information provided in the template below, a sustainability model study for RCauth is available in [\[SUSTAIN\]](#).

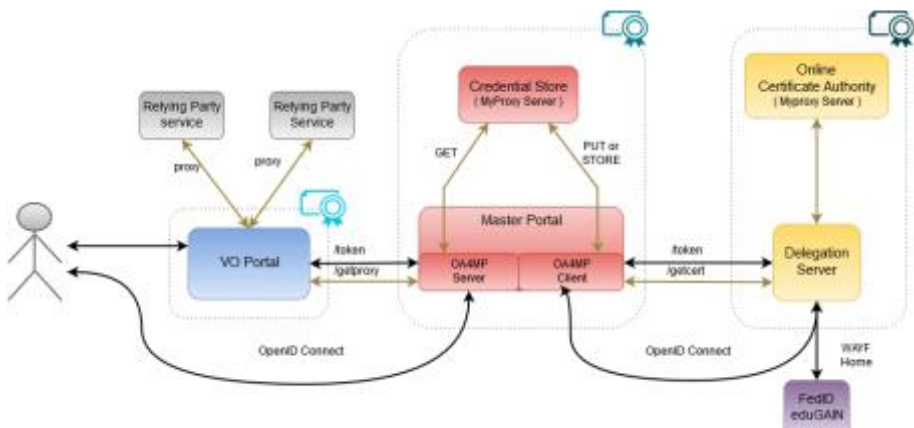
Aspects Considered	Reasons
Use Case and User Base	
<i>What is the pilot about?</i> <i>What problem does it solve?</i> <i>Which user group is it aimed at?</i> <i>What are typical use-cases?</i>	The “CILogon-like pilot for Europe” (of which RCauth.eu is the core translation component) is a token translation and credential management service that implements a bridge between the web-based R&E federation infrastructures (both based on SAML as well as OpenID Connect) and non-web scenarios.
<i>What is the estimated user base?</i>	RCauth.eu is positioned to serve research and generic e-infrastructures at a European level. The pilot, supporting initial operations for EGI and ELIXIR, is dimensioned to support up to ~ 4000 users (assuming once-weekly credential refreshment on Monday morning by all users), with a distributed scalability model that could support 100k+ users [SUSTAIN] .
<i>Are there already similar services?</i>	The CILogon-like pilot was modelled after (and produced in collaboration with) the CILogon service by the XSEDE and CTSC projects. A credential management service was added to support secure credential management by the European ESFRI cluster model (and reduce complexity for community science gateways and portals). Protocol support for OpenID Connect was provided to further ease deployment. These features are unique to the AARC Pilot.

Use Case and User Base	
<i>Is the service to be developed or procured?</i>	At the moment, all components are built and operated within the R&E community.
<i>Is federated access a requirement?</i>	Yes it is. In this specific case, the aim is to leverage federated access and generate an IGTF compliant certificates to access resource. The policy model supporting RCauth.eu also supports mechanisms to connect ‘legacy’ identity providers of last resort, based in the infrastructures.
<i>Is there sensitive data that the service needs to support?</i>	Yes, the data in the credential store is sensitive. In the recommended deployment model, this component is managed by the cluster infrastructures (so no central component will exist). The RCauth.eu translation service does not contain per-user sensitive data.
Operator	
<i>What are potential operators with a matching mission? Which operator is best suited for operating the service?</i>	In this specific case, the pilot has different components that can be operated in different ways, namely: <ul style="list-style-type: none"> • Each research infrastructure or an e-infrastructures to operate an instance of the service. • A research infrastructure to operate the whole service. • A consortium to operate the service. • Through a (commercial) third party.
<i>Who is going to support and train the users?</i>	The support for the RCauth.eu functions is provided indirectly through the connected Infrastructures. The responsibility is distributed so that end-user support will naturally fall to the research and e-infrastructures (potentially after triage by the community science gateways), and any central RCauth.eu service needs to only interact with verified (‘3 rd level’) issues mediated by the connected Infrastructures. User training is similarly devolved to infrastructures, and depends on the complexity of their own usage scenarios.
<i>Who should be responsible for the promotion of the service? (optional)</i>	The target audience for RCauth are the research and e-infrastructures. Through direct contacts, both EGI, EUDAT, and ELIXIR (European-wide), and national e-infrastructures (Dutch National e-infrastructure coordinated by SURF) have been engaged.
Costs estimation <i>What are the operator’s expected costs (in terms of cost of hardware/software and effort):</i> <ul style="list-style-type: none"> • For bootstrapping the 	Budgeting and resource requirements are highly dependent on the deployment model and desired service level. Potential deployment models are described in detail in [SUSTAIN] . Further details on costing and operational service requirements are considered private business information of the adopting infrastructures and are not discussed further in this document.

<p>service?</p> <ul style="list-style-type: none"> • For the annual operations? <p>When possible, find a key parameter that drives the costs (e.g. number of users, storage, hardware, procured service, software etc.) and state the constraints of your implementation</p>	
Sponsor	
<p>What are the plans for long-term cost recovery?</p>	<p>Cost recovery models depend on the deployment scenario(s) adopted. The sustainability model study identified three likely scenarios:</p> <ul style="list-style-type: none"> • One RCauth translation engine and credential store for each multidisciplinary e-Infrastructure – implying that only software maintenance needs to be sourced from an (external) technology provider. This model collectively saves on some resources, but bind research communities to a specific e-infrastructure. • One such engine per research infrastructure – the most costly option as it implies many (up to ~ 50) instances. • One single instance for Europe, with the credential management component distribute for availability. The last option is recommended, with cost-sharing based on an (in-kind or contract) recuperation model between the operating instance(s) and the ensemble of Infrastructures.
Governance, Policies and Processes	
<p>Are there specific policy aspects that should be taken into considerations?</p> <p>What are the specific security requirements?</p>	<p>In case there are two components that have security requirements:</p> <ul style="list-style-type: none"> • (the online CA): The CA should have very well-documented policies that are compliant with infrastructure requirements as specified through the IGTF, which are externally reviewed and reassessed periodically using, for example, a peer-review process and avails over specific hardware security modules (HSMs). • The credential repository that has to run in a secure environment, as it permits bulk access to a large number of user credentials.
<p>Is there sensitive data stored?</p>	<p>Yes the credential store contains the short-lived certificates and is, therefore, a critical component.</p>
<p>What are the availability</p>	<p>In specific scenarios the CILogon-like service may be a critical component for any work by the user. The service can be deployed in a high-availability mode (for</p>



<p><i>requirements?</i></p>	<p>which additional work beyond the pilot may be needed) or deployment can be scoped (one per infrastructure) so that service incidents have a more limited impact. Details are provided in the sustainability model study.</p>
<p><i>Are Service Level Agreements (SLAs) necessary or expected?</i></p>	<p>In the future SLAs will be required, as the RCauth CA is expected to be operational all the times.</p>
<p><i>What are the monitoring and accounting requirements?</i></p>	<p>At the moment, the general infrastructure needs monitoring, in particular, the CA.</p>
<p><i>What are the documentation requirements (user documentation, tutorials, administrators documentation, installation documentation)? (optional)</i></p>	<p>Documentation on how to install an instance of this pilot has been produced within the AARC project.</p>
<p>Service Implementation</p>	
<p><i>What is the current architecture? Are there dependencies with external tools/software/licences? What are conceivable deployment and operational scenarios?</i></p>	<p>The CILogon-like pilot can be decomposed into a few distinct service components, and the model study contains an inventory of possible deployment models for each of these service elements. The pilot comprises two elements: the Master Portal and the Delegation Server. The Master Portal is a bridging component between the identity service, any community assertion services (VOMS, not shown in figure), and the VO portals. It uses secure bilateral protocols to exchange information with both the VO portals and with the Delegation Server. Yet, in order to fulfil its role effectively, it also acts as a credential repository, and will hold long-term credentials for a (potentially large) set of users.</p> <p>The Delegation Server provides the actual token translation between the federated (SAML) user ID, the PKI certificate, and the OpenID Connect authentication by the user (via the master portal). In the PKI domain, it acts as a certification authority (CA) trusted third party and its credentials are to be accepted as authoritative by all resources and service providers in the infrastructure. The online CA is annexed to the Delegation Server but kept separate in order to ensure compliance with the minimum security requirements and IGTF guidelines.</p>

<p>How many elements compose the service? (provide schematic if possible, see example)</p>	 <p>Colours in the graphic above represent distinguishable service elements. For details we refer to the dedicated sustainability model study. The Master Portal, Credential Store, Online Certificate Authority and Delegation Server are components of the RCauth pilot. The blue “VO Portal” is outside the pilot and it is responsibility of the user community that implements it.</p>
<p>Technical requirements (VMs, storage, network...)</p>	<p>For security and IGTF accreditation reasons, specific hardware security modules are required for operation, and the credential store must be a specifically secured system. For requirements, refer to the RCauth.eu Certificate Policy and Practice Statement (CP/CPS) [RCauth].</p>
<p>What is the estimated sustainability of the software being used?</p>	<p>Software from the CTSC XSEDE CILogon and NCSA MyProxy project was extensively re-used in this service, and augmented by automatically-deployable models specific to the Master Portal. Although modifications to the software have been submitted to our upstream providers, selected component will need additional maintenance. It is necessary and expected that Infrastructures deploying the service will contribute (monetary or in-kind) to its maintenance and any necessary evolution.</p>

A.2 DARIAH Guest IdP

Aspect Considered	Reasons
Use Case and User Base	
<p><i>What is the pilot about?</i></p> <p><i>What problem does it solve?</i></p> <p><i>Which user group is it aimed at?</i></p> <p><i>What are typical use cases?</i></p>	<p>The DARIAH-DE Guest IdP is a so-called IdP of last resort for users that want to use DARIAH-Services but do not have access to a federated / institutional account. If these users are able to proof their affiliation to the target user group (of researchers / scholars in the field of digital humanities), they can get a dedicated DARIAH-DE account to access the DARIAH-DE services.</p>
<p><i>What is the estimated user base?</i></p>	<p>Current users: More than 3800 (as of March 2017).</p>
<p><i>Are similar services already available?</i></p>	<p>There are similar services but none that can fulfil the policy- and governance-needs of DARIAH, as</p> <ul style="list-style-type: none"> • DARIAH has to make sure that every user is part of the research community in the field of Digital Humanities. • Uses the available services only for the intended purpose – his research • Has to manage additional organisational processes and permission settings.
<p><i>Is the service to be developed or procured?</i></p>	<p>Built and managed by the community.</p>
<p><i>Is federated access a requirement?</i></p>	<p>Not applicable as users that arrive the guest IdP do not have any federated credentials. Users are provided with federated credentials through this service.</p>
<p><i>Are there sensitive data that the service needs to support?</i></p>	<p>The service handles personal data, but only ones with low protection requirements (name, email-address, organisation).</p>
Operator	
<p><i>What are potential operators?</i></p> <p><i>Which operator is best suited for the service?</i></p>	<p>DARIAH itself is a potential operator and the one with the best matching mission. Within DARIAH, as a consortium, DAASI International seems the most appropriate operator for the service and is willing to support it. Funding is dependent of the long-term funding of DARIAH as a whole.</p> <p>Other potential operators are other national members of the DARIAH-EU ERIC; in Germany, maybe the DHD association (Digital Humanities im deutschsprachigen Raum) or CLARIN.</p>

<p><i>Who is going to support and train the users? (optional)</i></p>	<p>The operator supports and trains users.</p>
<p><i>Who should be responsible for the promotion of the service? (optional)</i></p>	<p>DARIAH-DE promotes the service.</p>
<p>Costs estimation:</p> <p><i>What are the operator's expected costs (in terms of cost of hardware/software and effort):</i></p> <ul style="list-style-type: none"> • <i>For bootstrapping the service</i> • <i>For the annual operations?</i> <p><i>When possible, find a key parameter that drives the costs (e.g. number of users, storage, hardware, procured service, software etc.) and state the constraints of your implementation.</i></p>	<p>The Guest IdP operation currently has a financial cost of EUR42 000 per year, which is only possible because of funding coming from different organisations. The costs would be higher in another situation and are estimated at EUR85 000 per year.</p> <p>Key cost drivers are:</p> <ul style="list-style-type: none"> • Users. • Connected services. • Number of workflows (user registration, change of organisation, individual service request, ...).
<p>Sponsors</p>	
<p><i>What are the plans for long-term cost recovery?</i></p>	<p>The operator DARIAH-DE runs the service and bears the costs.</p>
<p><i>What are potential risks in service operation? Who bears these risks - operator or sponsor?</i></p>	<p>Risks are:</p> <ul style="list-style-type: none"> • Unexpected high user or service registrations / usages • Cyberattacks like DDOS etc. <p>The operator bears these risks in terms of quality. As no profits apply, no financial risks through contract liabilities or other liabilities are feasible.</p>
<p>Governance, Policies and Processes</p>	
<p><i>Are there specific policy aspects that should be taken into considerations?</i></p> <p><i>What are the security requirements?</i></p>	<p>There are many policy and security requirements that are discussed and determined in specific reports.</p>
<p><i>What are the availability</i></p>	<p>As the availability of the Guest-IdP is, at the same time, the threshold for any possible service SLA within DARIAH-DE, 24/7 availability is absolutely required and aimed for by</p>

<p><i>requirements?</i></p>	<p>the operator.</p> <p>Because of the current funding mechanism, compensation in case of any outages is not possible.</p>
<p><i>Are Service Level Agreements (SLA) necessary or expected?</i></p>	<p>Users expect at least 24/7 availability, but until now, there was no demand by the research community to enter into a SLA contract. Nevertheless, these SLAs are already written and ready to be signed, as soon as the demand is there and the organisational construct of DARIAH-DE allows for it.</p>
<p><i>What are the monitoring and accounting requirements?</i></p>	<p>Monitoring of the instance (hardware, operating system, etc.) as well as the service (regular functional probes) are in place.</p> <p>The main cost drivers (see above) are recorded and can be used for cost estimates and accounting purposes. Additionally, there are current discussions to use the Guest-IdP infrastructure as hub for accounting other services within DARIAH-DE.</p>
<p><i>What are the documentation requirements (user documentation, tutorials, administrator documentation, installation documentation, ...)? (optional)</i></p>	<p>There is documentation both for operating the service itself (from administrator documentation to end-user documentation) and for workflows and processes as they are implemented within DARIAH-DE.</p>
<p>Service implementation</p>	
<p><i>What is the current architecture?</i></p> <p><i>Are there dependencies with external tools/software/licences?</i></p> <p><i>What are conceivable deployment and operational scenarios?</i></p>	<p>see https://wiki.de.dariah.eu/display/publicde/DARIAH+AAI+Documentation</p>
<p><i>How many elements compose the service?</i></p> <p>(provide schematic if possible, see example)</p>	<p>The service is composed by:</p> <ul style="list-style-type: none"> • A self-service interface based on a DARIAH software component (didmos LUI) • An administration interface based on didmos LUI • A helpdesk infrastructure with workflow-capabilities based on OTRS • An Identity Provider based on Shibboleth • An RBAC system based on didmos Decision Point
<p><i>Technical requirements (VMs, storage, network...)</i></p>	<p>The infrastructure runs on 7 VMs (including staging systems for testing updates or new configurations). These currently all have a mid-ranged layout (2 CPUs, 50 GB HDD, 4GB RAM). The elements stated above mostly even independent of a special distribution</p>

	and can be operated on all common Linux environments today.
<i>What is the estimated sustainability of software being used?</i>	<p>All software being used is open source, which can be used and further developed by other operators as well. The main contributors of these software components are currently successful commercial companies with an interest to further develop and maintain them.</p> <p>The architecture and implementations are accessible and published.</p>

A.3 Social IDs (to SAML) pilot

Aspect Considered	Reasons
Use Case and User base	
<p><i>What is the service about?</i></p> <p><i>What problem does it solve?</i></p> <p><i>Which user group is it aimed at?</i></p> <p><i>What are the typical use-cases?</i></p>	<p>Including Guest Identities in the consuming of federated services through a Social to SAML proxy.</p> <p>It solves the problem of users in need of an identity enabling them to collaborate with other researchers already owning Federated credentials, for which a priori there is no possibility to make use of services offered; the pilot goes beyond pure inclusion of Social Identities, and allows managers of scientific collaborations to attribute an higher LoA through identity vetting, sponsorship and account linking to the ORCID registry.</p> <p>It is aimed at researchers who do not own federated credentials, but only social ones, in need of inclusion in scientific collaboration and use of eduGAIN-based Service Providers.</p> <p>A typical use case is a non-EU researcher working, for example, for ELIXIR or an LHC experiment at CERN and in need of access eduGAIN-based SPs.</p>
<i>What is the estimated user base?</i>	Non eduGAIN-IDs-owning researchers.
<i>Are there already similar services?</i>	There are similar pilots in the goals, although addressing different requirements (e.g. X.509 to SAML). This implementation is integrated through the IDP/SP proxy, and the COMANAGE AA allows for setting the required SAML attributes. Account linking to ORCID also allows, in some cases, to enhance the LoA of the managed social identities.
<i>Is federated access a requirement?</i>	Federated access is not required.
<i>Is the service to be developed or procured?</i>	The service is built by gluing together existing software components.
<i>Is there sensitive data that the</i>	No.



<i>service needs to support?</i>	
Operator	
<p><i>What are potential operators?</i></p> <p><i>Which operator is best suited for the service?</i></p>	<p>The potential operators are scientific collaboration managers in charge of including researchers in managed collaborations via COMANAGE.</p> <p>It is suitable for collaboration managers</p>
<i>Who is going to support and train the users? (optional)</i>	
<i>Who should be responsible for the promotion of the service? (optional)</i>	Research infrastructure managers and collaboration managers promote the service.
<p>Costs estimation</p> <p><i>What are the operator's expected costs (in terms of cost of hardware/software and effort):</i></p> <ul style="list-style-type: none"> • For bootstrapping the service • For the annual operations <p><i>When possible, find a key parameter that drives the costs (e.g. number of users, storage, hardware, procured service, software etc.) and state the constraints of your implementation</i></p>	<p>Very limited costs in terms of basic setup (Mid-size server: 4GB RAM, 4 vCPUs) . Proxy might require High Availability solutions</p>
Sponsor	
<i>What are the plans for long-term cost recovery?</i>	Costs should be covered by beneficiaries; research collaborations in the first instance.
<i>What are potential risks in service operation? Who bears these risks - operator or sponsor? (optional)</i>	Risks are associated to the proxying functionality. It is fundamental to have HA and DR available to avoid SPoF.
Governance, Policies and Processes	
<p><i>Are there specific policy aspects that should be taken into considerations?</i></p> <p><i>What are the security requirements?</i></p>	Yes. AARC is recommending settings for Guest IdPs.
<i>What are the availability requirements?</i>	Security management in a proactive fashion (patches, updates..)



Are Service Level Agreements (SLA) necessary or expected?	The service has to be up with 99.99 % uptime.
What are the monitoring and accounting requirements?	Network and main machine parameter monitoring are fundamental.
What are the documentation requirements (user documentation, tutorials, administrator documentation, installation documentation)? (optional)	
Service Implementation	
What is the current architecture? Are there dependencies with external tools/software/licences? What are conceivable deployment and operational scenarios?	Basic set-up using IDP/SP proxy and COMANAGE. Reported on https://wiki.geant.org/display/AARC/SocialIDs
How many elements compose the service?	https://wiki.geant.org/display/AARC/SocialIDs eduGAIN IdP; Social (Google) login; IDP/SP proxy, COMANAGE registry
Technical equipment (VMs, storage, network...)	1 VM - Reliable network connection
What is the estimated sustainability of software being used?	Will follow common understanding on AARC pilots involving s/w

A.4 WATTS pilot

Aspect Considered	Reasons
Use Case and User Base	
<p><i>What is the pilot about?</i></p> <p><i>What problem does it solve?</i></p> <p><i>Which user group is it aimed at?</i></p> <p><i>What are typical use-cases?</i></p>	<p>In this pilot, WaTTS (i.e. Token Translation Service developed by KIT) is used to enable the users to manage the SSH access to a number of trusted VMs from a single point in a secure and user-friendly manner.</p> <p>The problem solved is the users' need to access services that cannot directly utilise federated access and require that the users use security tokens, in this case SSH keys.</p> <p>The pilot is aimed at users and RI managers who wish to provide CLI access to users using federated identity.</p> <p>Typical use case is SSH access to VMs, where keys are managed through WaTTS.</p> <p>Full pilot description: https://wiki.geant.org/pages/viewpage.action?pageId=65733556</p>
<i>What is the estimated user base?</i>	Anybody who has a need for an SSH access. This is potentially EGI, EUDAT, cloud providers.
<i>Are there already similar services? If so, what does this new service add?</i>	Similar services include Moonshot or other SSH provisioning tools. This service is different from other services, as it allows users to login with different type of credentials.
<i>Is federated access a requirement?</i>	It is not, users are able to use their social identities. However, granting access to services is discriminated against LoAs, therefore, certain services might be inaccessible if users do not use accounts with sufficiently high LoA, where federated identity is potentially a requirement.
<i>What is the service to be developed or procured?</i>	The service is to be developed.
<i>Is there sensitive data that the service needs to support?</i>	The service needs access to users information, which in this case includes name, email and similar data. The service also receives an OIDC access token.
Operator	
<p><i>What are the potential operators?</i></p> <p><i>Which operator is best suited for operating the service?</i></p>	A WaTTS instance with the SSH plugin would be operated centrally. Since VMs might be managed at the level of the infrastructure, one instance per infrastructure is a potential use case. However, it can be deployed at a more local level, if there is such need to provision VMs.
<i>Who is going to support and</i>	KIT / Uros Stevanovic, Bas Wegh, Marcus Hardt

<i>train the users?</i>	
<i>Who should be responsible for the promotion of the service? (optional)</i>	Uros Stevanovic and Marcus Hardt
<p>Costs estimation <i>What are the operator's expected costs (in terms of cost of hardware/software and effort):</i></p> <ul style="list-style-type: none"> • <i>For bootstrapping the service?</i> • <i>For the annual operations?</i> <p><i>When possible find a key parameter that drives the costs (e.g. number of users, storage, hardware, procured service, software etc.) and state the constraints of your implementation</i></p>	<ul style="list-style-type: none"> • Costs / Initiating: Personnel ~6PM (well trained, specialised personnel) • Costs / Maintaining: 0.25FTE (trained admin) + 1PM/a security audits (highly specialised personnel) • Hardware: 1 VM for WaTTS, and provisional number of end VMs • Costs are low
Sponsor	
<i>What are the plans for long-term cost recovery?</i>	KIT can run the service in the longer-term; costs can be supported by an e-infrastructure such as EGI, for instance.
<i>What are potential risks in service operation? Who bears these risks - operator or sponsor? (optional)</i>	
Governance, Policies and processes	
<p><i>Are there specific policy aspects that should be taken into considerations?</i></p> <p><i>Are there specific security requirements?</i></p>	WaTTS does not keep users keys, and users could only upload their public SSH key. Security considerations include securing WaTTS instance.
<i>What are the availability requirements?</i>	If service is in production, then it should be available at all the times.
<i>Is there sensitive data stored by the service?</i>	



Are Service Level Agreements (SLA) necessary or expected? (optional)	
What are the monitoring and accounting requirements?	
What are the documentation requirements (user documentation, tutorials, administrator documentation, installation documentation)?	Admin docs => available Install docs => available, but incomplete User docs => should be self-explanatory
Service Implementation	
What is the current architecture? Are there dependencies with external tools/software/licences? What are conceivable deployment and operational scenarios?	Web service with web and REST interface All is based on standard Debian packages Deployment: Basic WattS + RCAuth plugin + MYProxy server + access to one or more third party VOMS servers
How many elements compose the service?	<ul style="list-style-type: none"> • WaTTS • SSH Plugin • VMs
Technical equipment (VMs, storage, network...)	1 VMs ~ 20GB disk (altogether). The web frontend benefits from good network connectivity, but there's less than 1GB traffic per month.
What is the estimated sustainability of software being used?	Developed at KIT, it will maintained if there is usage



Appendix B Managing Risks of Using Guest Identities

While the advantages of enabling guest identities are quite clear, it is also worth pointing out other aspects related to their integration. There are different requirements in terms of level of assurance that were gathered in other AARC document [MNA3.1], whilst [MJRA1.2] identified a set of risks associated to these assurance requirements.

The AARC team tried to explore an approach to assess how each possible guest IdPs support the identified requirements and to assign a score (1-5), as indicated in the table below. In the following table, the likelihood³ of the risk has been assessed as a number 1-5, using the following criteria:

1. Requirement is certain to be met correctly; there is supporting information.
2. Likely to be correct, or expected to be correct, but not documented or audited, or documentation is not available.
3. Possibly not done correctly, party may lack the skill, motivation, or may simply not have thought about implementing the requirement.
4. Medium-to-high expectation that a violation will happen, or requirement is not explicitly implemented but is fairly unlikely to be violated directly.
5. A violation is very likely to happen, e.g. where the requirement is not implemented and is as likely to be violated as not.

Document	Summary	Gov't	Community	Social	Commercial
MNA3.1-1	Unique account	1	2	1	1
MNA3.1-2	Persistent identifier	1	3	2	2
MNA3.1-3	ID vetting risks	1	5	5 ⁴	2
MNA3.1-4	Password practices	1	4	1	1

⁴ There are a few exceptions like Twitter's verified account (typically for celebrities that may be impersonated) - <https://support.twitter.com/articles/119135#>



Document	Summary	Gov't	Community	Social	Commercial
MNA3.1-5	Prompt closure	1	5	5	2
MNA3.1-6	Self-assessment	1	5	1	2
MJRA1.2-1	Clarity on use of id	1	3	2	1
MJRA1.2-3	Sustainable IdP	1	4	1	2
MJRA1.2-6	Incident handling ⁵	4	4	4	4

As this was mostly an exercise to provide some guidance for service providers, the assessment does not include the penetration level of some of the solutions (for example, government identities can generally be considered to be the most secure source for guest identities, but it is yet not a feasible approach to only rely on them).

⁵ The push towards SIRTfI in academic environments (federations) may spill over into community IdPs



References

- [AARC] <https://aarc-project.eu/>
- [BLUEPRINT] <https://aarc-project.eu/blueprint-architecture/>
- [CoCo] <http://www.geant.net/uri/dataprotection-code-of-conduct/v1/Pages/default.aspx>
- [D95/46/EC] <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML> [DARIAH]
<http://www.dariah.eu/>
- [DHd] <https://dig-hum.de/>
- [DNA3.5] https://aarc-project.eu/wp-content/uploads/2016/12/AARC-DNA3.5_Recommendations-for-Processing-Personal-Data_2016_11_07_v4_DG.pdf
- [eduGAIN] https://www.geant.org/Services/Trust_identity_and_security/eduGAIN
- [EGI] <https://www.egi.eu/>
- [IGTF] <https://www.igtf.net/>
- [IGTF-CSG] <http://wiki.eugridpma.org/Main/CredStoreOperationsGuideline>
- [INDIGO] <https://www.indigo-datacloud.eu/node/72>
- [FED-BP] <https://wiki.refeds.org/display/FBP/Federation-Best-Practice+Home>
- [FIM4R] <https://indico.cern.ch/event/301888/>
- [MJRA1.2] <https://aarc-project.eu/wp-content/uploads/2016/06/MJRA1.2-Design-for-Deploying-Solutions-for-Guest-Identities.pdf>
- [MNA3.1] <https://aarc-project.eu/wp-content/uploads/2015/11/MNA31-Minimum-LoA-level.pdf>
- [RCauth] <https://rcauth.eu/policy/>
- [RECOMMEND] MJRA1.2 <https://aarc-project.eu/wp-content/uploads/2016/06/MJRA1.2-Design-for-Deploying-Solutions-for-Guest-Identities.pdf>
<https://wiki.geant.org/download/attachments/57639704/AARC-sustainability-models-for-guest-idps.pdf?version=1&modificationDate=1461656536226&api=v2>
- [REFEDS] <https://indico.cern.ch/event/301888/>
- [RS] <https://refeds.org/category/research-and-scholarship>
- [SIRTFI] <https://refeds.org/wp-content/uploads/2016/11/Sirtfi-certification-v1.0.pdf>
- [SNCTFI] <https://wiki.geant.org/display/AARC/Snctfi>
- [SOCIALID] <https://wiki.geant.org/display/AARC/SocialIDCockpitPanel>



[SUSTAIN] <https://wiki.geant.org/download/attachments/56918657/AARC-sustainability-models-for-RCauth-20160506.pdf?version=1&modificationDate=1462630136894&api=v2>

[SURVEY] <https://geant.app.box.com/s/8f30ptw5houmauurfqfupw3ruz3x9enu>

Glossary

CA	Certificate Authority
CoCo	Code of Conduct
DeISU	DARIAH e-Infrastructure Service Unit
DR	Disaster Recovery
EI	e-infrastructure
ePPN	eduPersonsPrincipalName
IdP	Identity Providers
IGTF	Interoperable Global Trust Federation
GDPR	General Data Protection Regulation
HA	High Availability
HDD	Hard Disk Drive
HSM	Hardware Security Modules
LoA	Level of Assurance
REFEDS	Research and Education FEDerations
R&S	Research and Scholarship
RI	Research Infrastructures
SAML	Security Assertion Mark-up Language
SIRTFI	Security Incident Response Trust Framework for Federated Identity
SLA	Service Level Agreements
Snctfi	Scalable Negotiator for a Community Trust Framework in Federated Infrastructures
SP	Service Provider
SPoF	Single Point of Failure
SSH	Secure Shell
VM	Virtual Machines