~~Guidelines for evaluating the combined assurance of linked identities~~

DO NOT WORK ON THIS DOCUMENT
CURRENT WORKING VERSION:

AARC-G031 Guidelines for the evaluation and combination of the assurance of external identities

Introduction

The evolution of the R&E AAIs has to take into account the new environment for eGOV IDs that is being created by the eIDAS Regulation, and the AAIs that are being used in the private sector (both the enterprise and the so-called “social identities”).
In practical terms that means envisioning integration models and investigating interoperability issues.
In this context, Account Linking between R&E existing identities and cross-sector identities is a means to achieve integration, and it is a fundamental component to make different AAIs interoperable. On the other hand, identities coming from cross-sector AAIs may have a very low Assurance Level (AL), or on the contrary a higher AL compared to the one commonly used in the R&E space. In the first case, in order to use low-AL identities in R&E we need techniques and policies to elevate the AL, while in the second one we can combine the higher-AL identities with the R&E ones to achieve a higher AL to be used in a sensitive context like life science.
In this document we will investigate the use of Account Linking to evaluate the combined Assurance of identities, and we will take into account use cases related to different sectors.
To investigate LoA elevation itself, first of all we will define an LoA evaluation model based on the REFEDS Assurance Framework. In order to understand the possible outcomes of different LoAs combination, we will also define a combined LoA evaluation model.

1.1. Conventions

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in RFC 2119[1].

1.2. Definitions

1.2.1. Infrastructure identity

In the context of research collaborations, the user is typically assigned an identity by the infrastructure. This “infrastructure identity” consists of a personal, unique, non-reassignable, non-targeted identifier, and additional attributes containing profile information about the user, as well as group membership and role information. The infrastructure identity can be associated with a set of credentials issued by the Infrastructure itself.

[1] https://www.ietf.org/rfc/rfc2119.txt

2. Identity Linking

2.1. Definition

In the context of research collaborations, identity linking (commonly refer to as account linking) refers to the process of connecting the user's infrastructure identity with their external identities, i.e. identities created and assigned by Identity Providers that reside outside of the administrative boundaries of the infrastructure, such as institutional IdPs or social media IdPs. The identity linking process allows the user to access infrastructure resources as their infrastructure identity regardless of the identity used for authentication. It should be noted that the infrastructure identity can be used to obtain different types of credentials for accessing resources, for example, X.509 certificates, SSH keys or other access tokens. In fact, the user may not be aware of the credentials being used to access a specific resource, since in some cases the credentials are translated behind the scenes by the infrastructure.

2.2. Technical process

Account linking typically takes place as part of the user enrolment process, either explicitly or automatically, as described in the subsections that follow.

2.2.1. Explicit linking

In the explicit linking flow, the user requests that an additional identity be linked to their existing infrastructure identity. This flow requires the user to authenticate first with any of the identities already linked to their infrastructure identity (or with the infrastructure identity itself), and then to re-authenticate using the login credentials of the additional identity they want to connect. It should be noted that the administrators of the infrastructure identity management system can also manage identity links, usually to resolve enrolment issues, e.g. duplicate user registrations.

2.2.2. Automatic linking

The automatic linking process is triggered when one attribute, or a combination of attributes, of one identity correlate to one or more attributes of another identity that is already associated with a registered user. The correlation process may require exact matching of attribute values or tolerate some differences. In the latter case, this could allow for inconsistently capitalised or similar identity values. Automatic linking can prevent an individual from registering distinct infrastructure identities, either accidentally or on purpose. It can therefore be useful in an infrastructure with a strict policy against maintaining multiple user accounts. However, the risk here is that identities which should not be linked may accidentally be matched by this process. Therefore, automatic linking should not be considered unless either the correlation process requires an exact matching on attribute values expressing user identifiers that are personal, globally unique and non-reassignable, while also considering the level of assurance (LoA) associated with the matching attribute(s),

or the resulting account is directly derived from the user identifiers that are personal[2], globally unique and non-reassignable. Examples of attributes that may be considered for automatic linking include subject distinguished names of personal X.509 certificates and ORCID identifiers [ORCID]. In other cases, such as when detecting the same email address, the account linking process may be automatically triggered, yet it would require explicit user intervention before being applied due to the undefined reassignment practise for such attributes.

2.3 Reliability requirements

Requirements:

- An IdP able to do both strong identity vetting and MFA.
- An IdP provides strong identity vetting and second factor is delivered by the RI/EI.

- An IdP provides strong identity vetting (medium to high IAP value), the second factor is provided by an IdP with a low IAP value (for example social media platforms cannot signal if they employ MFA or not).

3. Assurance Evaluation

In the context of this document we will use the definition of Assurance as expressed in the REFEDS Assurance Framework [RAF].

Along the lines of other recent assurance guidelines [NIST.SP.800.63.3] and proposed standards [VOT], the RAF does not use the concept of level(s) of assurance, rather it splits assurance into separate components. RAF considers the following 4 components:

- Identity uniqueness
- Identity proofing and credential issuance, renewal and replacement
- Authentication
- Attribute quality and freshness

The components can eventually be collapsed to compose assurance profiles, each consisting of a set of values for one or more of these components. Currently two assurance profiles are defined in the RAF standard: Cappuccino and Espresso.

3.1. Combined Assurance Evaluation

The ground for combined assurance evaluation is that assurance components related to the same individual, but coming from different Credential Service Provider (CSP), are defined along the lines of the RAF, or are translatable to those definitions.

[2] A personal identifier is intended for use by a single person, as opposed to shared (or guest) user accounts such as "libraryuser1@university.org".

Care must be taken not to assume that a CSP as a whole complies with the defined assurance profiles, but rather that the identity (or class of identity) used by the individual does. The individual's organisation may not necessarily define the same levels of assurance requirements for all identities in their CSP, e.g. administrative staff in a medical institute may not necessarily need to have as high a level of identity assurance as the medical researchers who work with sensitive data. Home organisations must then clearly define which classes of identities will comply with each/any of the defined assurance profiles.

3.1.1. Identifier uniqueness

The RAF Identifier uniqueness component describes "how a CSP expresses that an identifier represents a single natural person and if that person remains the same over time" [RAF].

A value of `$PREFIX$/ID/unique` is assumed for the identity uniqueness component of all linked identities. Identities that don't qualify for `unique` MUST NOT be considered for account linking, nor for combined assurance evaluation.

3.1.2. Identity proofing and credential issuance, renewal and replacement

The value of the identity assurance component of the infrastructure identity is determined from the highest/strongest identity assurance value of the linked identities.

When combining IAP values, the Infrastructure MUST also evaluate the authentication assurance of the identity with the highest IAP value. If the authentication of that identity is based on a weak password, then there is a risk of impersonation. As a rule of thumb, a value of IAP higher than `$PREFIX$/IAP/low` SHOULD always be paired with an authentication assurance value equal to or greater than `https://refeds.org/profile/sfa`.

The value of the identity assurance of a linked identity must only be considered if that linked identity has been used within the last 12 months.

3.1.3. Authentication Assurance

The authentication assurance value is not affected by account linking, and cannot be the result of a combination of values. It's value is based solely on the authentication assurance of the identity used to authenticate.

3.1.4. Attribute quality and freshness

Use cases:

1. Home Organization IdP able and willing to transfer ePA attributes.
2. HO IdP not able to assert ePA - Affiliation later asserted by an infrastructure admin.
3. Self-asserted affiliation by the user.

3.2 Combined Assurance components evaluation matrix

	Linked Identity B	*Identity proofing and credential issuance, renewal and replacement*			*Attribute Assurance*	
Linked Identity A	**Infrastructure Identity**	$PREFIX$/IAP/low	$PREFIX$/IAP/medium	$PREFIX$/IAP/high	$PREFIX$/ATP/ePA-1m	N/A
Identity proofing and credential issuance, renewal and replacement	$PREFIX$/IAP/low	$PREFIX$/IAP/low	$PREFIX$/IAP/medium	$PREFIX$/IAP/high		
	$PREFIX$/IAP/medium	$PREFIX$/IAP/medium	$PREFIX$/IAP/medium	$PREFIX$/IAP/high		
	$PREFIX$/IAP/hihg	$PREFIX$/IAP/high	$PREFIX$/IAP/high	$PREFIX$/IAP/high		
Attribute Assurance	$PREFIX$/ATP/ePA-1m				$PREFIX$/ATP/ePA-1m	$PREFIX$/ATP/ePA-1m
	N/A				$PREFIX$/ATP/ePA-1m	N/A

3.3. Linked accounts and assurance components freshness

The infrastructure MUST define a policy for the updating of the values of the assurance components for linked identities, and it MUST define a maximum validity time.
The specific maximum depends on the requirements of the infrastructure. For example, relying on $PREFIX$/ATP/ePA-1m will probably require a maximum of around 1 month, while in other cases it could be a year.

4. Use cases

4.1. Social + eduGAIN

A user registers with an R/E-infrastructure using a social media identity, which typically is self asserted and lack any process of identity vetting. Subsequently, the user links her organisational identity (e.g. from eduGAIN) to her infrastructure identity. By linking the two identities, the user has proved that they both refer to the same person, but the infrastructure

needs to assign a value to each assurance component in order to make authorization decisions.

Assuming both providers meet the same requirements with respect to the uniqueness of the identifiers and the authentication strength, the infrastructure may assign a high LoA when the user logs in using the social media identity, since it has been linked to a high-LoA organisational identity that makes up for the lack of identity vetting.

Considerations:
Can we assume identifier uniqueness (including non-reassignability) and authentication assurance for well-known social/non-R&E user identifiers? See table below

	Google	Facebook	LinkedIn	ORCID	GitHub
Identity Uniqueness	X[3]				-[4]
Identity Proofing	-	-	-	?	-
Authentication Assurance	-[5]	X[6]			-[7]
Attribute Assurance	-	-	-		-

Assuming that the R/E-infrastructure effectively consider the two linked accounts for the evaluation of the LoA, let's see the outcome of the combined LoA:

eduGAIN Assurance Components	**Google Assurance Components**	**Combined Assurance Components if Google is used for authentication**
$PREFIX$/ID/unique	$PREFIX$/ID/unique	$PREFIX$/ID/unique
$PREFIX$/IAP/high	-	$PREFIX$/IAP/high
`https://refeds.org/profile/sfa`	-	-

[3] Google returns an OIDC public `sub` claim which is unique and non-reassignable by definition: https://developers.google.com/identity/protocols/OpenIDConnect#obtainuserinfo
[4] While account uniqueness in a precise moment in time is granted, GitHub accounts are reassignable after deletion, see https://help.github.com/articles/deleting-your-user-account/
[5] https://support.google.com/a/answer/33386
[6] https://developers.facebook.com/docs/workplace/authentication/password#passwordpolicies
[7] GitHub password policy requires a (minimum) 7 character long password containing at least one number. GitHub also supports two-factor authentication, see https://help.github.com/articles/about-two-factor-authentication/

$PREFIX$/ATP/ePA-1m	-	$PREFIX$/ATP/ePA-1m

4.2. eduGAIN + eduGAIN

Two ePSA attributes asserted (identity A and identity B), but just one `$PREFIX$/ATP/` value.

ADDITIONAL CONTENT

Some nasty details to consider in the design:

- Do the ID proofing levels expire or rotten over time? (you have created your infrastructure ID 15 years ago…)
- What is good enough for the actual account linking? Is the event of account linking so sensitive that it's security requires extra attention? Do you need to be able to log in to do the account linking or can it be done by demonstrating control of an e-mail address?
- Is it in the scope of this document to design a plan B for managing Home Organisation eP(S)A attribute for scenarios where the HO IdP is not able/willing to deliver that attribute.
    - Infrastructures need to serve also those users who don't have federation capability (i.e. SAML IdP) and who therefore need to use social Id to log in
    - For instance, in the ELIXIR AAI we have a plan that eP(S)A can be assigned to an infrastructure identity manually by a trusted person

4.3. eduGAIN+eIDAS

References

RAF	(Draft) https://docs.google.com/document/d/15v65wJvRwTSQKViep_gGuEvxLI3UJbaOX5o9eLtsyBI/edit
NIST.SP.800.63.3	NIST Special Publication 800-63-3, *Digital Identity Guidelines*, June 2017, https://doi.org/10.6028/NIST.SP.800-63-3
VOT	https://www.ietf.org/id/draft-richer-vectors-of-trust-07.txt
ORCID	https://orcid.org/