Account linking and LoA elevation use cases and common practices for international research collaboration

Published Date:	13-06-2017
Revision:	1.0
Work Package:	JRA1
Document Code:	AARC-JRA1.4H
Document URL:	https://aarc-project.eu/wp-content/uploads/2017/03/AARC-JRA1.4H.pdf

Table of Contents

1 Introduction

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. In this context, identity linking (also known 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 their external 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 Account linking use cases

This section presents the main use cases for account linking:

- Consistent user identification/representation.
- Accounting of resource usage.
- Traceability and security incident response.

2.1 Consistent user identification/representation

An end user typically maintains accounts on different external authentication providers, including:

- Identity providers managed by the user's home organisation.
- Guest identity providers, including social media (such as Google, Facebook and LinkedIn), research community/collaboration-specific identity providers, and national e-government identity providers.

Regardless of the authentication provider being used, the end user wishes to be identified consistently when accessing infrastructure services and resources. This is the main use case for account linking as it supports the reconciliation of identities from various authentication providers, allowing a user to authenticate using any of their identities and still be recognised by the infrastructure with the same user profile. For example, the user

can register to the infrastructure with their organisational ID, thus providing a strongly verified credential, but linking their social media ID – which can remain activated in the browser – they can easily "log in" to an infrastructure portal without having to authenticate.

2.2 Accounting of resource usage

Infrastructure providers may need to track the resources consumed by individual users. Accounting usage data usually associates a usage record to a user ID. Linking the user IDs from different IdPs and the certificate DN of the X.509 credentials allows the collection of all the data that is associated to the user and the different IDs that have been used to access the services. If this account linking is done at infrastructure level, the services do not need to change the way accounting data is produced (e.g. attach several user IDs to one usage record); the data can be merged at the time an accounting report for the user is generated. Assuming the accounting report generator has access to the information about the linked accounts – or just uses the unique identifier of the infrastructure identity – a user will be able to obtain an overview of all the resources consumed using any of their credentials.

2.3 Traceability and security incident response

In a federation of multiple service providers that allow different authentication mechanisms, the same users may use different credentials to access different services. In the case of a security incident affecting a service provider member of a federation, the federation may want to prevent a user involved in the incident from accessing services while the incident is being investigated. The service providers need to know which identities have been linked to the account involved in the security incident, either to suspend these identities or simply to investigate whether there are other suspicious activities associated with the user's account. Without account linking, the user's accounts would be registered in the federation as different users, making it very difficult to associate the activities performed with the different credentials.

(See also the infrastructure risks associated with the use of guest identities, summarised in [AARC-MJRA1.2], Appendix A.)

3 Account linking process

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

3.1 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.

3.2 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[1], 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 case, 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.

4 Reconciling identity information

Account linking requires merging attributes from different identities into the user's infrastructure identity. For multi-valued attributes of the infrastructure identity, the merging process can be based on a simple aggregation strategy, whereby the attribute values from all linked identities are copied to the user's

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

infrastructure identity. However, even then, a user may be allowed to choose their preferred value, e.g. a preferred email address. In the case of single-valued attributes, merging requires selecting a single value from all linked identities. Whether this choice is left to the user, or is selected based on assurance, or some other policy, needs to be decided by the infrastructure.

Some infrastructures recognised the need for managing provenance of attributes in account linking and surveyed existing work in attribute metadata [NISTIR-8112] to maintain the provenance, LoA [EUDAT-Attr], and user consent for the more important attributes. Other infrastructures, such as ELIXIR and BBMRI, have identified use cases that require attribute value selection upon access to resources; for example, to support a user with multiple roles/affiliations (e.g. multiple home organisations or projects they are affiliated with) who wants to log into a service that expects a single role/affiliation. Note that one should be careful about giving the user a choice in the presentation of authorisation attributes to services, in case users elevate their privileges beyond the level to which they are authorised. The classic example is if an IdP asserts membership of a "restricted" group which is denied access to a service, and the user can choose to not assert this membership by selecting membership information from another (linked) IdP.

5 LoA elevation

Each of the external identities linked to the infrastructure identity is usually associated with a different LoA based on various properties of the authentication provider. Currently, the LoA assigned to the infrastructure identity is typically derived from the LoA associated with the authentication provider used by the user when accessing infrastructure resources. However, many infrastructures have identified the need to support the re-evaluation of the infrastructure identity LoA based on the LoA information associated with all linked identities. Specifically, the (re)evaluation model should take into account all four aspects of the LoA (see also [AARC-DNA3.1]) associated with each linked identity:

1. Identifier uniqueness (including the reassignment policy in place)
2. Identity proofing and credential issuance, renewal and replacement
3. Authentication
4. Attribute quality and freshness (primarily pertaining to the home organisation and affiliation information)

Another aspect that may be considered is the operational security of the Identity Provider that may have an impact on the LoA of the asserted identities. Sirtfi [SIRTFI] is a framework that can be used to indicate the operational security of the Identity Provider.

Models for addressing such LoA re-evaluation include:

- Linked high-LoA identity.
- Step-up authentication.

- Origin information.

Each of these is considered below.

5.1.1 Linked High-LoA Identity

The following LoA elevation flow considers all components of the LoA associated with linked identities: A user registers for an infrastructure identity with a low-LoA identity, e.g. from a social media identity provider lacking identity vetting. Subsequently, the user links their high-LoA organisational identity to their infrastructure identity. However, by linking the two identities, the user has proved that they are the same user. 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.

5.1.2 Step-Up Authentication

There are also types of linked identities that do not support sufficiently strong authentication methods for high-risk access use cases, despite being otherwise trustworthy (e.g. from the point of view of the identity vetting process). In such cases, the user may register a second authentication factor to enhance the strength of the authentication method and effectively the associated LoA (step-up authentication).

5.1.3 Origin Information

Another model for determining the LoA of a linked identity is by examining the origin information associated with the asserted attributes. While there is currently no standard way to convey such information, some identity providers have defined attribute value metadata aiming to support cross-organisation confidence in attribute assertions. One such example is ORCID [ORCID], which allows researchers to create their account either by self-registration or through institutional login (through the eduGAIN inter-federation service). In general, all the information related to affiliations, publications, awards and grants is provided by the users themselves, so the assurance level is rather low. To provide validated assertions about the users' data, ORCID started the "Collect & Connect" program [ORCID-CC], through which accounts can be connected with the researchers' home organisations; publications with the publishers; and grants and awards with funders. In this way the assertions about affiliation, authorship, and awards can be verified, and updated as well, by each authoritative source.

This feature can be exploited through ORCID's APIs [ORCID-API], which allow the retrieval of ORCID IDs and related records. Specifically, the retrieved information (e.g. affiliation) is accompanied by the source: in the case of self-asserted information, the source points back to the ORCID user; when the information is inserted and verified by the researcher's home organisation, the source points to the home organisation itself. There are, however, some limitations:

- Only the information that has been made public by the user is retrievable (at least using the ORCID public APIs).
- Matching of ORCID identifiers for home organisations with, for example, SAML 2.0 /eduGAIN entityIDs is not straightforward.

The Umbrella Collaboration is currently investigating the possibility of driving ORCID adoption within their partner organisations (14 photon and neutron sources and their aligned partners across Europe) by potentially providing it as an additional attribute (in an eduTEAMS attribute authority) on login. This is currently subject to a GÉANT eduTEAMS pilot [UMBRELLA]. Additional attributes that are useful to Collaboration partners and services using Umbrella ID via eduGAIN are under consideration.

6 References

[AARC-DNA3.1] Deliverable DNA3.1: Differentiated LoA recommendations for policy and practices of identity and attribute providers
https://aarc-project.eu/wp-content/uploads/2017/04/DNA3.1-Differentiated-Assurance.pdf

[AARC-MJRA1.2] MJRA1.2: Design for Deploying Solutions for "Guest Identities"
https://aarc-project.eu/wp-content/uploads/2016/06/MJRA1.2-Design-for-Deploying-Solutions-for-Guest-Identities.pdf

[EUDAT-Attr] EUDAT Attribute metadata - background
http://doi.org/10.23728/b2share.20c1c0c8ba254e768fbcb67724918936

[NISTIR-8112] Attribute Metadata: A Propose Schema for Evaluating Federated Attributes, NIST Internal Report 8112
https://pages.nist.gov/NISTIR-8112/NISTIR-8112.html

[ORCID] ORCID website
https://orcid.org/

[ORCID-API] ORCID web page: The ORCID API
https://orcid.org/organizations/integrators/API/

[ORCID-CC] ORCID web page: Collect & Connect
https://orcid.org/content/collect-connect/

[SIRTFI] Sirtfi website
https://refeds.org/sirtfi

[UMBRELLA] GÉANT eduTEAMS Umbrella pilot wiki page
https://wiki.geant.org/display/gn42jra3/Umbrella

7 Glossary

API	Application Program Interface
IdP	Identity Provider
LoA	Level of Assurance
SAML	Security Assertion Markup Language
SP	Service Provider