Guidelines on attribute aggregation

Published Date: 13-06-2017
Revision: 1.0

Work Package: JRA1
Document Code: AARC-JRA1.4B
Document URL: https://aarc-project.eu/wp-content/uploads/2017/03/AARC-JRA1.4B.pdf

Table of Contents

1 Introduction 3

2 Models of Attribute Aggregation 4
- 2.1 Push 4
  - 2.1.1 Examples 4
- 2.2 Pull 4
  - 2.2.1 Examples 5
- 2.3 Provision 5
  - 2.3.1 Examples 5

3 Guidelines 5
- 3.1 Persistent, unique identifiers are critical when linking records 5
- 3.2 Explicit consent for data sharing should be obtained 6
- 3.3 Attributes stored at an AA, IdP or SP post-aggregation should expire 6
- 3.4 Check attributes supplied by the user's IdP/OP and redirect users to aggregation sources if additional information is required 7
- 3.5 Consider moving aggregation "Business Logic" away from the SP 7
- 3.6 Scoped attribute values 7
- 3.7 Be cautious when using EduPersonEntitlement to store URIs 8
- 3.8 Filter attributes according to source 9
- 3.9 Attribute vocabularies should be harmonised by the aggregator 9

4 References 10

5 Glossary 10

Table of Figures

Figure 2.1: Push Model 4

Figure 2.2: Pull Model 4

Figure 2.3: Provision Model 5

1 Introduction

A user's home institution IdP can provide attribute information to the relying party/service provider he/she is accessing. However, many federated IdPs will not send enough information to meet the requirements of all RPs/SPs. Some IdP operators will not release the required information: they may have data protection concerns, restrictive policies or just slow procedures. Other IdP operators may not have the information at all, and are unable or unwilling to create or manage it. Research collaborations may have their own data for users and groups that they wish to use alongside federated authentication. It is common for VOs to create their own group and entitlement information for access control and management.

It may be important to garnish attributes received from IdPs with additional information about the originating organisation (especially for assurance purposes or to deal with differing attribute vocabularies and data quality risks)

Service providers can work-around this problem by gathering the extra information they require from other sources and aggregating it with attributes supplied by the user's federated IdP.

Although this document usually refers to SAML concepts, attribute aggregation occurs whenever more than one source of data is needed for authorisation, and can take place using any combination of authentication protocols. Aggregation can occur during "token translation", within proxies or at the destination service.

This document discusses attribute aggregation scenarios that can be applied in international research collaborations. Attribute aggregation can take place at proxy, SP or TTS services, in-line with the Blueprint Architecture.

2 Models of Attribute Aggregation

2.1 Push

Attributes can be aggregated by an IdP/SP proxy service or other authentication service before they reach the RP/SP. The SP is not directly involved. The proxy can aggregate attributes from several different sources with attributes from the user's home IdP. Push aggregation can also occur when a client x509 certificate is created with aggregated attributes and provided by a web browser or desktop application.

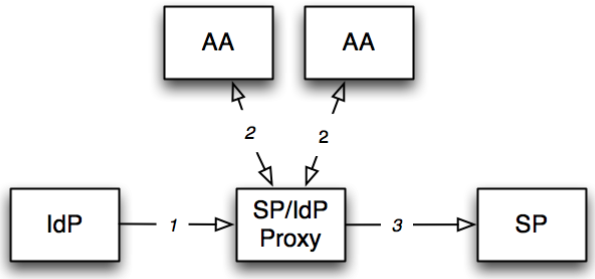

Figure 2.1: Push Model

2.1.1 Examples

SAML: The EGI CheckIn Service is an AAI proxy that can aggregate information from a user's origin IdP, EGI managed data, and from other sources. RP/SPs using the EGI IdP will receive aggregated information as part of the user's authentication.

2.2 Pull

An SP can request additional attributes for a user after receiving attributes from the user's IdP during authentication. These can be fetched from a SAML Attribute Authority, 3rd party LDAP database or another source.

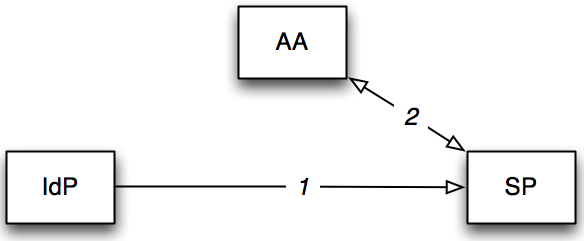

Figure 2.2: Pull Model

2.2.1 Examples

SAML: The DARIAH SP requests additional attributes from the Dariah AA in addition to the attributes that came from the origin IdP.

VOOT: The Foodle scheduling service aggregates attributes from federated logins with a call to a NREN-run groups management service using the VOOT protocol.

2.3 Provision

Similar to Pull. An SP can maintain its own store of information on users, supplied before the user authenticates. Attributes are aggregated either during authentication or afterwards by the application. This is commonly used by "enterprise" cloud applications that cannot rely on IdP-supplied data. This is not a scalable or federated approach and usually only one source of attributes is used.

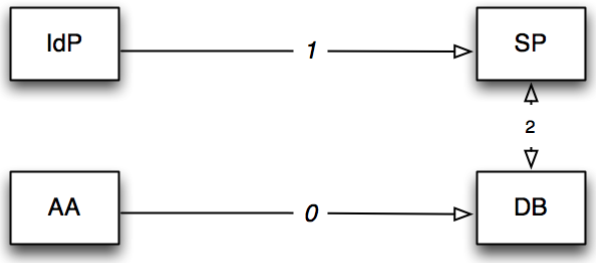

Figure 2.3: Provision Model

2.3.1 Examples

SAML: NewRelic uses SAML for authentication but only makes use of the mail attribute. Users are also required to register or be provisioned, and most user data is managed locally.

SAML + LDAP or OIDC + LDAP: NextCloud, like many web applications, can use federated authentication but also pull information from an LDAP directory. This is particularly useful for group management.

3 Guidelines

3.1 Persistent, unique identifiers are critical when linking records

- **Institutional identifiers**: eduPersonPrincipalName (ePPN) [I2-EPPN] is widely available and required by the REFEDS R&S entity category. It should be a good key to link records from different sources. However, recycling/reallocation of ePPN at some institutions creates a data protection risk. Migration to the use of eduPersonUniqueId [I2-EPUI] is preferred and should be supported by R&S.

- **Social/professional identifiers**: ORCID identifier [ORCID] (presented as eduPersonOrcid) [I2-EPO] appears to be a viable way to link to user-asserted data, and to indicate that accounts at different organisations are used by the same person.

3.2 Explicit consent for data sharing should be obtained

- It is important that users are aware of what personal information is being stored and accessed at a second service.
- Consent to share an identifier is not consent to aggregate data using that identifier. For example, a user may give consent for their ORCID identifier to be shared by their IdP, but may need to give further consent for aggregation of their ORCID data.
- The user should be informed about the attributes that will be aggregated. The user's consent to release attributes, which is usually collected by the authentication service, must be obtained in compliance with the General Data Protection Regulation (GDPR) [GDPR].
- Unnecessary data collection should be avoided. Again, this is in accordance with the GDPR.

3.3 Attributes stored at an AA, IdP or SP post-aggregation should expire

- Deprovisioning is very important. Failure to deprovision can create privacy and security risks for both individuals and organisations.
- IdPs and AAs should ideally provide expiry dates for attributes with each assertion – schacExpiryDate is an appropriate existing attribute type for this purpose.
- Aggregators should expire cached or stored records in accordance with any expiry information from the originating IdP.
- Aggregators should expire records with no explicit expiry date either in accordance with existing data protection guidelines for their organisation, or within 3 months of an update.

3.4 Check attributes supplied by the user's IdP/OP and redirect users to aggregation sources if additional information is required

- The Shibboleth SP AttributeChecker feature allows SPs to redirect to another source if inadequate data is sent by the user's IdP. This can be used to redirect a user to register with an Attribute Authority to provide (and give consent to) additional attributes.

3.5 Consider moving aggregation "Business Logic" away from the SP

If aggregation is done at the SP/RP from similar, reliable, equally trusted IdPs (maybe from within the same federation) then the aggregation can be kept simple and there's no need for more advanced logic. Attributes can simply be gathered and passed on to the application or HTTP server's access control.

The future direction of FIM (especially regarding assurance levels) requires some business logic so that data can be harmonised depending on its source IdP. At the moment, not all SP software can dynamically rewrite attribute data.

Complex aggregation rules should be moved outside the SP software:

- Rules can be moved into a proxy (especially appropriate for Push aggregation)
- Rules can be moved into the application (the best option for Pull aggregation)

3.6 Scoped attribute values

- Use of @domain scoping is limited by the strict scope-origin filtering that should be done by SAML SPs for security. A proxy may not be able to pass to an SP an attribute that is scoped to a source IdP, as the SP will, by default, only trust the original IdP to provide attributes with that scope.
- If information about the source IdP is not required and attributes have been harmonised, then scopes of attributes from suitable sources can simply be rewritten to originate at the aggregator. For example, student@aa1.edu would become student@proxyidp.com. Locally unique identifiers, such as ePPN, must not be used to create new aggregator-scoped identifiers, and if a new identifier is created, the source identifier must always be traceable.
- Registering all the origin AA and IdP's scopes in the aggregator's metadata is also possible, and may be practical even for large numbers of source IdPs for a proxy service only supporting SPs outside of a

federation (such as within a research organisation). Federations are unlikely to allow aggregating proxies to share scopes with institutional IdPs, as the aggregator would be able to impersonate any IdP it shares a scope with.

- URIs containing domains are naturally scoped. See *Guidelines on expressing group membership and role information* [AARC-JRA1.4A] for examples involving groups.

- The aggregator must verify that scopes entering the aggregator are from valid IdPs, and belong to the legitimate source.

3.7 Be cautious when using EduPersonEntitlement to store URIs

- The SAML eduPersonEntitlement attribute [I2-EPE] is intended to contain one or more URIs that indicate a specific entitlement to a resource. The very flexible nature of URIs makes eduPersonEntitlement an often effective workaround to some of the aggregation limitations of SAML assertions.

- However, this may lead to eduPersonEntitlement being used to represent the aggregated values of many other attribute types such as groups, organisation membership, roles and institutional affiliations, rather than abstract resource access rights. This can create maintainability problems.

- Try to create useful entitlements at the aggregator that are derived from source attributes, rather than storing other aggregated source attributes in eduPersonEntitlement.

- Store values aggregated as URIs in more appropriate attributes if a suitable attribute is available and the original data is needed, rather than an entitlement. Examples include identifiers, affiliation, assurance levels, and groups.

- Research communities can create their own local schemas and new attributes to store aggregated values.

- Groups are frequently used to indicate shared access entitlements, and so membership of such a group can often be safely expressed with a simple entitlement URI.

- Care must also be taken to check and filter values when passing eduPersonEntitlement through the aggregator to SPs.

3.8 Filter attributes according to source

- High-assurance, low-assurance and user-asserted attribute data should not be mixed without careful filtering.
- Filtering may also be needed to remove unknown or inconsistent values (if normalisation is not possible).

3.9 Attribute vocabularies should be harmonised by the aggregator

- The aggregator should, whenever possible, tidy and simplify the wide range of possible attribute values into a smaller, known, and more consistent set. This is especially important if a diverse set of IdPs and AAs are being used.
- The aggregator should become a Single Source of Truth ("SSOT"). There is a risk that an SP using both processed attributes (from an external aggregator, or aggregated itself) and attributes taken directly from an origin IdP, may use unprocessed data by assuming it comes from the aggregator. It should be safer for an SP to only use certain attributes from a single trusted aggregator.
- As already mentioned, creating new harmonised entitlement values from various source attributes may be more efficient and reliable than processing the source attributes and passing them on to RPs/SPs.

4 References

[AARC-JRA1.4A] Guidelines on expressing group membership and role information
https://aarc-project.eu/wp-content/uploads/2017/03/AARC-JRA1.4A.pdf

[GDPR] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

[I2-EPE] eduPersonEntitlement description
http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201310.html#eduPersonEntitlement

[I2-EPPN] eduPersonPrincipalName description
http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201310.html#eduPersonPrincipalName

[I2-EPO] eduPersonOrcid description
http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201602.html#eduPersonOrcid

[I2-EPSA] eduPersonScopedAffiliation description
http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201310.html#eduPersonScopedAffiliation

[I2-EPUI] eduPersonUniqueId description
http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201310.html#eduPersonUniqueId

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

5 Glossary

AA	Attribute Authority
AAI	Authentication and Authorisation Infrastructure
EGI	European Grid Infrastructure
FIM	Federated Identity Management
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
IdP	Identity Provider

LDAP	Lightweight Directory Access Protocol
NREN	National Research and Education Network
OIDC	OpenID Connect
ORCID	Open Researcher and Contributor ID
REFEDS	Research and Education FEDerations group
R&S	Research and Scolarship
RP	Resource Provider
SAML	Security Assertion Markup Language
SP	Service Provider
TTS	Token Translation Service
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
URN	Uniform Resource Name
VO	Virtual Organization
VOOT	Virtual Organization Orthogonal Technology