GAGE IdM Engagement Final Report

July 1, 2020 - January 13, 2021

Feb 23th, 2021

[Status: Final]

Eventual distribution: Public

Joshua Drake, Adrian Crenshaw, Susan Sons

About Trusted CI

The mission of Trusted CI is to provide the NSF community with a coherent understanding of cybersecurity, its importance to computational science, and what is needed to achieve and maintain an appropriate cybersecurity program.

About CI CoE Pilot

The CI CoE is a nationally recognized leader that facilitates discovery across NSF facilities and supports the cyberinfrastructure ecosystem that includes people, practical knowledge, and processes, and that enables research communities to attain the NSFs vision of a nation that is the global leader in research and innovation.

Acknowledgments

The CI CoE Pilot team would like to thank Trusted CI for their partnership in the identity management working group, governance and guidance in the engagement process and use of their reporting and planning templates during the engagement planning and execution. Additionally we’d like to thank Jim Basney for providing subject matter expertise on CILogon and Scott Koranda for guidance on implementing and integrating the COmanage registry and OIDC claims.

Engagements are inherently collaborative; the authors would like to thank the GAGE team including Doug Ertz, Jim Riley, Ngoc Nguyen, Mark Power, Scott Schriener, David Mencin, Donna Charlevoix, Kathleen Hodgkinson, and Susan Jeffries for the collaborative effort that made this engagement possible.

This document is a product of the CI CoE Pilot and Trusted CI Identity Management Working Group in collaboration with the US ARF. The CI CoE Pilot is supported by the National Science Foundation (NSF) under Grant Number OAC-1842042[1]. Trusted CI is supported by the NSF under Grant ACI-1547272[2]. For more information about Trusted CI, please visit: http://trustedci.org/. For more information about CI CoE Pilot, please visit: https://cicoe-pilot.org/. GAGE is supported by the NSF under Grant Number 1724794[3]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[1] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1842042
[2] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1547272
[3] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1724794

Executive Summary 3

1 Background 3

2 Factual Summary 4

2.1 Chronicle of GAGE Interviews 4

2.1.1 Doug Ertz and Jim Riley 4

2.1.2 Doug Ertz 5

2.1.3 Jim Riley 5

2.1.4 Donna Charlevoix 5

2.1.5 Susan Jeffries 5

2.1.6 Kathleen Hodgkinson 6

2.2 Proof of Concept Design Considerations 6

2.4 Implementing the Proof of Concept Design 7

3 Recommendations 8

3.1 Proof of Concept Design description 8

3.1.1 OIDC Claims Implementation Decisions 11

4 Conclusion 11

Appendix A - GAGE Organizational Access Model 13

Appendix B - The AARC Blueprint Architecture Framework 14

Executive Summary

UNAVCO/GAGE (hereafter GAGE) engaged with the Trusted CI/CI CoE Pilot Identity Management (IdM) working group from July to December 2020 in order to investigate, design, and implement a proof of concept identity management solution for a research data portal that could consume federated identities in order to grant access to GAGE research data and track which users were requesting access to that data.

Over the period the IdM team interviewed GAGE staff across several departments and investigated tools and services to include in a proof of concept design. Once the design was chosen, the GAGE and IdM teams worked together to implement it. The chosen design is based on the AARC Blueprint Architecture (AARC BPA)[4] and uses CILogon[5] and COmanage[6] to consume federated identities from the InCommon federation[7] and store them as objects within COmanage to enable GAGE to add custom attributes in order to grant specific access to hosted resources.

The use of these custom identity assertions can be tracked by GAGE for reporting purposes and, also, the services and software used in the design may be expanded to support additional authentication types in the future, such as LDAP, SAML or SSH keys. While many of the components in this design can be implemented and configured in an "off-the-shelf" fashion, GAGE will need to devote significant time and effort to modeling the organization in COmanage and developing an OIDC back end interface to issue and track tokens for its use-cases.

1 Background

The Geodetic Facility for the Advancement of Geoscience (GAGE) project is one of NSF's two premier geophysical major facilities which supports geoscience and geoscience education. It is operated by UNAVCO Inc, and provides support to the NSF investigator community for geodesy, earth sciences research and education[8]. The GAGE facility is responsible for maintaining GPS and other instruments used to measure the movement of earth's tectonic plates. Data collected from these instruments is collected and distributed as both raw and product data to researchers and the public upon request.

[4] https://aarc-project.eu/architecture/
[5] https://www.cilogon.org
[6] https://spaces.at.internet2.edu/display/COmanage
[7] https://incommon.org/federation/
[8] https://www.unavco.org/what-we-do/gage-facility/

In the first half of 2020, GAGE reached out to Trusted CI for an engagement that would assist them with identifying the users of GAGE data in order to comply with a request from the NSF. In their application they stated that they would like to investigate federated identity providers to allow users to self-identify using existing institutional credentials and avoid creating and managing GAGE specific accounts. In partnership with the CI CoE Pilot, the jointly operated Identity Management Working Group was tasked to engage with GAGE to help navigate the implementation and usage of a federated identity service for this purpose.

2 Factual Summary

The Identity Management Working Group Team (IdM team) began meeting with representatives from GAGE in June of 2020 to collect objectives for the engagement and begin drafting an engagement plan. The engagement plan was approved on July 15th with the following objectives:

- Improve the state of Identity Management practice and infrastructure at GAGE by producing a plan and proof of concept for a new IdM system.
- Provide resources to assist the wider NSF major facility community in solving similar problems

A plan of work to achieve these objectives was laid out in timeline form, consisting of two phases, an initial series of interviews and discovery, and a second phase of designing and testing a proof of concept for using federated identities to track research data usage.

2.1 Chronicle of GAGE Interviews

Interviews with GAGE personnel were conducted between July 8th and August 12th.

2.1.1 Doug Ertz and Jim Riley

On July 8th the engagement team met with GAGE Project Manager Doug Ertz and Technical Lead Jim Riley to discuss the nature of services provided by the GAGE project to the NSF community with the following takeaways:

- Most GAGE users today, and all GAGE users in the future will log in via a web portal in order to view and download GAGE data. The NSF would like GAGE to track who is accessing the data via federated identity.
- Existing FTP servers will be decommissioned.
- GAGE intends to concentrate many if not most services in the cloud in the future and is working to determine best practices for cloud security and IAM with an emphasis on simplified workflows.

- GAGE is currently using G suite for email, document storage, and federated identity.

2.1.2 Doug Ertz

On July 15th the engagement team met again with Doug Ertz to discuss and determine what systems need to be integrated with the proposed IdM solution. Priorities for integration are:

- Client side web application for finding, viewing and downloading research data
- FTP servers (PureFTP or server at IU running SFTP)
- Enterprise systems currently using Active Directory(AD) LDAP authentication

2.1.3 Jim Riley

On July 22nd the engagement team met with Jim Riley to have a discussion concerning the research data web site and services.

- The researcher data portal is currently in development at GAGE
- The portal will use a client side web application developed in angular which runs in the browser to authenticate users and provide access to various research data.
- The preferred workflow is that users could browse the site and search for data without being authenticated, but would need to log in in order to download data.

2.1.4 Donna Charlevoix

On July 29th the engagement team met with GAGE Communication and Education Director Donna Charlevoix to determine which forms of communication GAGE uses to contact stakeholders.

- GAGE keeps in contact with the scientific community online in a variety of ways including email, social media, and web presence.
- GAGE also uses some more traditional educational pathways including providing curriculum guidance and by offering internships.
- GAGE’s primary stakeholders are USGS, NOAA, and JPL, but they also connect with independent researchers from academic institutions around the world, as well as the general public.

2.1.5 Susan Jeffries

On August 5th the engagement team met with software engineer Susan Jeffries to discuss additional requirements for GAGE systems. Susan is primarily responsible for GAGE’s business and inventory systems, which includes data from both GAGE systems and client equipment.

- The current inventory system was developed at GAGE out of a need to track equipment at other entities GAGE/UNAVCO was working with. It has evolved over time to include support for project support requests, equipment inventory and management, registry tracking for events and meetings, and accounting management.

- It currently uses AD for authentication, which has led to a lot of non-UNAVCO employees being added to AD.

2.1.6 Kathleen Hodgkinson

On August 12th the engagement team met with Kathleen Hodgkinson, Data Products Manager to determine how different components of the system connect.

- GAGE uses a system known as “real-time (RT)” to receive seismic data from its remote instruments and record that data in retrievable databases.
- The RT system connects to an RT caster (sensor) and requests data using the NTRIP format over an http connection.
- RT data is stored and can be accessed directly in the application or used to generate reports.
- Some sets of data are restricted to certain sets of users, but most data is not restricted.
- Access to the RT and reporting systems is handled within the application itself via credentials stored in a SQL database
- These credentials are only used for connecting to RT casters.
- In the future it is planned that access to the data sets and reports will be handled by the web application that Jim Riley’s team is developing.

2.2 Proof of Concept Design Considerations

At the conclusion of the interview and information gathering process the engagement team presented GAGE with IAM options for further investigation. During the second half of August the teams investigated various services and toolkits to be used in the proof of concept design including the CILogon proxy service, COmanage identity registry, Grouper, Auth0, Okta, OpenID Connect, SAML and Shibboleth.

Through the interview process the IdM team identified several driving factors in GAGE’s current and future IdM needs.

- The design should be cloud compatible
- The design should utilize simple workflows
- The design should allow authentication through multiple standards (LDAP, key based, OIDC Tokens)
- The design should use open standards to avoid vendor lock-in

In addition to the need to identify who is using their data, GAGE also has several legacy and adjoining systems in use or under development which would benefit from being able to use federated identities for authentication, such as the real-time system and the business unit applications developed in house. There is also a future need to consider a single identity management system that can facilitate both external and internal authentication for local and cloud based applications. While the design described below does not specifically address that consideration, the architecture and products chosen were selected with those future considerations in mind.

At the end of these discussions the IdM team presented a proof of concept design that would allow GAGE's new data portal (named DAIv3) to utilize OIDC tokens from CILogon using federated user identities. This system would leverage COmanage in the background to track and uniquely identify GAGE users, and could also be expanded to provide authentication credentials for other GAGE systems using OIDC, LDAP or key based authentication, as well as allowing interoperability with Active Directory, MS LDS[9], and LDAP systems[10]. For more detailed information on the proof of concept design, see Section 3.1.

2.4 Implementing the Proof of Concept Design

Once the design had been approved by GAGE the teams worked together from September to December to implement the design. In September the teams met with Senior Research Scientist and COmanage consultant Scott Koranda to review the functionality of COmanage, the proposed use cases for GAGE, and to configure access to a testing environment for CILogon and COmanage.

On September 16th the teams gained access to COmanage and began to model the GAGE organization groups and users. The proposed solution makes use of user groups in COmanage to add attributes to OIDC tokens, which the application can consume to grant access to various resources. While only one group was used for the proof of concept implementation, five separate groups and associated access permissions were planned for future expansion (see Appendix A).

Both teams agreed that the best target for integration with the concept design was the GAGE web application for research data access, DAIv3. The DAIv3 application is a client side web application using the AngularJS framework. At this point two integration options were considered:

- Using an apache OIDC module[11] on a separate backend server to provide OIDC tokens to the user's browser.

[9] https://docs.microsoft.com
[10] https://www.openldap.org/
[11] https://www.mod-auth-openidc.org/

- Developing a custom backend application based on the globus reference implementation in flask[12] which can handle requests for OIDC tokens and handle tracking token requests.

Due to time and resource constraints during the engagement period we decided to use the simpler apache module so that we could have a working OIDC authentication workflow operational before 2020 holidays disrupted the meeting schedule. However, for the production implementation we recommend that a custom version of the Globus Toolkit reference implementation be included in the final production design.

Using an apache server and the mod_auth_oidc module, Jim Riley set up an OIDC URI and configured CILogon and COmanage reference URIs to pull groups attributes from COmanage and include them as assertions for the OIDC claim token. Test users could then be created in COmanage using federated identity via CILogon and used to test enrollment workflows, administrator assigned and self-service election to access groups, and the passing of group membership attributes to applications via the OIDC token assertions.

To track users based on the OIDC token claims, the teams determined that logic would need to be included on either the backend apache server or within the DAIv3 application to track who was requesting download access to research data sets. The most scalable solution we discussed involved using flask and the globus reference implementation for OIDC to create a separate backend service to handle OIDC claim requests and track which applications were requesting access. A full discussion of this implementation is included in Section 3.1.1.

3 Recommendations

The following section details the proof of concept design that the IdM team presented to GAGE, as well as changes based on lessons learned during the test implementation process.

3.1 Proof of Concept Design description

The scope of this proof of concept design encompasses the systems and services that need to be in place to consume a federated identity from an identity provider, store a unique identity in a GAGE Comanage registry instance and add custom assertions to it, and to generate and pass OpenID Connect (OIDC) tokens to other applications to grant access to resources hosted by GAGE. This concept design also considers GAGE's future need to operate software in the cloud without being tied to a specific cloud vendor, and allows for expansion of both identity

[12] https://docs.globus.org/api/auth/reference/#user_authorization_and_authentication_with_oauth2oid

providers (federated and local enterprise sources) and additional methods of providing digital identities to resource providers beyond OIDC.

The first requirement of the design is a method of retrieving federated identities from trusted institutions. While several options exist in this space, we recommend an AARC blueprint architecture (BPA)[13] proxy service like CILogon[14] (see Appendix B). The AARC BPA is “a set of building blocks that can be used to implement federated access management solutions for international research collaborations” that “lets software architects and technical decision makers mix and match tested components to build customized solutions for their requirements.”

CILogon is a service which implements the AARC BPA in order to allow users to consume SAML and OIDC identities from identity providers such as InCommon, Shibboleth, Google, GitHub, ORCID, etc. CILogon is specifically designed for research cyberinfrastructure operators and operates as a subscription based service.

To store and add attributes to collaborator federated identities, the design uses COmanage Registry[15]. COmanage Registry is an identity enrollment and lifecycle management application that can consume federated identity assertions from CILogon or other AARC BPA identity provider services and store them as an identity object in a registry so that the COmanage operator can add unique attributes for granting access to hosted resources at resource providers and other hosted services.

COmanage registry can be run as a locally hosted instance, a cloud hosted instance, or hosted by COmanage. For the test implementation we used the COmanage hosted option, but any implementation strategy could be used for a production instance.

When implementing COmanage, GAGE should take time to carefully model its organization and the logical divisions between different types of users and different types of access that GAGE users may be granted. During our test implementation process we identified five such access groups (see Appendix A), but several different implementation models should be considered and tested during the production implementation.

COmanage uses several enrollment models to consume federated identities and create local identity objects for the user. GAGE can design enrollment workflows for their users which best

[13] https://aarc-project.eu/architecture/
[14] https://www.cilogon.org/
[15] https://spaces.at.internet2.edu/display/COmanage

suit the needs of GAGE and the level of access the user desires. Possible enrollment flows include, but are not limited to:

- Self-service enrollment
- Email invitation based enrollment
- Shared link enrollment
- Administrator enrollment
- Legacy identity import

Based on testing during the proof of concept implementation we recommend GAGE offer external users the ability to self enroll to GAGE's COmanage registry, and provide links from the COmanage homepage for users to request additional access via self-service enrollment with administrator approval. This would allow users to generate an entry in the GAGE COmanage registry without administrator intervention, but would still allow for GAGE personnel to approve any form of privileged access if desired. As part of implementing and customizing COmanage, it may also be beneficial to train 2-3 staff members on COmanage through InCommon[16].

Once a user has enrolled in COmanage and been placed in a group, COmanage can be configured to add the user's list of group memberships as an attribute to their user object in a field named "isMemberOf". This attribute is then included as an OIDC token assertion or as a field in an LDAP registry entry and passed to other systems that consume OIDC tokens or LDAP attributes.

In order to set up a full OIDC workflow, GAGE must host a URI for ODIC claims to be directed to from COmange and CILogon. For the proof of concept implementation we used an apache server running basic auth and an OIDC module called mod_auth_oidc with a minimum of customized configuration. In this configuration users directed to the GAGE OIDC URI from COmanage or from the resource providing application would receive an OIDC claim token in their browser if they are registered with the GAGE COmanage instance. If the user is not registered the application or OIDC URI can implement logic to direct the user to the self service portal for enrollment or privilege escalation requests.

3.1.1 OIDC Claims Implementation Decisions

As described in section 2.4 above, for the proof of concept deployment we implemented OIDC using an apache server and the mod_auth_oidc module[17], with a minimal amount of custom

[16] https://www.incommon.org/academy/
[17] https://www.mod-auth-openidc.org/

configuration according to CILogon's recommendations[18]. This method is sufficient for generating an OIDC claim token in the user's browser, but is limited in options for managing the TTL for the token, and requires the resource provider website or application to implement its own logic for tracking token usage for reporting purposes.

To have a more complete design which provides more control over the claim tokens and allows centralized reporting on which identities were requesting access to various GAGE resources, we recommend an implementation model which handles OIDC claims using a lightweight web application to complete the OIDC three way handshake with COmanage and the resource providing application and to track which applications are requesting tokens for a given user identity.

The Globus Toolkit includes a reference implementation[19] for such a service which could be further developed to suit GAGE's needs. This approach should allow an application or website at GAGE to direct authentication requests to an API, and fulfill those requests with OIDC tokens generated from CILogon using the custom attributes shared in COmanage. This approach is identity provider agnostic should GAGE wish to work with a different provider in the future such as Auth0 or Okta.

This approach has the added benefit of allowing GAGE to implement identity tracking at the authentication service rather than at the resource provider level. The authentication API can issue short lived, minimum permission tokens that may only be used for the currently requested task, while tracking all such requests as they are fulfilled in one central place for ease of reporting. While it was not possible to implement this approach during the engagement due to time and resource constraints, the IdM working group would be interested in working with GAGE in a future engagement to further explore this topic.

4 Conclusion

To implement the proposed proof of concept design in a production environment, there are a few obstacles left to address that were not within the scope of the current engagement.

As discussed in section 3.1.1, the implementation of an OIDC infrastructure at GAGE that can handle claim token requests in a robust manner and track their usage, either at the OIDC claim token service layer or the requesting application layer is central to fulfilling the tracking part of

[18] https://www.cilogon.org/oidc

[19] https://docs.globus.org/api/auth/reference/#user_authorization_and_authentication_with_oauth2oidc

the services purpose. This part of the implementation is likely to require the most effort and time of a production implementation.

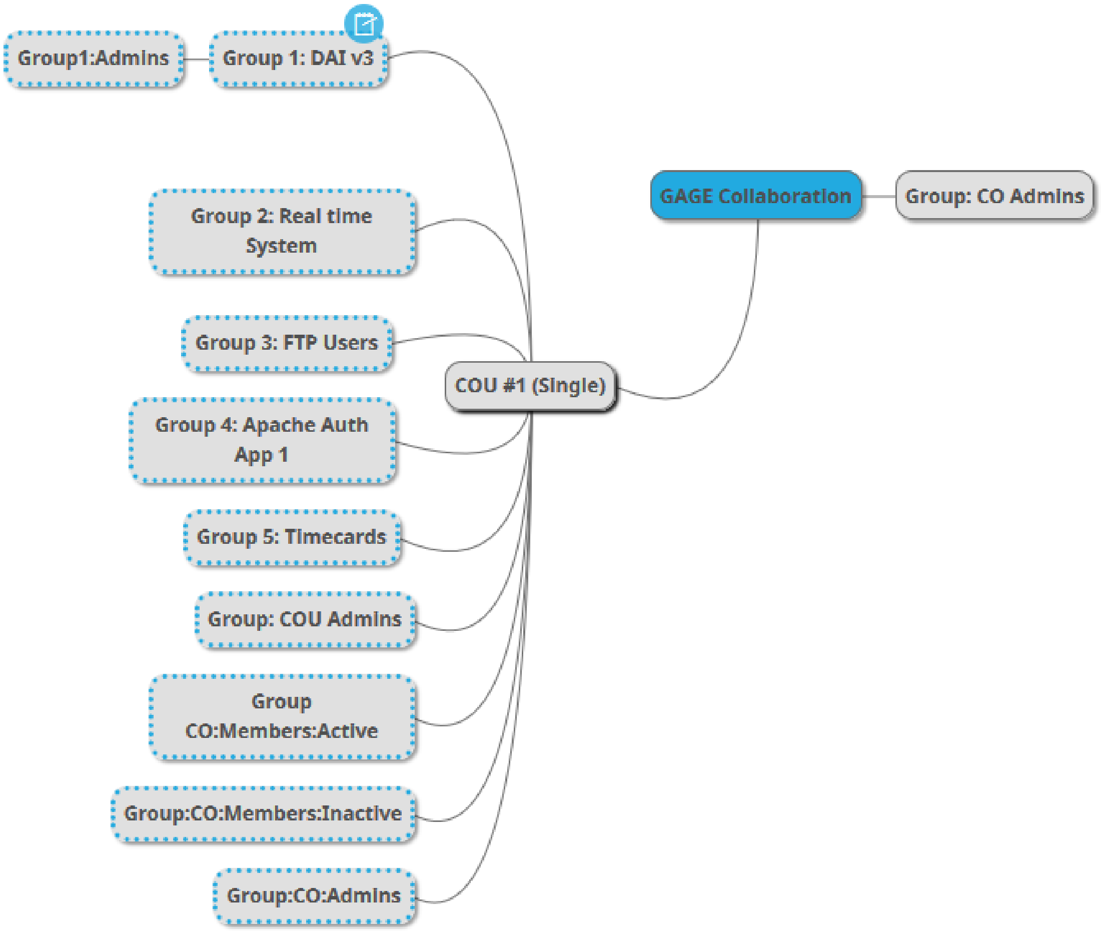
Secondly, the development of a model organizational structure for GAGE in COmanage should be given priority consideration before implementation due to its foundational importance to other aspects of the design implementation. If needed, GAGE could work with COmanage implementation engineers[20] for additional consulting if they elect to implement the COmanage registry.

Finally, additional consideration should be given to the most scalable and adaptable method of implementing the OIDC claim service and COmanage registry given the rapid changes anticipated at GAGE in the coming year. The decision of which identity provider to use and whether to host the OIDC claim service and COmanage locally or in the cloud should be carefully considered in light of what role those two services are anticipated to fulfill in the near and long term environment.

[20] https://sphericalcowgroup.com/comanage

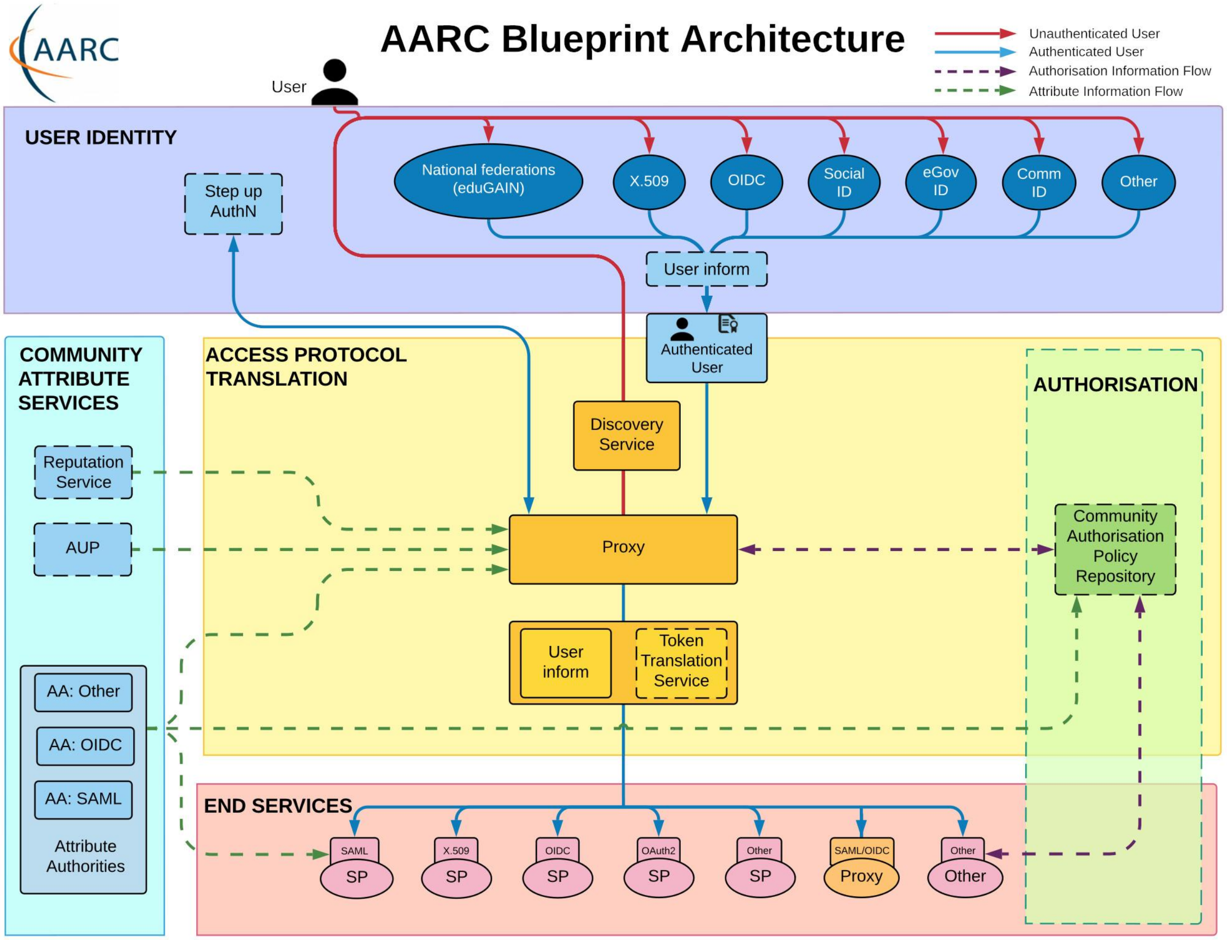
5 Appendix

Appendix A - GAGE Organizational Access Model



GAGE organizational diagram depicting a flat group structure of five access permission groups under a single organizational unit.

Appendix B - The AARC Blueprint Architecture Framework


AARC
AARC Blueprint Architecture
Unauthenticated User
Authenticated User
Authorisation Information Flow
Attribute Information Flow
User
USER IDENTITY
Step up AuthN
National federations (eduGAIN)
X.509
OIDC
Social ID
eGov ID
Comm ID
Other
User inform
Authenticated User
COMMUNITY ATTRIBUTE SERVICES
ACCESS PROTOCOL TRANSLATION
AUTHORISATION
Discovery Service
Reputation Service
AUP
Proxy
Community Authorisation Policy Repository
User inform
Token Translation Service
AA: Other
AA: OIDC
AA: SAML
Attribute Authorities
END SERVICES
SAML
SP
X.509
SP
OIDC
SP
OAuth2
SP
Other
SP
SAML/OIDC
Proxy
Other
Other