30-06-2024

Deliverable 8.1:

AARC TREE Strategy document - Who, what, why?

Publication Date	30-06-2024
Due Date	30-06-2024
Authors:	L.Florio (NORDUNet), D. Groep (Nikhef), N Liampiotis (GRNET), D. Kelsey (STFC), I. Fava (EGI)
Version:	1.0
Document Code:	AARC-TREE D8.1
Publishing Organisation:	NORDUNet
DOI:	N.A.

Abstract

This document describes the strategic approach of the AARC TREE project and a list of activities to advance the AARC BPA Framework and to deliver recommendations to ensure the sustainability of Authentication and Authorisation Infrastructures for the scientific community.

The AARC-TREE project is co-funded by the European Union under the HORIZON-INFRA-2023-DEV-01 call

Table of Contents

1. About AARC

The authentication and authorisation infrastructures (AAIs) for research worldwide have for several years now based their architectures on the “AARC Blueprint Architecture” (AARC BPA) and the suite of accompanying guidelines. The first version of the AARC BPA was developed under the first AARC project (2015-2017), further enhanced under the second AARC project (2017-2019) and subsequently maintained by the AARC Community.

The AARC BPA model, fostered by the accompanying “Engagement Group For Infrastructures - AEGIS, has been the reference model for the European Open Science Cloud AAI, many national research AAIs, and research and e-infrastructures in Europe, the Americas, and the Asia-Pacific region. However, with the increased scope and complexity of novel federated identity models, come new challenges.

This document highlights crucial factors that play a key role in enabling broad access to services that are relevant for the Research and Education (R&E) community and presents the AARC TREE strategy to address them. The document also highlights how the AARC TREE project supports the AARC community in their journey to enhance the AARC BPA (and its guidelines) so that the AARC BPA can continue to remain a reference architecture for those developing and operating AAIs in the research community.

In the context of this document, the terms research infrastructure or research communities are used interchangeably to indicate thematic collaborations (whether legal organisations or not) that manage resources for users working in that specific field. The e-infrastructures, in the R&E community are those organisations that offer horizontal services in terms of networking, identity, computing and data management to researchers.

2. AEGIS and the AARC Community

Before delving into the specific challenges that the AARC TREE project wants to support, it is important to stress the role of the AARC Engagement Group for Infrastructures (AEGIS).

AEGIS [AEGIS], the forum for research and e-infrastructures that deploy an AARC-BPA compliant Authentication and Authorisation Architecture (AAI), was created at the end of the second AARC project to address the need for sustainability of AARC results. Since 2019, AEGIS meets monthly to endorse AARC guidelines that are discussed in the AARC policy and AARC architecture Working Groups, which are also a spin-off of the previous AARC projects.

This has created a community of operators of AARC BPA-compliant AAIs that includes engineers, policy, security and identity management experts that recognise the need to adopt common standards and community best practices when deploying AAIs for research collaborations. This community has continued to work on the AARC BPA, on the AARC guidelines, and have promoted the need for interoperable frameworks for AAIs for science to enable research and collaboration at scale.

The AARC community recognises that researchers across all scientific disciplines depend increasingly on online access to resources and international collaborations for their daily work. National, European and international Research Infrastructures provide the digital environment within which researchers can find, access, and share data and resources in ways that had never been possible before. Seamless authentication and authorisation to access data and resources as if they were available right there next to the researcher, is key for enabling e-Science at scale.

Obviously, the AARC community does not work in isolation. On the contrary, great effort has been made to ensure that the AARC work is promoted internationally and inputs are sought by relevant stakeholders such as the ESFRI Clusters [ESFRI], EOSC [EOSC], FIM4R [FIM4R], REFEDS [REFEDS], IGTF [IGTF] and other relevant groups.

The strategy of the AARC TREE project is inherently linked with the overall strategy of the AARC community, which is about ensuring that AARC BPA and the guidelines can continue to offer a de facto standard for those who deploy AAIs in the R&E community.

3. Why Federated Access

The AARC BPA champions federated access as a user-friendly yet secure way to access resources at scale. Federated access works by redirecting users to their home organisation or Identity Provider (IdP),   such as their university or research lab, for authentication when they try to access a resource. Upon successful authentication, a limited set of information about the users is transferred to the resource where the authorisation takes place.

Federated authentication is still the preferred access technology as it enhances the user experience: users benefit from a seamless experience by logging in once and accessing multiple applications and services without repeated logins. Federated access also reduces the risk of phishing and password theft by minimising the number of times users need to enter their credentials, which they do only at their known, preferred Identity Provider; the transfer of users' personal information to resource providers is regulated by the principle of data minimisation, which means that only necessary information is transferred.

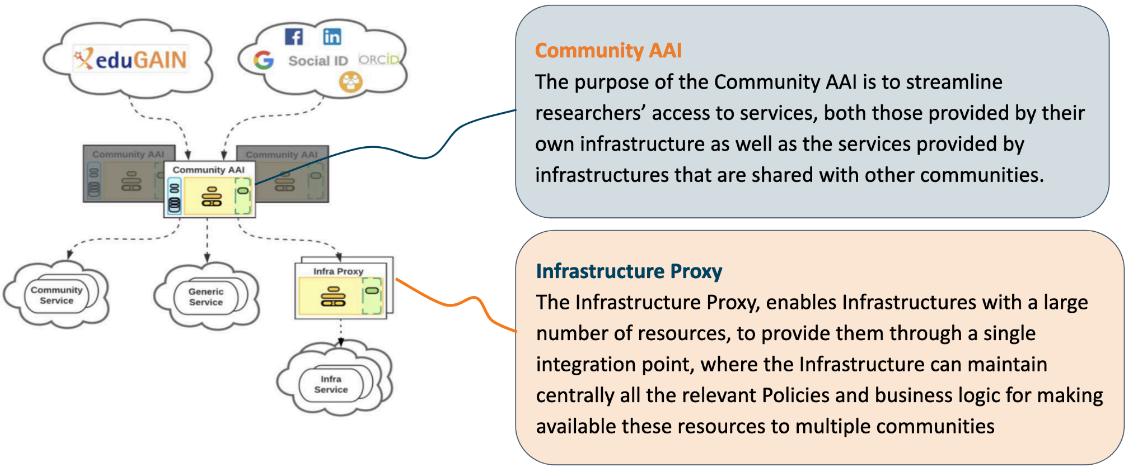
AARC builds on the eduGAIN inter-federation infrastructure, that connects a large number of trusted Identity Providers (that participate in their national federations) worldwide and makes it possible for users in the academic community to collaborate and access services and resources globally, using the same credentials that they use at their local institutions. AARC expanded eduGAIN to better support international research collaborations that span across multiple countries and national identity federations. The main added value of the AARC BPA is that research collaborations can manage access to their own resources based not only on the information provided by the users' IdP but also on the role that the user has in a specific collaboration; this information is managed by the collaboration directly.

4. Current challenges for the AARC community

The main challenges are summarised below.

4.1. Evolution of the BPA and Technical Interoperability Framework

The single ‘AAI proxy’ model of the initial AARC blueprint – which combined identity sources, community collaboration management, authorisation controls, and service provider connections - evolved in [AARC-BPA-2019] to include both ‘**infrastructure proxies**’ to provide coherency on the service provider side, as well as ‘**community proxies or Community AAIs**’ focussing on membership management (see Figure 1).



Community AAI
The purpose of the Community AAI is to streamline researchers’ access to services, both those provided by their own infrastructure as well as the services provided by infrastructures that are shared with other communities.

Infrastructure Proxy
The Infrastructure Proxy, enables Infrastructures with a large number of resources, to provide them through a single integration point, where the Infrastructure can maintain centrally all the relevant Policies and business logic for making available these resources to multiple communities

Figure 1: AARC BPA 2019

Yet the challenges keep coming at an ever-increasing pace, both in terms of complexity, technology, policy and in the range of communities that can benefit from this coherent ‘AARC approach’ to federated access management.

In addition to SAML federations, such as those for research and education that are part of eduGAIN, other technical models are gaining adoption, such as OpenID Connect Federations. A multitude of identity sources (academic, governmental, and self-sovereign) are being used and need to coexist and be linked together.

Different research communities and infrastructures may have different authorisation policies that govern how access to resources (including sensitive data) is granted. Mapping group memberships and roles managed by the community to capabilities required for accessing infrastructure resources can be a complex task. Furthermore, the lack of standardised mechanisms for requesting specific identity claims or claim values can lead to inconsistencies and confusion for users when accessing resources across different infrastructures. These challenges can create barriers to efficient and effective sharing of resources, hindering collaboration and cooperation across scientific disciplines

4.2. Global Policies for the AARC BPA

Policies and good practices enable interoperability and distribute responsibilities between the participants in the AARC Blueprint Architecture model. A clear set of guidelines is foundational for the mutual trust between the participants, but the evolution of the BPA will change the responsibilities and introduce more parties to the AAI besides the 'community' and 'infrastructure' proxies: multi-tenant proxy-as-a-service operators, but also other sources of identity assurance from government e-identity systems, translational trust based on new (OpenID) federation models, and even the user itself in case of self-sovereign identities. Hence, updates to the Blueprint Architecture require an evolution of associated policies and good practices for Research Infrastructure proxies and for the community organisation and identity assurance in the infrastructure.

Good practice policy evolution needs to address several new challenges, including

- baseline operational security trust in the evolved 'multi-hop' proxy infrastructures: as more parties are involved in e.g. security incident response, timeliness of communications and 'chain responsibility' come under stress;
- the balance between transparency versus coherency for proxies: can a proxy operator take liability for all its back-end services, or all its connected identity sources, and provide a coherent but opaque translation, or should, on the contrary, the source and assurance level of the parties on the other side be exposed, and does that depend on the research community use case;
- streamlining the user interaction across proxies so that the number of interstitial screens and interruptions is limited: the compound nature of proxies today leads to each party independently collecting regulatory compliance information via the display of privacy notices, acceptable use policy and terms and conditions of use, and collection of informed consent where applicable;
- more diverse communities, including smaller collaborations and more dynamic and evolving membership structures in projects: with the BPA giving a central role to the proxy and its coherent attribute services, the community plays a more important role in defining its own structure, but the smaller communities lack the effort and people to invest time in expressing their internal structure for access to services, leading to the unclear definition of their community structure with unclear or ad-hoc access to needed services as a result;
- difficulty in sourcing assured identity information from institutional identity providers in line with the requirements of high-security infrastructures: whereas the original federation architecture assumes that the home institution through its identity provider is an authoritative source of identity information, the assurance level - while 'probably' good - is not consistently expressed and hence not a reliable element in access decisions for sensitive research infrastructures.

The existing Policy Development Toolkit, which aimed to provide some reference documents for research infrastructure whilst very positively received, was found to be too complex and will need to address these challenges.

4.3. Sustainability for AAIs for Research and Education

Running an AARC BPA complaint AAI is a non-trivial task. Many research infrastructures have chosen to outsource the operations of their AAI to e-infrastructures (such as GÉANT and EGI) while retaining control of the policies.

Even with the authentication being done by the users' home organisations and the operations being 'outsourced' to other parties, there is still a need for research collaborations to manage controlled access to their services and to understand the relevant AARC BPA guidelines.

Operating services at an international scale requires resources and long-term planning. Solutions with sustainable cost models that can meet today's and tomorrow's research requirements are needed. There is consensus that the global interoperability of the AAI infrastructures must be strengthened to avoid fragmentation and unnecessary duplication.

Bringing the RIs, e-Infrastructures, including the right people with operational experience as well as all relevant stakeholders together is a must to align strategies to integrate new technologies, better interoperate and share resources across thematic areas. Whilst efforts to this end are taking place, there is still a fragmentation of how the activities are carried out. A more holistic approach that starts from the research needs in terms of access and sharing of resources, how these needs can be monitored and how to deploy solutions that address these needs is needed.

4.4. Diversifying Adoption of the AARC BPA

Whilst the primary target for the AARC community is to enable support for eScience, it is vital for AARC BPA to have an inclusive approach and to recognise the wider developments in the identity space.

On one hand, the updated AARC BPA will reflect the needs of Research infrastructures and strive to address the main requirements of the different players in the AAI research landscape.

On the other hand, the AARC TREE project will actively liaise with other initiatives via its partners to promote the AARC BPA model and seek for strategic alliances that span beyond the R&E sector.

For instance, support for Digital Identity Wallets can provide researchers with a convenient and secure way to manage their personal information, increasing trust and enhancing security when accessing resources. AARC TREE can explore how to extend AARC interoperability to other relevant areas, such as 'SIMPL' data spaces which also rely on decentralised identity management.

5. AARC TREE strategy to address the challenges

AARC TREE together with the AARC community will address the challenges above by delivering the following three Key Exploitable Results:

- Updated AARC BPA
- Updated interoperability framework
- Recommendations for sustainable adoption of the AARC BPA

5.1. Updated AARC Blueprint Architecture Framework

The AARC BPA has evolved into something more sophisticated than just technical architecture. With the aim to strive for deployment, it has transformed into a framework that encompasses policies, technologies, operational aspects, security and community best practices.

One of the main results of the AARC TREE project will be a new version of the AARC BPA that will meet new requirements of research infrastructures and explore new needs arising from EU and global initiatives and national infrastructures. The surveys planned for 2024 will be a great instrument to gather emerging requirements.

The updated BPA specifically will offer up-to-date guidance on how to support new technologies, standards, and best practices, informed by lessons learned from previous implementations of the BPA. AARC TREE will leverage the state of the art of technologies such as OpenID Connect Federations, advanced federated authorisation mechanisms and decentralised identity management as well as provide solutions for the harmonisation of community user attributes.

Different research communities and infrastructures may have different authorisation policies that govern how access to resources (including sensitive data) is granted. Mapping group memberships and roles managed by the community to capabilities required for accessing infrastructure resources can be a complex task. The updated version of the AARC BPA will develop common procedures for authorisation.

5.2. Updated interoperability framework

The AARC interoperability framework consists of policy and technical guidelines to support a consistent implementation of the AARC BPA.

The methodology to achieve acceptance of the evolved guidelines relies on consistent participation of the stakeholder groups through AEGIS, FIM4R, ESFRI (the majority of the ESFRI clusters is represented in AARC TREE), WISE, IGTF, and the joint user engagement actions coordinated with the GEANT Project’s *Enabling Communities* for Research activity.

The AARC TREE project is committed to developing the guidelines in an open and transparent manner. This approach fosters community adoption and ensures the long-term sustainability of the project's outcomes. Here's how openness will be implemented:

- Public development process: All project outputs will be developed publicly, with clear version history and a trackable process.
- Open communication channels: The project will utilise open mailing lists to facilitate communication and information sharing.
- Open access submissions: Project outputs will be submitted to open-access repositories like Zenodo.

To ensure the effectiveness of the updated guidelines, the project will conduct real-world validations through cross-RI pilots and develop an automated validation suite. Partner feedback and continuous input from the research and education community will be crucial for successful adoption of the developed guidelines.

5.2.1 Technical interoperability framework

Technical interoperability is shaped by guidelines that detail specific architectural aspects. Over the years, AARC projects first and AEGIS later have developed a number of guidelines. One of the tasks of AARC TREE is to define a minimum set of recommended AARC TREE technical guidelines. In addition, a new BPA will require additional guidelines.

The lack of standards and uniform approach for expressing certain community user attributes leads to inconsistencies in how these attributes are communicated across different infrastructures. This can hinder the seamless flow of information between them, causing errors and other operational inefficiencies that can negatively impact user experience and overall interoperability. To this end, the task will develop guideline documents for standardising how to express community attributes in commonly used protocols, including OpenID Connect and SAML. These guidelines will cover topics such as expressing authenticating authority information, service account information, and identity assurance information.

To support the growing interest in OpenID Connect Federation. AARC TREE will develop a deployment profile for the OpenID Federation specification that will allow entities based on OpenID Connect to participate in Identity Federations without the need for protocol translation. The profile will cover topics such as expressing compliance with entity categories (e.g. REFEDS Personalized) and security frameworks (e.g. Sirtfi), best practices for managing OAuth2.

5.2.2 Policy interoperability framework

Policy interoperability is achieved through comprehensive policy guidelines developed over previous iterations of AARC. AARC TREE will build upon this foundation by providing new guidelines and updating the existing Policy Development Kit (PDK). The PDK offers policy templates that give RIs a head start in deploying the AARC BPA, ensuring they do not have to reinvent the wheel. These templates serve as a starting point, streamlining the adoption of consistent and interoperable policies across RIs.

AARC TREE takes a two-pronged approach to the policy development kit evolution.

The first approach enhances support for the proxy *operators* on how they can act consistently in the mesh of identity providers, services, community- and infrastructure proxies, and the different types of stakeholders (research-performing organisations and universities, AAI service providers, data and computational infrastructure organisations, governmental organisations, and the end-users themselves). This includes also the requirements for secure operations of the proxy, since operational trust and baselining is a prerequisite for the AAI proxy operators to be accepted in the ecosystem as trusted third parties. We pursue this by creating extended guidelines on Attribute Authority (Proxy) Operational Security, including token validity periods, conveyance of transparency for access attributes and 'upstream' provider categorisation. Also, primarily for proxy operator implementation guidance, but ultimately for the benefit of the end-user, we have scheduled guidance on the presentation of (joint) Terms and Conditions and Privacy notices, and how to combine and update these as the proxy attaches new services (or new communities).

The second approach emphasises the community-centric elements of the AAI trust fabric: community self-organisation, guidelines on access management structures in communities, and enabling higher levels of identity assurance for individuals in those communities. This avenue foresees guidance documents on a simplified policy development kit, enhancing the differentiation of policy, procedures, and best practices - with a particular view to smaller and emergent communities (for which the term 'community' may not even be the appropriate term). Based on demonstrators and the evolution of the BPA architecture, it will also include new guidance on user-centric identity and assurance, and how new federation models affect the trust chain from end-user, via a chain of proxies, towards services.

Both aspects are driven by the use cases collected in the AARC TREE surveys, although the second (community-centric) guidelines are by their nature more affected by the survey results than the baseline security and proxy operations - for which the requirements are more readily collected via AEGIS. The implementation of the community-centred elements and the PDK will hence follow later in the project, once results from the surveys are available and analysed. To ensure the guidelines match community needs, for both approaches high-level 'conceptual framing' will be added, either as an extended introduction or as a separate 'Informational' AARC document in the series.

5.3. Compendium for sustainable adoption of the AARC BPA

The AARC TREE project will address the challenge of supporting the deployment of AARC BPA-compliant AAIs in the research and education environment and will promote the AARC BPA in other sectors. The surveys planned in the AARC TREE project will enable the team to gather intelligence on the eScience needs and challenges the research communities are facing in sustaining the operations of complex AAIs.

The project has allocated effort to use the results of the surveys to compile a **compendium** that will be maintained beyond the lifespan of the AARC TREE project. The compendium will provide recommendations to align strategies to ensure interoperability across AARC BPA AAIs. It also aims to address sustainability aspects related to the operation and the adoption of AARC guidelines and standards in AARC BPA AAIs.

The Compendium will maintain the right balance between technical and policy direction to appeal to a diverse audience and provide the necessary guidance for readers, whether they are technical architects or policy makers.

6. Conclusions

The document presented the AARC TREE strategy and a list of activities to implement it.

The document intentionally provides extensive information to ease the reading and the understanding, particularly for those not familiar with AARC BPA and its guidelines.

The document has primarily addressed the “what” and “why” to explain the rationale behind having a strategy. The “who” is less explicitly defined since it encompasses all entities involved in executing the strategy. Initially, the AARC TREE consortium plays a key role by proposing the strategy. Additionally, AEGIS, the ESFRI clusters, e-infrastructures, and other research collaborations are part of the “who” because they influence the strategy’s direction through their feedback and ensure the utility of the AARC results by adopting them.

The AARC TREE team will present the AARC strategy to various stakeholders to seek feedback on the main actions. The compendium will address the Inputs received and will derive the necessary recommendations.

References

[AARC Projects]	https://aarc-project.eu/about/documents/
[AEGIS]	https://aarc-project.eu/about/aegis/
[ESFRI]	https://www.esfri.eu/
[FIM4R]	https://fim4r.org/
[REFEDS]	https://refeds.org/