

Clouds for Crowds—Implementing Federated AAI for the Digital Humanities

Tibor Kálmán¹ , Sally Chambers² , and Peter Gietz³ 

¹ GWDG, Göttingen, Germany

² DARIAH, Paris, France

³ DAASI International, Tübingen, Germany

Abstract

Technology has been inherent to the digital humanities from the outset. However, with the increase in data-driven research, Research Infrastructures (RI) such as DARIAH (Digital Research Infrastructure for the Arts and Humanities) need to ensure secure access to the data, tools and workflows they offer. Furthermore, collaboration with other RIs through initiatives such as OSCARS (Open Science Clusters's Action for Research and Society) and compatibility with the European Open Science Cloud (EOSC) are becoming increasingly important. It is within this context that this article aims to highlight the necessity and advantages of implementing federated identity management and authorisation. It describes the technological background of such an Authentication and Authorisation Infrastructure (AAI) solution for the digital humanities. Finally, it motivates the DH Tech community to adopt the AARC Blueprint Architecture supported by a Compendium being developed in the context of the AARC-TREE project.

Keywords: DARIAH AAI, AARC TREE, Authentication and Authorisation Infrastructure, AAI, AARC Blueprint Architecture

1 Research Infrastructure for the Arts and Humanities

DARIAH (Digital Research Infrastructure for the Arts and Humanities) [6] is a European research infrastructure that provides a platform for researchers in the arts and humanities to access, share, and preserve digital resources and tools [4]. DARIAH aims to enhance and support digitally-enabled research in the arts and humanities, fostering collaboration, innovation and knowledge-sharing across disciplines and borders [5]. The infrastructure offers a range of services, including repositories, virtual research environments, and authentication and authorisation mechanisms, to support the research lifecycle and facilitate the creation, publication, and preservation of digital scholarship. By providing a shared infrastructure and community-driven approach, DARIAH enables researchers to work together more effectively, leveraging digital methods and tools to advance our understanding of human culture and society. The technical foundation for this shared infrastructure is provided by the DARIAH AAI (Authentication and Authorisation Infrastructure).

This paper provides an overview of the DARIAH AAI from the authors' perspective. One author is one of the developers of the DARIAH AAI, one is responsible for its current operation, and one holds a management position within DARIAH. This paper provides insights into the technical aspects of the DARIAH infrastructure with the goal to encourage others to build on top of DARIAH AAI by following AAI best practices, and to showcase a sustainable integration pathway for DH services and DH communities into the global research infrastructure landscape.

Tibor Kálmán, Sally Chambers, and Peter Gietz. "Clouds for Crowds—Implementing Federated AAI for the Digital Humanities." In: *Digital Humanities Tech Symposium 2025*, ed. by Julia Damerow and Rebecca Sutton Koeser. Vol. 2. Anthology of Computers and the Humanities. 2025, 32–37. <https://doi.org/10.63744/a9oVejGuN3UN>.

© 2025 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

2 AAI for DARIAH

The DARIAH AAI [7] is a federated identity management system designed to facilitate secure and seamless access to data, tools and workflows for researchers in the humanities and arts. By leveraging federated authentication and authorisation, service providers can benefit from improved security, reduced administrative burdens, and enhanced user experience. Federated authentication allows users to access multiple services with a single set of credentials (e.g. their home institution login), eliminating the need for multiple usernames and passwords. This approach also enables service providers to outsource identity management to trusted identity providers, reducing the risk of password breaches and minimising the administrative effort required to manage user accounts. Furthermore, federated authorisation enables fine-grained access control, allowing developers to define and enforce precise access policies based on user attributes, group memberships and roles. From the user perspective, the DARIAH AAI also allows for single sign-on (SSO), i.e. the user authenticates once after connecting to a first application and is then automatically authenticated for all other federated applications they want to use the same day. By adopting the DARIAH AAI, service providers and developers can focus on delivering high-quality services to their users, while relying on a robust and scalable infrastructure to manage authentication and authorisation.

3 Technological Background

The technological background of DARIAH AAI is provided using established technologies and standardised protocols for federated identity management, as summarised below. More background information can be found in Chapter 3.3.3, *Access control to research data in the frame of FAIR principles and open access* [13].

SAML (Security Assertion Markup Language). The DARIAH AAI uses SAML as the primary protocol for exchanging authentication and authorisation information between identity providers and service providers. SAML allows for the secure exchange of user attributes and authentication statements between parties.

OpenID Connect (OIDC). In addition to SAML, the DARIAH AAI also supports OIDC, an authentication protocol built on top of the OAuth 2.0 framework. OIDC provides a simpler and more lightweight alternative to SAML for authentication and authorisation.

OAuth 2.0. The DARIAH AAI also supports OAuth 2.0 as the authorisation framework for managing access to protected resources. OAuth 2.0 provides a standardised mechanism for clients to request access to resources on behalf of the user.

Shibboleth and SimpleSAMLphp. The DARIAH AAI is based on the Shibboleth and simple-SAMLphp open source software systems, which provide a robust and scalable framework for managing identity federation. Both products support SAML and OIDC, and can provide a single sign-on (SSO) experience across both protocols.

Role-based Access Control (RBAC). The DARIAH AAI implements RBAC as a policy-based approach to access control. It allows for fine-grained access control decisions based on user roles, affiliations and permissions.

LDAP (Lightweight Directory Access Protocol). The DARIAH AAI uses LDAP as a directory service for storing and managing user attributes and identity information, as well as group memberships and roles.

SP/IdP Proxy (Service Provider / Identity Provider Proxy). DARIAH's SP/IdP proxy simplifies the connection of services and identity sources to the infrastructure. This mechanism enables any Dariah related attributes (e.g. member of a specific research project), to be provided, in addition to the attributes provided by the home organisation.

The DARIAH AAI is designed to support federation, which means that it allows multiple identity providers and service providers to participate in the infrastructure. This enables users to

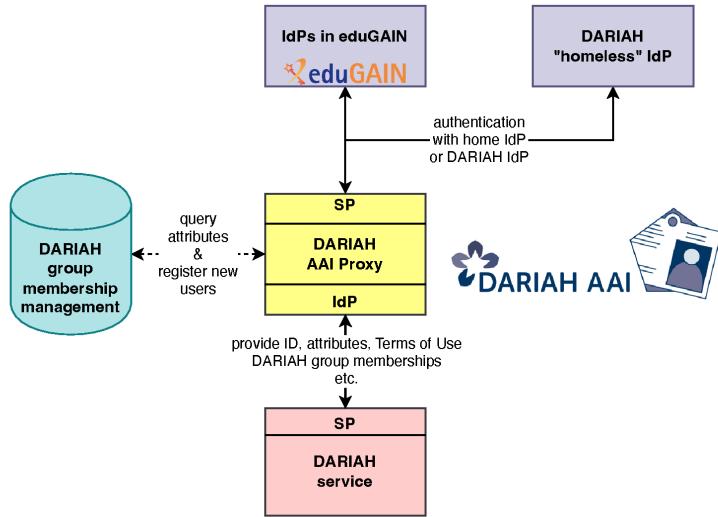


Figure 1: The Architecture of the Dariah Authentication and Authorisation Infrastructure.

access multiple services with a single set of credentials, while also allowing service providers to outsource identity management to trusted identity providers. It interconnects, for example, with eduGAIN [10], the interfederation of national research federations world-wide. Thus a researcher from a country that is part of the interfederation can access federated services world-wide by using the credentials managed by their home organisation.

Figure 1 depicts the Dariah AAI architecture. The Dariah SP/IdP Proxy allows authentication for (1) users from every IdPs interfederated in eduGAIN, as well as for (2) so-called “homeless” users from the Dariah “homeless” IdP, where any humanities researcher can register for an account including independent researchers without an institutional affiliation. For users of both categories (“eduGAIN” or “homeless”), Dariah-specific rights and privileges can be managed with the Dariah group membership management. This enables the proxy to transfer user attributes from the IdP where the authentication took place, as well as membership attributes from the Dariah virtual organisation to the Dariah services.

By leveraging the above mentioned standard technologies and protocols, the Dariah AAI provides a robust, scalable and secure system for managing identity and access control in the context of the research landscape. The majority of the design decisions made for the Dariah AAI are compliant with the globally recognised AARC Blueprint Architecture.

4 AARC Blueprint Architecture

The AARC (Authentication and Authorisation for Research and Collaboration) Blueprint Architecture [2] is a comprehensive framework for designing and implementing federated authentication and authorisation infrastructures for research communities. Developed through three phases of AARC projects [1], the Blueprint provides a set of guidelines, policies and architectures for building scalable and secure identity management systems. Dariah AAI implements the main components of AARC, including: **(1) Policy Framework**, which defines the rules and guidelines for identity management and access control; **(2) Proxy-based Architecture**, which outlines the technical design and components of the infrastructure; **(3) Standards and Protocols**, which specify the interoperability standards for identity federation and attribute-based access control; and **(4) Governance**, which establishes the organisational and operational structures for managing the infrastructure. By following the AARC Blueprint, research organisations and communities can es-

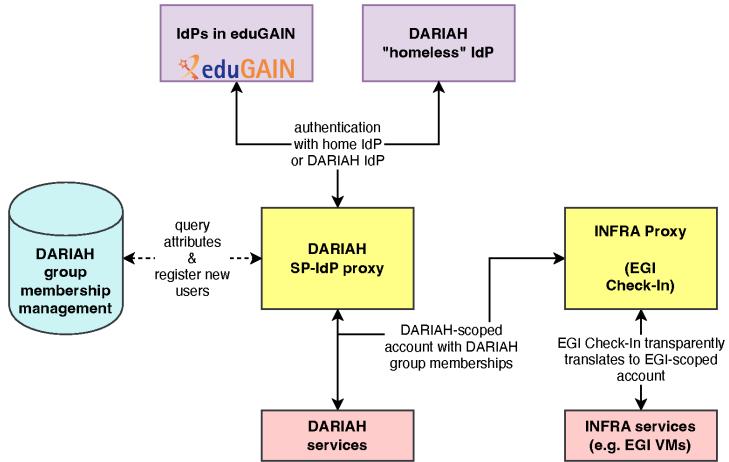


Figure 2: Interoperation between Dariah-AAI and Services of a Research Infrastructure.

Establish robust and trustworthy identity management systems (like Dariah AAI), to enable secure and seamless collaboration and access to shared resources and services.

The current AARC TREE project (AARC Technical Revision to Enhance Effectiveness) [3] drives the next phase of integration for research infrastructures. It expands federated access management to integrate user-centred technologies and access to federated data and services (authorisation), as well as defining common strategies for the development, deployment and sustainability of AAIs. A key aim of the AARC TREE project is to foster the uptake of the AARC Blueprint Architecture and Guidelines by research infrastructures and their communities such as Dariah. This work is supported by the development of a Compendium of AARC best practices and recommendations for a common long-term strategy for AAI services in pan-European Research Infrastructures in Europe. Furthermore, this work will lay the foundations for ensuring that the AAIs are compatible with the European Open Science Cloud.

5 EOSC and pan-European e-Infrastructures

The European Open Science Cloud (EOSC) [9] is a virtual environment that brings together research communities, data repositories, and service providers to enable seamless access to research data, services, and resources. The EOSC Federation [8] enables the distributed management of such data, knowledge and resources for European research, fostering collaboration, innovation, and the uptake of open science.

At the heart of the EOSC Federation are the EOSC Nodes, which are national or thematic hubs that provide access to research services, data, and resources. EOSC Nodes are responsible for aggregating and providing services, data, and resources from their respective domains, making them available to the broader research community through the EOSC Federation. Each Node acts as a gateway to the EOSC, providing a single point of access to a wide range of research services, including data repositories, computing resources, and software tools. By connecting to the EOSC Federation through the Nodes, researchers can access a vast array of resources and services, facilitating collaboration, data sharing, and innovation across disciplines and borders. The EOSC Nodes play a crucial role in realising the vision of the EOSC, enabling the creation of a unified, open, and accessible research environment for European scientists. The basis of the EOSC Federation is the EOSC AAI as described in [12], which interoperates with the Dariah AAI.

Dariah AAI's interoperability with pan-European e-infrastructure has already been suc-

cessfully validated by various proof of concepts implementations (PoC). For instance, [11] showcases how Dariah-AAI users could access services from other e-infrastructures, by using the Dariah AAI proxy as presented in the Figure 2. The Figure also shows how the Dariah-AAI proxy provides a powerful layer of identity harmonisation (account linking capabilities) by interacting with other infrastructure proxies (like EGI’s Check-In service or the EOSC EU Node’s MyAccessID service).

With the European Open Science Cloud entering its build-up phase, such federated access is crucial for enabling humanities researchers to connect to the wealth of data, services and resources available. Based on these undertakings it is now possible to productively login to the EOSC EU Node, via the Dariah AAI.

6 Conclusion

Robust federated identity management and authorisation for research infrastructures such as Dariah is becoming increasingly essential. By ensuring that the Dariah-AAI is compatible with the AARC Blueprint Architecture and related initiatives such as EOSC, its adoption not only helps to simplify the management of access control and authorisation for DH service providers, but also reduces administrative burdens and most importantly, improves the overall user experience for researchers in the humanities and arts.

Acknowledgments

This work was partly funded by the European Union (AARC TREE project, grant agreement no. 101131237). We would also like to acknowledge the support and contributions of the AARC community, and we are grateful for the feedback provided by the Dariah community.

References

- [1] AARC Community. “AARC webpage”. Accessed: Aug. 2025. URL: <https://aarc-community.org>.
- [2] AARC Community members and AppInt members. “AARC Blueprint Architecture 2019 (AARC-G045)”. Nov. 2019. DOI: 10.5281/zenodo.3672785.
- [3] AARC Technical Revision to Enhance Effectiveness. “AARC TREE webpage”. Accessed: Jul. 2025. URL: <https://wiki.geant.org/display/AARC/AARC+TREE+Project>.
- [4] Barbot, Laure, Dolinar, Maja, Gray, {Edward J.}, Grisot, Cristina, Illmayer, Klaus, Kurzmeier, Michael, and McGillivray, Barbara. “Contextualizing Research Tools & Services Through Workflows in the SSH Open Marketplace”. English. In: *Journal of Open Humanities Data* 10 (Feb. 2024). ISSN: 2059-481X. DOI: 10.5334/johd.192.
- [5] Chambers, Sally et al. “OSCARS D2.1 - Clusters’ Services and Data Sources Portfolios”. Feb. 2025. DOI: 10.5281/zenodo.14881627.
- [6] Dariah (Digital Research Infrastructure for the Arts and Humanities. “Dariah webpage”. Accessed: Jul. 2025. URL: <https://www.dariah.eu/>.
- [7] Dariah (Digital Research Infrastructure for the Arts and Humanities). “Documentation for Dariah-AAI”. Accessed: Jul. 2025. URL: <https://doc.de.dariah.eu/DARIAH-AAI-Documentation/>.
- [8] EOSC Association. *EOSC Federation Handbook*. Zenodo, Mar. 2025. DOI: 10.5281/zenodo.14999577.
- [9] European Open Science Cloud. “EOSC webpage”. URL: <https://www.eosc.eu/>.

- [10] GÉANT Association. “EduGAIN webpage”. Accessed: Aug. 2025. URL: <https://edugain.org>.
- [11] Hübner, D. and Igoumenos, I. “Showcase of the DARIAH-EGI pilot”. FIM4R meeting, Vienna. Feb. 2019. DOI: 10.5281/zenodo.3672785.
- [12] Kanellopoulos, Christos et al. “EOSC AAI Architecture 2025”. Version March 2025. May 2025. DOI: 10.5281/zenodo.15388270. URL: <https://doi.org/10.5281/zenodo.15388270>.
- [13] Tóth-Czifra, Erzsébet et al. “Research Data Management for Arts and Humanities: Integrating Voices of the Community”. Chapter 3.3.3, Access control to research data in the frame of FAIR principles and open access, p. 67-75. Aug. 2023. DOI: 10.5281/zenodo.8059626.