EOSC-Life

EOSC-Life: Building a digital space for the life sciences

D5.3 – Access and User Management System for Life Science – the blueprint update

WP5 – User management and access services
Lead Beneficiary: MU and Instruct
WP leader: Ludek Matyska (MU) and Susan Daenke (Instruct)
Contributing partner(s): BBMRI, INFRAFRONTIER, Instruct, ABO, USMI, MU, CSC

Authors of this deliverable: **Mikael Linden** (CSC), **Dominik Frantisek Bucik** (MU), **Philipp Gormanns** (INFRAFRONTIER), **Natalie Haley** (Instruct), **Jani Heikkinen** (CSC), **Petr Holub** (BBMRI), **Pasi Kankaanpää** (ABO), **Daniel Kouril** (MU), **Martin Kuba** (MU), **Slavek Licehammer** (MU), **Ludek Matyska** (MU), **Tommi Nyrönen** (CSC), **Michal Prochazka** (MU), **Robert Reihs** (BBMRI), **Paolo Romano** (USMI), **Fiona Sanderson** (Instruct), **Athresh Shigaval** (ABO), **Callum Smith** (Instruct), **Jonathan Tedds** (ELIXIR)

Contractual delivery date: **31 March 2021**
Actual delivery date: **24 March 2021**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Table of Contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Executive Summary

This deliverable is the updated blueprint of the Life Science AAI, a common authentication and authorisation service for the European life science research infrastructures. The Life Science AAI provides a way to coordinate the way how user identity and access is managed in research services and data in the (community-specific) federated Infrastructure. The service is managed by the life sciences community and operated by the e-infrastructures, including GEANT and EGI.

This deliverable is an update to the deliverable D5.1 "Access and User Management System for Life Science – the blueprint" [LIND19] that laid the ground for the production deployment with e-infrastructures. In this deliverable and its appendices, only those sections of D5.1 are updated where the requirements have changed or become more clear, while the rest of the document remains unchanged. This deliverable can be read as an independent document, without reading D5.1 first.

This deliverable and its appendices first introduces the technical and non-technical requirements for the AAI service, including the components operated by the e-infrastructures, and further clarifies how that relates to and integrates to the AAI service components operated by the life sciences research infrastructures. Relevant external work, including the EOSC AAI, is then introduced. A migration plan for the existing AAI services, including BBMRI-ERIC AAI, ELIXIR AAI and ARIA, is presented. Finally, the deliverable provides an overview of Life Science AAI's potential relying services within the EOSC-Life project and beyond.

Project Objectives and Introduction

With this deliverable, the project has reached/this deliverable has contributed to the following objectives:

a. To specify/define a convergent access and user management system to enable multi-RI applications and workflows that build on existing approaches (LS AAI, ARIA, Negotiator service, etc) and support access to sensitive data with their specific requirements

Detailed Report on the Deliverable

1. Background

The need to authenticate researchers and manage their access rights is common to many research infrastructures. Research infrastructures need to protect access to confidential (such as samples from human patients or information about ongoing or proposed research projects) or expensive resources (such as sophisticated instruments or computing capacity). This requires sufficient information: who the users are (identity proofing and user authentication), whom are

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

they representing (affiliation with a home organisation) and what resources they can access (authorisation). These services are called an authentication and authorisation infrastructure (AAI). For more information on the AAI terms, concepts and paradigms, refer to section 1 of CORBEL D6.5 [CORB19b].

Providing services for researcher authentication and authorisation fits well the research infrastructures' mission to support researchers' work. Research infrastructures are permanent entities facilitating research projects in collaboration with the research communities. They are well connected in their domain and able to understand the common needs of their user communities. On the other hand, AAI is not a core business for research infrastructures; encouraging collaboration in the research and education sector with other actors (such as e-infrastructures) who have a long history of developing the underlying AAI technologies and services.

The work towards a Life Science AAI started in May 2016 with a common meeting of the e-infrastructure-coordinated AARC project and the BMS RIs participating in CORBEL project WP5. As an outcome, the RIs started to collate the use case requirements from participating RIs and compiled them to a requirements specification for a pilot[1] that took place in the context and on the funding provided by the AARC2 project. The pilot infrastructure was operated by e-infrastructures (EGI, EUDAT and GEANT) whereas the CORBEL WP5 community remained as the stakeholder of the pilot.

During the pilot, CORBEL WP5 updated the Life Science AAI requirements based on the pilot experiences and supplemented them with non-technical requirements. The technical and non-technical requirements presented in deliverable D5.1[LIND19] have then served as key contributions for the deployment and operation of the Life Science AAI, as described in deliverable D5.2 [LICE21]. The updated technical requirements are presented in Appendix A. The non-technical requirements are presented as Appendix B of D5.1 and are not updated by this deliverable.

Some BMS RIs have been operating their RI-wide AAI services for several years, including BBMRI-ERIC AAI (709 users and 13 production services by the end of February 2021) and ELIXIR AAI (5863 active users and 102 production services at the end of 2020). While those AAIs have been important for understanding the BMS community's needs on AAI, the eventual goal of LS AAI is to enable their migration to the Life Science AAI during the EOSC-Life project. It is important to ensure these services' forward compatibility with and smooth migration to the LS AAI. However, some RI specific components may still remain under the control of the RIs, as will be described in this deliverable.

2. LS AAI service ecosystem

This section introduces the Life Science AAI ecosystem, which consists of

[1] https://aarc-project.eu/aarc-in-action/corbel/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- the AAI services operated by e-infrastructures. This composes the core LS AAI services, which require operational excellence for the AAI components, underlying (normally open source) products and the availability of the environment;
- the RI services contributing to the Life Science AAI. These services build on top of the core Life Science AAI and require the expertise of the life science community, and its specific needs and standards;
- external AAI services, including the EOSC AAI.

These entities are shortly introduced in this section.

2.1. Services operated by the e-infrastructures

Requirements on the services subsumed from e-infrastructures consist of technical requirements, (section 2.1.1. and Appendix A) and non-technical requirements (Appendix B of D5.1) which are both shortly highlighted here with references to more information in the appendices.

2.1.1. Technical requirements

The technical requirements for the Life Science AAI articulate functionality the BMS community is expecting from the LS AAI operated by the e-infrastructures. The technical requirements do not impose any particular technical architecture and the e-infrastructures have met the requirements with a technical architecture that is compatible with the AARC blueprint architecture [AARC19] and has been presented in deliverable D5.2 [LICE21]. This deliverable updates the original requirements presented in D5.1 [LIND19].

The Life Science AAI issues a new **identifier** called a Life Science ID (Appendix A, section 2.1) to a user who registers to the Life Science AAI (section 3.1) and accepts its usage policy (section 3.4). Users authenticate using authentication providers, such as their home universities (section 3.2) or the Life Science AAI's Hostel IdP (section 3.3) which can be linked to their Life Science ID (section 3.5). To cater services with specific assurance needs, Life Science AAI supports an assurance framework (section 3.7) and has a step-up authentication service (section 3.8).

Life Science AAI decorates the user IDs with **extra attributes**, such as the user's home organisation (section 4.1) and researcher status (section 4.3). The Life Science AAI can also manage user's group memberships (section 4.4) and permissions to access controlled access datasets (section 4.5). These attributes can then be consumed and used for access control enforcement by the services relying on the LS AAI.

The Life Science AAI exposes two main **interfaces for the relying services** (section 6.1): SAML 2.0 (Security Assertion Markup Language) and OpenID Connect (OIDC)/OAuth2 which is better suited for API/CLI access. The deliverable also adds support to GA4GH Passports, which is technically a thin profile layer on top of the standard OIDC. After authenticating the user using an authentication provider, the Life Science AAI returns user attributes to the relying services. However, some relying services may prefer to use X.509 certificates (section 6.3) for user authentication or receive (or update) user attributes using a back-end synchronisation (section 6.5), which are supported, too. Life Science AAI also supports provisioning and deprovisioning of user accounts and authorisation data to services. Services can react to changes in user accounts even if the user is not directly interacting with them.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

2.1.2. Non-technical requirements

The non-technical requirements on the Life Science AAI were presented in Appendix B of D5.1 and are not updated in this deliverable. They do not specify new functionality for the Life Science AAI but further clarify what conditions the service operated by the e-infrastructures must meet. In particular, it introduces

- split of the service components to three service categories with different annual **availability** requirements: red (99.9%), yellow (99.5%) and green (98%) categories. This indicates where the e-infrastructures are expected to optimise the architecture for high availability;
- how the **monitoring** of the Life Science AAI, its availability, in particular, should be organised and reported;
- how the **governance** of the Life Science AAI is expected, including how the interaction between the BMS and e-infrastructure communities is channelled through representative coordinating bodies;
- how the requirements imposed by the **GDPR** are covered, including defining the e-infrastructures as a data processor for the life science community who is the data controller;
- the 600,000 euro **budget** reserved by the EOSC-Life project for e-infrastructures for the operations of the Life Science AAI. The funding model after EOSC-Life project remains to be agreed on during the project;
- considerations on the **future evolution** of the e-infrastructure service offering.

Deliverable D5.2 [LICE21] presented how the GDPR requirements have been solved for the run of the project. The post-project governance and funding of Life Science AAI remains to be solved by the end of the project.

2.2. RI services contributing to LS AAI

The e-infrastructures are experts on operations of the core AAI services, but the BMS research infrastructures are better suited for managing the service components that require wider substance matter understanding and contacts with the user communities. This section describes the service components operated by the BMS research infrastructures. Those components can, for instance, contribute an attribute to the users' Life Science ID.

2.2.1. ARIA

ARIA[2] is a collection of web services designed, built, and provided by Instruct[3] to research infrastructures, facilities, and user communities. As a cloud service, ARIA has the opportunity to centralise access offerings from multiple biomedical science research domains to provide cross-disciplinary scientific proposals and truly integrative science. ARIA has seen expansive adoption and is now in use by a number of different RIs, networks and facilities for managing access. ARIA was also used for managing the CORBEL Open calls and is the proposed platform to host the EOSC-Life open calls.

[2] http://aria.structuralbiology.eu/
[3] https://instruct-eric.org/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

The ARIA cloud service has had federated identity management in production for 5 years, utilising a similar model to that proposed here where the service can link together accounts from multiple sources to a single identity. ARIA IdP leverages this position of linked identities to be able to provide consistent identifiers of a single user enhanced with attributes from the central ARIA identity service. Users can self-register within ARIA with an automated email verification followed up by Instruct-ERIC manual verification of the user's position within their defined institute. Additionally, the user specifies projects and infrastructures with which they are associated. Finally, users can be given management permissions of access routes (RIs such as EU-OPENSCREEN), facilities (e.g. Diamond Light Source, NeCEN), machines and technologies (e.g. scientific instruments like electron microscopes) which dynamically populate their group memberships. Through centralised exposure of these dynamic group memberships access to different tools and services are managed without the overhead of complex group management.

Adoption of Life Science AAI will be critical towards its success and ARIA represents one of the largest active user communities that will be directly consuming the new AAI service. The migration will be phased to ensure no interruption of the process is caused to ARIA users and to allow for growth of trust to the new identity platform. The group management of Life Science AAI will be tightly integrated to the ARIA group management to allow for seamless transfer of users between the two systems and to ensure that no additional management overhead is incurred through the transition. Where possible alignment of roles and group memberships will be made to ensure interoperability of ARIA membership roles and where feasible Life Science AAI roles will be enabled. Project participation is a key area of ARIA that ensures GDPR and regulatory compliance for a homogenous user-base which would be a strong area to focus on cross-RI harmonisation.

Instruct-ERIC has a lot of ties with physical infrastructure providers for researchers such as large synchrotrons (Diamond Light Source) which need to provide terminal desktop logon for a huge number of access visitors. It is hoped that Life Science AAI will be expanded to adopt non-web desktop authentication technologies to enable streamlined access for users and give a common user identifier throughout the access process. Life Science AAI has the opportunity to enable large gains in data sharing and interoperability as a huge benefit to consistent identity throughout the process of access.

2.2.2. BBMRI-ERIC Negotiator

BBMRI-ERIC Negotiator[4] is a tool that implements policy and procedures to access data and/or biological material from BBMRI-ERIC partner biobanks. BBMRI-ERIC Negotiator has currently 201 biobankers registered to represent 1317 biobank collections, and overall there are close to 646 registered users. The service covers all 20 member countries of BBMRI-ERIC RI and IARC, a subsidiary of the WHO international organization. This service is about negotiating access to health-related human data (particularly sensitive personal data following GDPR) and biological samples in a trustworthy manner. The Negotiator is used by the biobank representatives to negotiate and eventually decide on whether access to the particular data and sample sets can be granted [CORB19b]. As the service guards access to precious human biological material and highly

[4] https://negotiator.bbmri-eric.eu/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

regulated sensitive associated data, the service exhibits substantially lower numbers of accesses and registered users compared to the services dealing with open data.

User authentication is a critical part of the functionality of Negotiator, as the biobanks need to be able to trust the authenticity of the users and their requests. Furthermore some sensitive data may be accidentally communicated as a part of the negotiation, e.g., in case of rare diseases where data can be potentially revealing and can be used to correlate and/or infer additional information about research participants; therefore it is important to have identified users and that all users need to be bound to confidentiality.

BBMRI-ERIC has its own Authentication and Authorization Infrastructure[5] (BBMRI-ERIC AAI) that manages the user identities and authentication. It uses external identity providers coming from eduGAIN, as well as an internal identity provider called LifeScience Hostel[6], which is a catch-all provider for all the users whose home institutions are not participating in eduGAIN. BBMRI-ERIC's experience shows that approximately 60% of such medical researchers come via LifeScience Hostel; this is not only because of the above-mentioned lack of home institutions participating in eduGAIN, but also because it is practically impossible to use information from some participating institutions if they fail to release even the basic attributes (typical for most of the Dutch institutions), or due to other fundamental problems (e.g., a changing user identifier on every authentication instance as observed for some German institutions).

BBMRI-ERIC AAI maintains information about the attributes released by users' home organisations as well, i.e. the status of users' institutional affiliations, and level of assurance of users' identity, if available. Furthermore, it maintains additional attributes internally which are specific for the use within the BBMRI-ERIC infrastructure: a typical example is a complex hierarchy of groups defining entitlements of users with respect to biobanks and their collections and also biobank networks. These groups are then propagated into the relevant services such as the Negotiator, either via push or pull mechanism (push is usually preferred to avoid online dependencies among the system, thus minimizing the risk of downtime caused by several online interdependent services).

Before being able to use the Negotiator, all users are required to agree to the Terms and Conditions (T&Cs) for using BBMRI-ERIC Services to make sure the user agrees with the above-stated confidentiality and privacy protection principles. Acceptance of these T&Cs is stored as another attribute of the user, also allowing for versioning: if a new version of the T&Cs becomes available, the users are requested to approve those before proceeding to the target service.

BBMRI-ERIC AAI allows two types of authentication clients: SAML and OpenID Connect; practical experience from the BBMRI-ERIC ecosystem shows that programmers largely prefer the OpenID Connect interface.

Negotiator interfaces, inter alia, the BBMRI-ERIC Directory[7], which is the main service ensuring findability of European biobanking resources. The Directory has been recently integrated with the BBMRI-ERIC AAI, too, in order to allow users to manage and edit information about the biobanks and their collections of biological material and data, for which they are responsible (either

[5] https://perun.bbmri-eric.eu/

[6] LifeScience Hostel is already using the LifeScience name to facilitate the transition to the LifeScience AAI and to minimize disturbance to the users.

[7] https://directory.bbmri-eric.eu/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

biobank/collection representatives, or representatives of BBMRI-ERIC National Nodes, i.e., member states). Integration of other BBMRI-ERIC services with the AAI is underway, such as BBMRI-ERIC Helpdesk[8] and BBMRI.uk National Directory[9].

Integration of the BBMRI-ERIC Negotiator with the pilot LifeScience AAI has been tested as a part of the AARC2 pilot in 2018. This integration used a developer's testing version of the Negotiator and turned out to be successful.

The main advantage of the transition to the full-scale LifeScience AAI would be in merging the user bases across different Research Infrastructures in the life sciences domain. A typical example where practical benefits materialize is the European Joint Programming on Rare Diseases, where ELIXIR and BBMRI-ERIC (among others) collaborate to develop a Virtual Platform. The larger user base should also allow for faster on-boarding of home organizations of users if it requires manual approval of institutional identity provider managers.

2.2.3. Resource Entitlement Management System (REMS)

Resource Entitlement Management System (REMS)[10] [LIND13] is an open source product implemented by ELIXIR-Finland to manage access to resources, in particular to controlled access data. A researcher who has identified the datasets of interest fills a data access application, attaches a research plan and the list of research group members and submits the application. REMS then circulates the application to the data owners (commonly represented by Data Access Committees, DACs) for review and approval. REMS delivers the data access rights to the system component enforcing access, such as a dataset download site or a secure cloud where the data is made available for the researchers. REMS also provides necessary reporting for the audit trail.

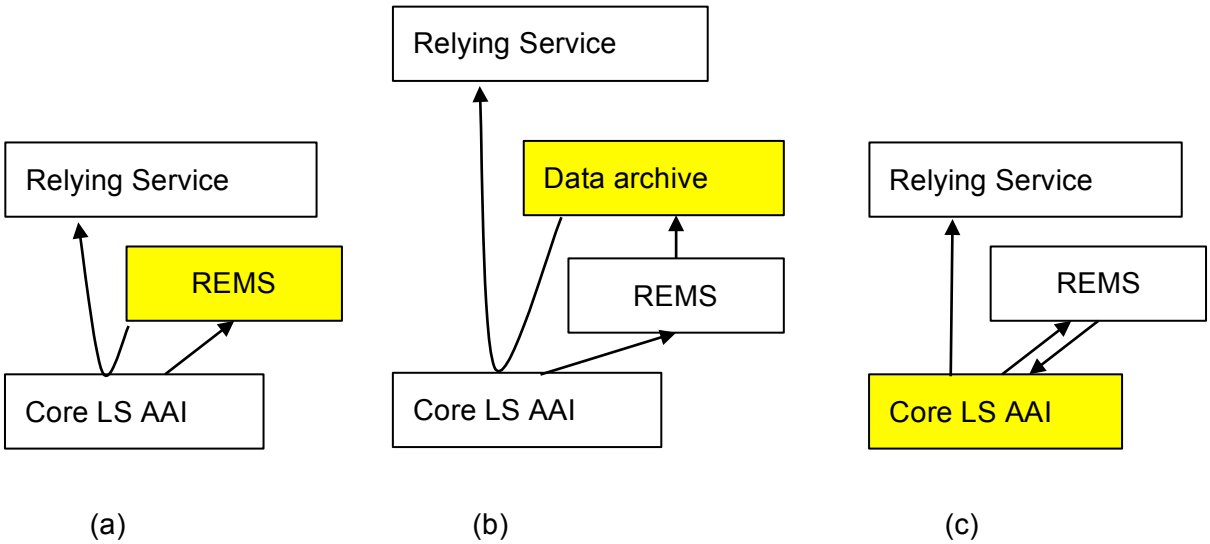

Figure 1: Deployment scenarios for REMS in the Life Science AAI. The yellow components indicates the authoritative source of permissions.

[8] https://helpdesk.bbmri-eric.eu/
[9] https://directory.biobankinguk.org/
[10] https://github.com/CSCfi/rems

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Figure 1 presents three possible, non-exclusive ways to integrate REMS to the Life Science AAI. In all alternatives, REMS first uses the Life Science AAI to receive a researchers' authenticated Life Science ID and, based on the approval process, attaches a list of permitted datasets to it. In scenario (a) REMS is the authoritative source of permissions (indicated in yellow). At the time a user logs in, Life Science AAI fetches their fresh and accurate permissions from REMS and exposes them to the relying service (potentially using standards, such as OIDC and GA4GH Passports). This approach has been demonstrated in the Nordic Tryggve project[11]. In scenario (b) REMS is seen as a sub-component of the data archive that holds the datasets. REMS feeds information on approved data access applications to the data archive which is always the authoritative source of fresh permissions towards the relying services. This approach has been demonstrated with the EGA data archive in the ELIXIR EXCELERATE project [LIND18]. In scenario (c) REMS is seen as a sub-component of the Core Life Science AAI service to which it feeds information on approved data access applications. This approach is deployed in ELIXIR AAI for a specific attribute where REMS is used by peer researchers to vouch for a user's status as a bona fide researcher.

2.3. Integration to external AAIs

In the research and education sector, there are other AAI services to which the Life Science AAI will be integrated. Some of them are operated by other research infrastructures (such as Umbrella[12] AAI operated by the photon/neutron community), some are operated by generic e-infrastructures (such as eduTEAMs of GEANT, Check-in of EGI and B2Access of EUDAT).

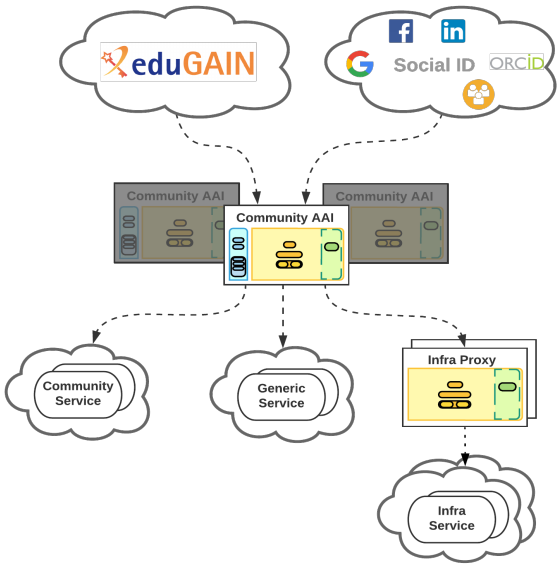

Figure 2: AARC2 Blueprint architecture [AARC19, p.9] introduces an approach where Community AAIs (such as Life Science AAI) manage users' identities and roles, group memberships and permissions and e-infrastructure AAIs act as Service Provider proxies for e-infrastructure services.

[11] https://neic.no/tryggve
[12] http://pan-data.eu/Umbrella

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

To facilitate the integration of the AAI services, AARC2 project has introduced an AAI blueprint architecture [AARC19] (Figure 2) which separates community and e-infrastructure AAI services. As the research communities normally manage their community memberships and the services available for the community, they are in the best position to manage researcher identities, roles and other attributes, group memberships and permissions to resources. The e-infrastructures consume the information managed by the community AAI services in their Service Provider proxies to enforce access to the e-infrastructure services permitted for the communities.

In this approach, Life Science AAI is positioned as a community AAI service for the life science community. E-infrastructure services then integrate to Life Science AAI as a relying service, acting as a proxy for multiple e-infrastructure services. Life Science AAI may further integrate to other community AAIs (such as Umbrella) with whom it has a significant number of common users. The integration can be done, for instance, by the EOSC AAI Federation that is introduced in the next section.

2.4. EOSC AAI Federation

In January 2021, the European Commission published the EOSC Authentication and Authorization Infrastructure report [EC21] that was prepared by the AAI Task Force of the EOSC Architecture Working Group. The report establishes a global ecosystem for identity and access control infrastructures for the European Open Science Cloud (EOSC).

The EOSC AAI is an identity federation to which organisations providing services to the EOSC can join as a member. Membership to the EOSC AAI Federation is open to organisations that have a relationship with the EOSC and that abide by the EOSC Rules of Participation and the EOSC AAI organisational baseline [EC21].

The EOSC AAI builds on the AARC Blueprint architecture described above. Members can register Identity Providers (in particular, Community AAIs) and Service Providers (in particular, Service Provider proxies) to the EOSC AAI Federation provided they follow the defined technical and security baseline. Members can also ask their Providers to be imported from a peer federation (e.g. eduGAIN) if they are already registered to one. The providers must support SAML and/or OpenID Connect protocols, as described in the Technical framework of the EOSC AAI Federation [EC21].

When it becomes operational, the Life Sciences community can join the EOSC AAI Federation and register the Life Science AAI as an Identity Provider. This will enable Life Science AAI users to access the EOSC services they are permitted to access. Via the EOSC AAI Federation, also the relying services of Life Science AAI can be made available for the end users of other communities, when applicable.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

2.5. Relevant external work

This section introduces external factors that are relevant for the Life Science AAI. Those can, for instance, manage standards that the Life Science AAI needs to implement or take into account in its own design.

2.5.1. Federated identity management and identity federations

The federated identity management concept emerged in the early 2000s to cater access management scenarios where a user wants to use a single identity to access (initially web-based) services in different security domains on the Internet. Nowadays, the most widely used federated identity management protocols are SAML (Security Assertion Mark-Up Language) and OpenID Connect [CORB19b].

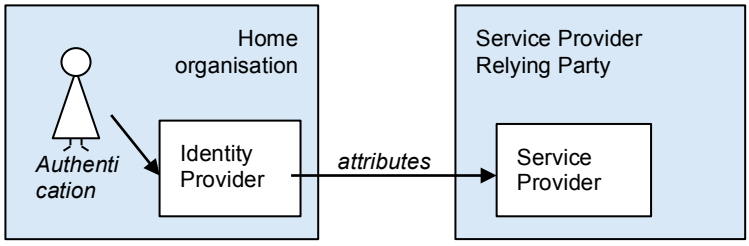

Figure 3: In federated identity management, a service (service provider) delegates user authentication to an identity provider server managed by the user's home organisation [CORB19b].

Figure 3 describes a scenario where the users' Home organisation issues a username and password (or other means for authentication) that can be used to access all relying services (sometimes called service providers, SP). When users need to log in to a service, the home organisation's identity provider server authenticates them and releases their attributes (such as a unique identifier and role description) to the relying service [CORB19b].

Federated identity management has many security benefits if the home organisation is the organisation the user is affiliated with (e.g. the university or research institution employing the researcher). The user's home organisation is usually in a position to perform reliable identity proofing for the user and make sure their attributes are fresh and accurate. When the user departs, the home organisation can close their account promptly which is sufficient to close their access to all relying services [CORB19b].

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

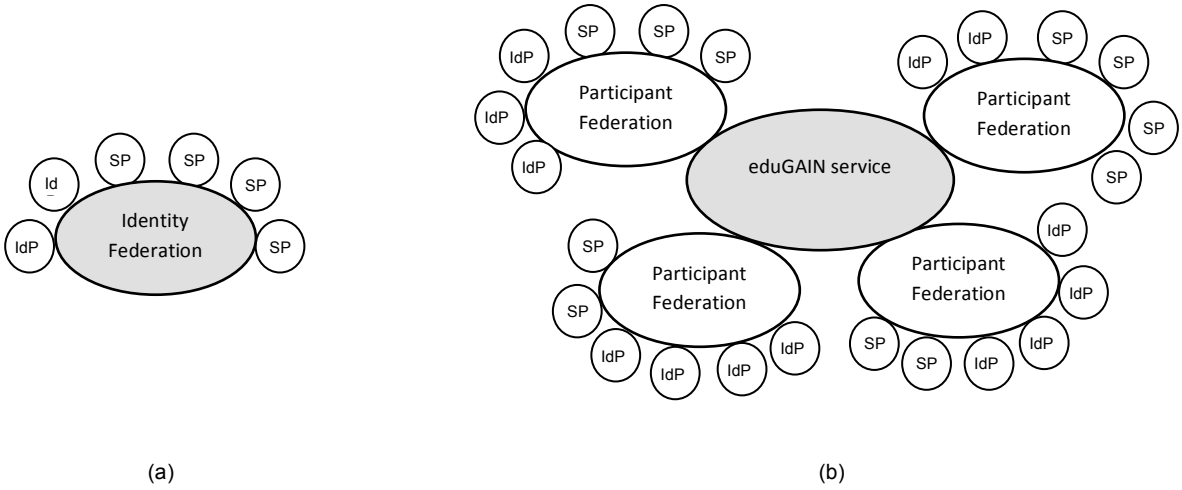

Figure 4: An identity federation (a) connects identity providers (IdP) and service providers (SP). eduGAIN interfederation service (b) further connects identity federations. [CORB19b]

There are thousands of research and education institutions globally. To organise the policies and practices under which federated identity management can be done, national research and education

networks have established identity federations (REFEDS13) which are associations of organisations that come together to exchange information as appropriate about their users and resources to enable collaborations and transactions. This is depicted in Figure 4(a). An identity provider registered to a federation can enable their users to log in to the service providers registered to the federation and service providers accept user logins from the federation's identity providers. GEANT e-infrastructure is further running a service called eduGAIN (Figure 4(b)) which enables interfederation for identity and service providers belonging to different federations [eduGAIN,CORB19b].

To benefit from the identity proofing and affiliation information managed by researcher's home organisation, the Life Science AAI relies on home organisation login via the eduGAIN interfederation service. For smooth integration to home organisation login, the Life Science AAI needs to subscribe to certain practices that the federations are using

- REFEDS Research and Scholarship[14] services, indicating that the service is operated for the purpose of supporting research and scholarship interaction.
- GEANT Data protection Code of Conduct[15], indicating that the service is following practices derived from the European data protection laws
- SIRTFI[16], indicating that the service follows community practices for information security, in particular those related to management of security incidents

[13] http://www.refeds.org/
[14] https://refeds.org/category/research-and-scholarship
[15] https://wiki.refeds.org/display/CODE
[16] https://refeds.org/sirtfi

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

2.5.2. Federated Identity Management for Research (FIM4R)

Federated identity management for research (FIM4R[17]) is a global cross-discipline network of identity and access management (IAM) professionals in research infrastructures and services. After its first meeting at CERN in June 2011, FIM4R has had 16 face-to-face meetings to compare and articulate different disciplines' needs on IAM, in particular on federated identity management [CORB19b].

FIM4R works on voluntary capacity collating participants' IAM requirements and presenting them to stakeholders for action. FIM4R's latest contribution is the Federated Identity Management for Research Collaborations version 2 white paper [FIM4R18] which presents 40 recommendations to the stakeholders, including the identity providers, identity federations and eduGAIN, research communities, research service providers, software developers and standards bodies [CORB19b].

2.5.3. Global Alliance for Genomics and Health (GA4GH)

Global Alliance for Genomics and Health[18] is a policy-framing and technical standards-setting organisation, seeking to enable responsible human genomic data sharing. It is an international organisation that is relevant for the BMS RIs working on human data, such as BBMRI and ELIXIR.

GA4GH is organised to work streams. The DURI (Data Use and Researcher ID) work stream has delivered a GA4GH Passports[19] specification describing the syntax and semantics for five claims[20] (dubbed as visas) for a researcher:

- AffiliationAndRole visa to express a user's role in their home organisation.
- AcceptedTermsAndPolicies visa to describe that this person or their organisation has acknowledged specific terms and conditions which contribute to their permission to access services. For instance, the researcher has made an attestation, as further specified by GA4GH, that they will refrain from trying to re-identify individuals from the samples they access.
- ResearcherStatus visa to indicate that this person has been acknowledged to be a researcher, for instance by their home organisation or a peer.
- ControlledAccessGrants visa to present a list of data objects this person has been granted access to. For instance, the researcher has requested access and presented their research plan to a competent data access committee who has approved the request.
- LinkedIdentities visa to describe other identifiers this researcher is known to use.

The Security Work stream has delivered a GA4GH AAI OpenID Connect Profile[21] specification which presents how the claims should be mounted on the OIDC protocol for delivery from an OIDC provider (dubbed as an OIDC broker) to a Relying service (dubbed as a Claims Clearinghouse). The Life Science AAI is a potential operator of such an OIDC broker.

[17] https://fim4r.org/about/
[18] https://www.ga4gh.org/
[19] https://github.com/ga4gh-duri/ga4gh-duri.github.io/blob/master/researcher_ids/ga4gh_passport_v1.md
[20] *A claim* is an OpenID Connect term for what is called *an attribute* in SAML (and this document).
[21] https://github.com/ga4gh/data-security/blob/master/AAI/AAIConnectProfile.md

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

3. Migration of existing research infrastructure AAIs

After the Life Science AAI is launched and has demonstrated sufficient stability and operational excellence, the existing BMS research infrastructure AAIs are invited to migrate their users and services to the Life Science AAI. The decision to migrate will be done by the sponsoring research infrastructures and requires that the Life Science AAI provides at least the same level of functionality.

This section describes the migration process of the BMS AAIs, focusing on BBMRI-ERIC AAI, ELIXIR AAI and ARIA, which are found to be the key AAI services that need to be migrated.

3.1. Migration of BBMRI-ERIC and ELIXIR AAI

BBMRI-ERIC AAI and ELIXIR AAI are similar enough to share the same migration approach. For the migration, they are integrated to the Life Science AAI using a hybrid architecture depicted in Figure 5. The migration consists of four phases which are described below.

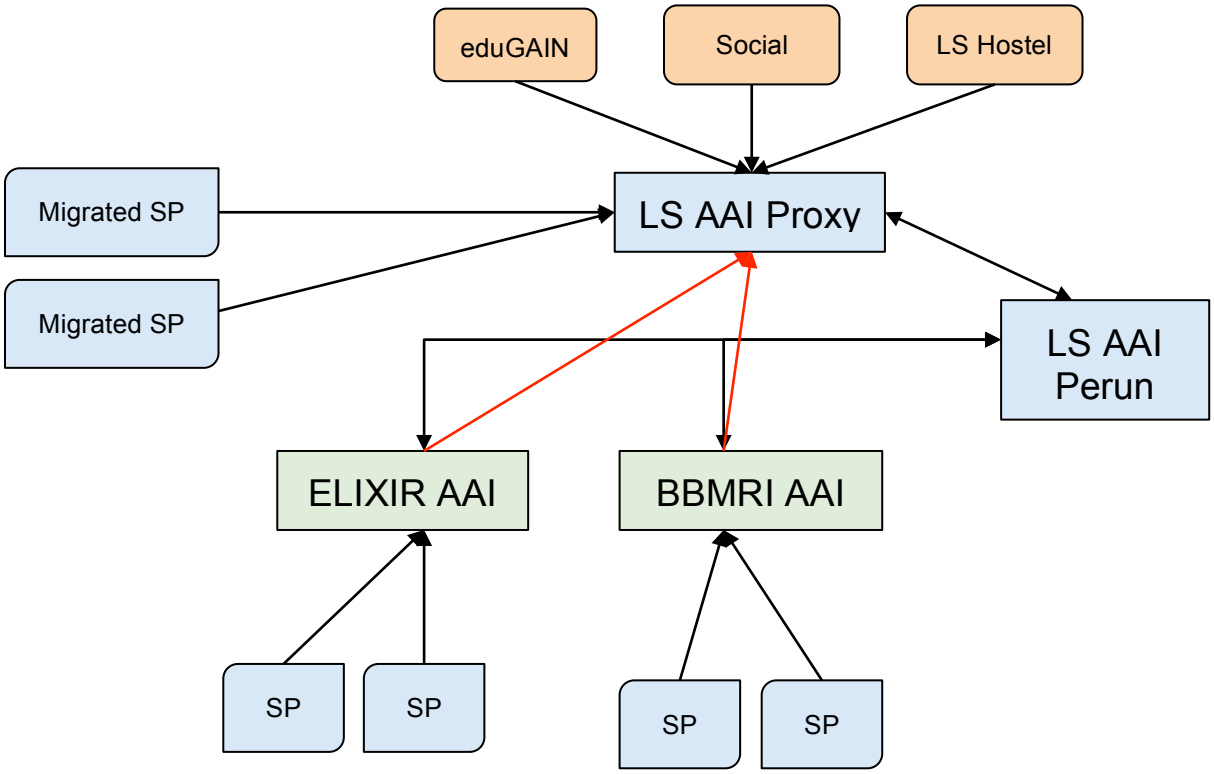

Figure 5: Migration of ELIXIR AAI and BBMRI-ERIC AAI to the Life Science AAI.

Preparation phase. The identity provider proxies of ELIXIR AAI and BBMRI-ERIC AAI are configured to use Life Science AAI as their identity provider proxy. The Life Science AAI becomes one of the hundreds of supported external authentication providers for ELIXIR AAI and BBMRI-ERIC AAI. If a user selects the Life Science AAI as their login method in ELIXIR AAI or BBMRI-ERIC AAI, the login flow will traverse the two proxies.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Users migration phase. All user records and groups from ELIXIR AAI and BBMRI-ERIC AAI are moved to the Life Science AAI. User identifiers are changed to the Life Science ID format and identifier collisions are resolved. ELIXIR AAI and BBMRI-ERIC AAI are switched to use LS AAI as their identity management back-end. ELIXIR and BBMRI-ERIC AAIs switch off their identity provider discovery service and start to transparently pass all users to the Life Science AAI for identity provider discovery.

Services migration phase. Services are asked to migrate from ELIXIR AAI and BBMRI-ERIC AAI to the Life Science AAI. This requires that the services are reconfigured to use the technical endpoints of the Life Science AAI identity provider proxy and that the services transform the user identifiers in their existing user databases to the Life Science IDs. This phase can be relatively long and the services can be assisted in the migration, one-by-one.

Finalisation phase. Finally, the ELIXIR AAI and BBMRI-ERIC AAI proxies can be removed from the Life Science AAI and decommissioned.

During the process, attention needs to be paid to

- informing the users and providing them an opportunity to opt-out. Although there is no material change in the purpose of processing their personal data, the data controller will change at least for the BBMRI-ERIC AAI. The users need to be informed beforehand about the migration to the Life Science AAI and they need to be provided an opportunity to opt out.
- the users committing to the Acceptable Usage Policy of the Life Science AAI the first time their login goes through the Life Science AAI.
- ensuring that the Life Science AAI supports at least the same authentication providers than ELIXIR and BBMRI-ERIC AAI. Otherwise, the users won't be able to find their familiar authentication provider in the Life Science AAI discovery services.
- that some authentication providers use targeted user identifiers (SAML2 persistent nameID) which means that their identifiers do not match in ELIXIR or BBMRI-ERIC AAI and Life Science AAI. This implies that the migration of those user records cannot be done automatically. In ELIXIR AAI, this has been estimated to affect about 3% of the user population.

3.2. Migration of ARIA

ARIA runs a completely independent authentication and authorisation infrastructure that cannot be migrated wholesale into LS AAI. Instead a phased migration is proposed with the existing software of ARIA being integrated to LS AAI with increasing degrees by which users are directed to LS AAI.

Integration Phase - Authenticating separately with LS AAI or with existing ARIA account. Authentication workflow is adapted to allow Life Science ID to be selected alongside ARIA ID directly - allowing users to opt-in to Life Science ID.

Migration Phase - Authenticating using LS AAI with existing ARIA account. Given successful integration the authentication workflow for the ARIA service will need to be adjusted visually to support the authentication with LS AAI for users with existing ARIA accounts. LS AAI will be visually prominent within the ARIA login process and alternative logins will be available through a click-through link.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Adoption Phase - Authenticating using LS AAI with existing/new LS AAI account. Should the majority of users in ARIA use the LSAAI authentication mechanism over time then we will disable all other authentication mechanisms and exclusively adopt the LS AAI login.

4. Relying services of the Life Science AAI

Relying services make use of the authentication and authorisation services of the Life Science AAI. They are the consumers that enforce access based on the attributes managed by the Life Science AAI. The relying services can also inject some user attributes back to the Life Science AAI (such as, the compute quota a user has left in a private cloud), but those attributes are not visible to other relying services.

This section describes some key relying services of the Life Science AAI from the other Work Packages (WP) of EOSC-Life project and beyond.

4.1. Other EOSC-Life activities

EOSC-Life WP2 "**Tools collaboratory**" aims at making tools and workflows interoperable and reusable in the EOSC across RIs. WP2 will integrate the tools and workflows (e.g. Galaxy[22] [RASC19]) to controlled access data in cooperation with WP5 and WP7. Life Science AAI can facilitate also other WP2 services when user authentication is needed.

WP3 "**Demonstrators and Open Calls** for User Projects" supports a set of selected demonstrators and organises open calls for specific topics to provide a fast, coordinated and user-oriented way to start building RI's connection to EOSC. Some of the demonstrators and open call projects are expected to need the authentication and authorisation services provided by the Life Science AAI.

WP7 "**Cloud deployment services**" is to integrate to community and commercial clouds to provide cloud resources to other WPs and to adopt related interoperability standards. That requires user authentication and authorisation by the Life Science AAI, including managing access to sensitive data.

4.2. Research infrastructure activities

This section introduces potential other relying services of the Life Science AAI from the research infrastructures in broad categories. For instance, in the end of 2020, ELIXIR AAI has 102 registered relying services[23].

Collaborative tools, such as wikis and web portals, often authenticate a user to provide personalised user experience based on their role in the research infrastructure. Many of the collaborations are not particularly sensitive, but some of them also impose requirements on the

[22] http://galaxyproject.org/
[23] https://login.elixir-czech.org/services

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

assurance of user authentication and sophisticated management of authorisation. Examples services are Euro-BioImaging Web Portal[24], ELIXIR Intranet[25] and ELIXIR e-Learning Platform[26].

Instruments, such as imaging tools and genome sequencers are expensive and limited resources and therefore need access management. Also, the data they produce needs access control if it is sensitive (e.g. human data) or contains research results that are not published yet. An example service is Instruct ARIA.

Data Archives and biobanks require access management if they hold sensitive data from humans. In the classic approach, the secondary use of controlled access data requires permission from the original data collector, commonly represented by a data access committee. A more lightweight approach (so called registered access) to less sensitive data (e.g. information on allele frequencies in a cohort) has also been proposed. Example services are the European Genome-phenome Archive (EGA)[27] and BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort)[28].

Clouds and other computing environments integrate computing to data. Users need to receive a resource allocation to be able to initiate jobs which then start to consume the resource quota granted for them. If the computation is done on sensitive data, additional access controls and security enhancements are expected. Examples of services are de.NBI cloud[29] [BELL19], EMBL-EBI Embassy[30] which already support ELIXIR AAI and Helix Nebula Scientific Cloud[31] and Google Cloud Platform[32].

Workflows and pipelines are examples of analysis tools processing data, potentially running in a cloud. Users need to authenticate to keep their data and results separate from others. If the size of the job is non-trivial, also a separate authorisation to run the analysis may be needed. Analysis on sensitive data requires additional safeguards. An example of tools and pipelines is Metapipe for Marine metagenomics [RAKN18] and Galaxy[33] [RASC19].

References

AARC19 AARC Community. AARC Blueprint Architecture 2019. November, 2019. https://doi.org/10.5281/zenodo.3672784

BELL19 Belmann, P., Fischer, B., Krüger, J., Procházka, M., Rasche, H., Prinz, M., Hanussek, M., Lang, M., Bartusch, F., Gläßle, B., Krüger, J., Pühler, A., Sczyrba, A. de.NBI Cloud federation through ELIXIR AAI. June,

[24] https://www.eurobioimaging.eu/
[25] https://elixir-europe.org/intranet
[26] https://elixir.mf.uni-lj.si/
[27] https://ega-archive.org/
[28] http://www.bbmri-eric.eu/scientific-collaboration/colorectal-cancer-cohort/
[29] https://www.denbi.de/cloud
[30] https://www.embassycloud.org/
[31] https://www.hnscicloud.eu/
[32] https://cloud.google.com
[33] http://galaxyproject.org/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

2019. https://doi.org/10.12688/f1000research.19013.1

CORB19 Haley, N. CORBEL Report on Common Access Framework Concept. May, 2019. http://doi.org/10.5281/zenodo.2662134

CORB19b Linden, M., Boiten, J., Courtot, M., Holub, P., van de Geijn, G., ; van Enckevort, D., Lappalainen, I., Nyrönen, T., Parkinson, H., Reihs, R., Senf, A., Spalding, D., Swedlow, J., Swertz, M., Törnroos, J., Kankaanpää, P., van Iperen, E. CORBEL Prototype implementation of distributed automated data access request, review and authorization and delivery systems. June, 2019. http://doi.org/10.5281/zenodo.3238496

EC21 European Commission. Directorate-General for Research and Innovation. Directorate G — Research and Innovation Outreach. Unit G.4 — Open Science. EOSC Authentication and Authorization Infrastructure. Report from the EOSC Executive Board Working Group (WG) Architecture AAI Task Force (TF). January, 2021. https://op.europa.eu/s/oMAa

eduGAIN eduGAIN Policy Framework: Constitution. 1 May 2017.

FIM4R18 Atherton, C., Barton, T., Basney, J., Broeder, D., Costa, A., van Daalen, M., Dyke, S., Elbers, W., Enell, C., ; Fasanelli, E., Fernandes, J., Florio, L., Gietz, P., Groep, D., Junker, M., Kanellopoulos, C., Kelsey, D., Kershaw, P., Knapic, C., Kollegger, T.,
Koranda, S., Linden, M., Marinic, F., Matyska, L., Nyrönen, T., Paetow, S., Paglione, L., Parlati, S., Phillips, C., Prochazka, M., Rees, N., Short, H., Stevanovic, U., Tartakovsky, M., Venekamp, G., Vitez, T., Wartel, R., Whalen, C., White, J., Zwölf, C. Federated Identity Management for Research Collaborations. June, 2018. http://doi.org/10.5281/zenodo.1296031

LICE21 Licehammer, S., Linden, M., Smith, C. EOSC-Life Access and User Management System for Life Science – the implementation and usage report. EOSC-Life deliverable D5.2. February, 2021. https://doi.org/10.5281/zenodo.4559400

LIND13 Linden, M., Nyrönen, T., Lappalainen, I. Resource Entitlement Management System. Selected papers of TNC2013 conference. http://tnc2013.terena.org/getfile/870

LIND18 Linden, M., Jalkanen, T., Törnroos, J. Sensitive dataset access management. ELIXIR Poster. June, 2018. https://doi.org/10.7490/f1000research.1115547.1

LIND19 Linden M., Bucik D., Gormanns, P., Haley N., Heikkinen J., Holub P., Kankaanpää p., Kouril, D., Kuba, M., Licehammer, S., Matyska, L., Nyrönen, T., Prochazka, M., Reihs, R., Romano, P., Sanderson, F., Shigaval, A., Smith, C., Tedds, J. Access and User Management System for Life Science - the Blueprint.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

EOSC-Life deliverable D5.1. September, 2019. https://doi.org/10.5281/zenodo.3386307

RAKN18 Raknes, I., Bongo, L. META-pipe Authorization service. https://doi.org/10.12688/f1000research.13256.2

RASC19 Rasche, H. UseGalaxy.eu: Community, Training, Infrastructure, and Users. ELIXIR Poster, July 2019. https://doi.org/10.7490/f1000research.1117097.1

Abbreviations

AAI	Authentication and Authorisation Infrastructure
API	Application Programming Interface
BMS	Biological and Medical Sciences
CLI	Command-line Interface
DAC	Data Access Committee
DURI	Data Use and Researcher Identity
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulation
IAM	Identity and Access Management
ID	Identifier
IdP	Identity Provider
LS	Life Sciences
OIDC	OpenID Connect
RI	Research Infrastructure
SAML	Security Assertion Mark-up Language
SP	Service Provider
WP	Work Package

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Delivery and schedule

The delivery is delayed: No

Adjustments made

None

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Appendix 1: Technical requirements specification

Life Science AAI

Date	Editor	Change
10 June 2019	Mikael Linden	Approved remaining changes to append the document to EOSC-Life deliverable D5.1
11 March 2021	Mikael Linden	Approved the updated document in the Technical WG call to be appended to EOSC-Life deliverable D5.3

Table of contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- 4.1. Home Organisation Affiliation(s) of a user
- 4.2. Home Organisation properties
- 4.3. Researcher status and attestations
- 4.4. GA4GH Passports
- 4.5. Groups
- 4.6. Dataset authorisation
- 4.7. Resource capabilities
- 4.8. Other attributes
- 4.9. Life Science ID user profile panel

5. Access control
   - 5.1. Active role selection
   - 5.2. Access control enforcement during login
   - 5.3. Life Science AAI Test environment

6. Technical interfaces
   - 6.1. Federated login and attribute release
   - 6.2. Attribute retrieval from external sources
   - 6.3. Credential translation
   - 6.4. User synchronisation
   - 6.5. Provisioning

7. Logging, statistics and data retention

8. Information security

9.Usability
   - 9.1. Ease of use
   - 9.2. Usability expert review
   - 9.3 Consistency

10. Capacity

11. Accessibility & Compatibility

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

1. Introduction

1.1. This document

This document presents the requirements for a Life Science AAI, the common Authentication and Authorisation service portfolio for the research infrastructures participating in the EOSC-Life project and beyond. The document intends to serve the design and deployment of the Life Science AAI.

This document is prepared by the Work Package 5 of the CORBEL project together with the AARC project and amended by EOSC-Life WP5. The work is based on the use case gathering among the AAI experts of the participating research infrastructures and the Life Science AAI pilot in the AARC2 project.

This document describes the Life Science AAI requirements as understood by the contributing research infrastructures at the time of writing. Some requirements may change over time as the needs of the Life Science community evolve and as the contributors learn them better. This document tries to highlight the key factors relevant for the success of the Life Science AAI.

In addition to this Technical requirements specification, there are other documents that describe other aspects of the Life Science AAI, including

- Requirements on service levels
- Requirements on organisational and legal aspects, including data protection

1.2. Terms

AAI	Authentication and authorisation infrastructure. The services described in this document for the Life Science community.
Account	A user account in an authentication provider external to the Life Science AAI, such as the researcher's Home Organisation or a Commercial company.
Authentication provider	An organisation external to Life Science AAI that manages users' Accounts and authenticates them in the Life Science AAI.
Bona Fide researcher	A researcher in good standing. An extra user attribute issued and managed by Life Science AAI, as described in section 4.3. Relying services may decide to make use of the Bona Fide attribute in access control enforcement.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Home Organisation	The university, research institution, company or other organisation that employs the user or where the user is otherwise affiliated with. Potentially the user's Authentication provider.
Identity, ID	Collection of attributes belonging to a certain user.
Identifier	An attribute that uniquely identifies a user.
Life Science ID	An umbrella term referring to both a Life Science user ID and a Life Science service ID.
Life Science service ID	A Life Science ID which is used by services which need to authenticate with other services. A Life Science service ID is owned by at least one Life Science user ID holder who is responsible for the service ID.
A Life Science user ID	A Life Science ID which the Life Science AAI issues to a natural person who registers to the Life Science AAI.
Relying party	An organisation that manages a Relying service.
Relying service	A service that makes use of the authentication and authorisation services of the Life Science AAI.

2. Identity and identifiers

2.1. Life Science ID

There are two kinds of identities which are commonly referred to as "Life Science IDs" or simply "users":

- Life Science user IDs
- Life Science service IDs

Any natural person can register a Life Science user ID. Shared accounts (such as, "operations manager in-duty") are not allowed. To register a Life Science ID a user needs to commit to an Acceptable Usage Policy (section 3.4).

A Life Science service ID can be assigned to a service. They are distinguishable from the Life Science user IDs assigned to natural persons.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

2.2. User identifiers

Each user is assigned two identifiers: one Life Science Identifier and one Life Science username. Both identifiers are non-reassignable (i.e. their value cannot be later recycled to another user).

Life Science identifier is an opaque and non-revocable identifier (i.e. it cannot change over time)

- It carries the syntax of eduPersonUniqueID, which consists of "uniqueID" part and fixed scope "lifescience-ri.eu", separated by at sign
- The uniqueID part contains up to 64 alphanumeric characters (a-z, A-Z, 0-9)
- N.B. eduPerson defines the comparison rule caseIgnoreMatch for eduPersonUniqueID, implying there must be no two users whose Life science identifier collides in a case insensitive comparison
- Example: 28c5353b8bb34984a8bd4169ba94c606@lifescience-ri.eu

Life Science username is a user selected, human-readable, revocable identifier (i.e. the user can change it)

- It carries the syntax of eduPersonPrincipalName, which consists of "user" part and fixed scope "lifescience-ri.eu", separated by at sign
- the user part (syntax derived from Linux accounts (reference)) begins with a lower case letter or an underscore, followed by lower case letters, digits, underscores, or dashes. In regular expression terms: [a-z_][a-z0-9_-]*?
- Intended use: when user's unique identifier needs to be displayed in the UI (e.g. wikis or Unix accounts)
- The usernames beginning with an underscore are dedicated to Life Science service IDs.
- Example: mike@lifescience-ri.eu

The Life Science identifier and Life Science username "test@lifescience-ri.eu" are test accounts reserved for testing and monitoring the proper functioning of the Life Science AAI. The Relying parties should not authorise it to access any valuable resources.

2.3. Cardinality of identities

Each user is supposed to register only one Life Science ID which follows them during their career although they may change their affiliation (It is believed that it would be confusing for the users themselves to have several, causing extra workload in the AAI helpdesk).

The Life Science AAI will implement checks to prevent users incidentally creating parallel Life Science IDs (for instance, name and e-mail address comparisons when a new Life Science ID is registered). However, there is no way to fully prevent a user having several parallel Life Science IDs.

The administrator must have the capacity to delete a Life Science ID if a user has unintentionally created several.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

3. Registering with and authenticating to Life Science AAI

3.1. Registering a Life Science ID

Registering a Life Science ID is triggered by the user themselves by

- The user browsing to "register" page, or
- A Relying service redirecting a user to register page

To start the registration process, the user needs to

1. Select their authentication provider (see the next section) and authenticate at it
2. Commit to the Acceptable Usage Policy (section 3.4) and
3. Enter their e-mail address and other necessary personal data on themselves (at least select their Life Science username)
4. Demonstrate they control the e-mail address they entered.

3.2. Supported Authentication providers and their discovery

For user authentication the Life Science AAI supports following authentication providers. The user is supposed to have an account in at least one of them and the users are supposed to link that account to their Life Science user ID:

- Identity Providers managed by researchers' Home Organization (via eduGAIN interfederation service)
- Research infrastructures (such as, ARIA)
- Commercial (such as, Google)
- ORCID
- Hostel Identity Provider (see the next section)

Apart from the Hostel Identity Provider, the Life Science AAI does not issue passwords for Life Science user IDs. Life Science service IDs can have credentials (e.g. password) associated.

The Discovery service (the UI for a user to select their Authentication provider) displays

- The user's previously used authentication provider(s) (up to 3),
- The recommended authentication provider if specified by the relying service (e.g. users authenticating via Life Science AAI to use ARIA SP should see ARIA IdP highlighted as a recommended authentication provider),
- The eduGAIN Identity Providers which
  - signal support to REFEDS Research and Scholarship entity category, or
  - signal support to GEANT Data Protection Code of Conduct entity category, or
  - have been demonstrated to release the necessary attributes. The release of such necessary attributes is checked by a user logging in to the Life Science AAI's dedicated "attribute release test" page. Users are encouraged to perform this attribute release test by clicking an "Add my institution" button in the bottom of the discovery page.
- Other authentication providers listed above

Following attributes are required from Identity Providers in eduGAIN:

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Unique user identifier (eduPersonUniqueID, SAML subject-id, eduPersonTargetedID, SAML Persistent NameID or SAML pairwise-id)
- User's name
- User's e-mail address
- Affiliation (eduPersonScopedAffiliation or eduPersonAffiliation)
- schacHomeOrganization

3.3. Hostel Identity Provider

The Life Science AAI manages a Hostel Identity Provider for those users who cannot use any other Authentication providers listed in the previous section. The users can self-register to the Hostel Identity Provider which issues them a username and password. The username is the user's Life Science username.

It must be possible to upgrade a self-registered user identity in Hostel Identity Provider to a verified identity (IAP/medium or IAP/high, see section 3.7) if one of the designated persons in trusted organizations (typically one of national nodes of RIs) carries out the identity proofing for the Hostel identity holder. Such verification process must be documented by that designated organization. The Hostel Identity Provider must keep logs on the upgrade process for the audit trail.

The Hostel Identity Provider must provide authentication that qualifies to the REFEDS Single-Factor Authentication profile (section 3.7).

3.4. Acceptable Usage Policy (AUP)

The Acceptable Usage Policy of the Life Science AAI may change from time to time. Any time a user logs in, the Life Science AAI verifies if the user has committed to the latest AUP version and, if necessary, asks them to do it before they can continue. User's decision to commit to the AUP is recorded for the audit trail.

3.5. Account linking for Life Science user IDs

A user can link multiple accounts from multiple authentication providers (see section 3.2) to their Life Science user ID. Linking a new account is carried out by

- at first logging in using a previously linked account and subsequently the new account, or
- by demonstrating control of an e-mail account, using a procedure that is as secure as above

Account linking can be triggered by

- The user logs into their "Life Science ID user profile panel" (section 4.9) where they can manage their Life Science ID and start linking a new account, or
- The user is trying to log in using a previously unknown account after which the Life Science AAI provides them with two alternatives: "Create a new Life Science ID" (section 3.1) or "Link an existing account".

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

The Life Science AAI sends a notification of ID linking to its holder's registered e-mail address in Life Science AAI and to the e-mail address associated to the new linked ID, if possible.

A user can unlink an account in the "Life Science ID user profile panel". After unlinking an account the user cannot use that account any more for login. The user cannot unlink their last account. If a user loses access to their last account, the Life Sciences AAI operations shall have a possibility to help them, but only according to the defined procedures which ensures there will be no security risk and only with explicit agreement from the user. All the steps must be audited.

3.6. Account management for Life Science service IDs

Each Life Science service ID must have at least one associated Life Science user ID that belongs to a natural person who manages the account and takes responsibility of the activity done using the service ID.

Any of the managers can

- Invite new managers
- Remove managers

3.7. Assurance framework

Life Science AAI supports issuing the following REFEDS Assurance Framework (RAF[34]) ver 1.0 values to the Life Science IDs and releases them to Relying services:

eduPersonAssurance (ePA) value	Implementation in Life Science AAI	Rationale
`$PREFIX$`	Always true	Life Science AAI fulfills RAF conformance criteria
`$PREFIX$/ID/unique`	Always true	(Unique-1) will be satisfied by policy (see section 2.1) and the AUP (Unique-2) will be satisfied by e-mail handshake when user registers to Life Science AAI (section 3.1) (Unique-3) will be satisfied by policy (see section 2.2) (Unique-4) will be satisfied by Life Science attribute profile

[34] https://refeds.org/assurance

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

`$PREFIX$/ID/eppn-unique-no-reassign`	Always true	Life Science AAI never reassigns ePPN (section 2.2)
`$PREFIX$/ID/eppn-unique-reassign-1y`	Always missing	Excluded by the previous row
`$PREFIX$/IAP/low`	Always true	Guaranteed by the e-mail handshake (section 3.1) when user registers to Life Science AAI
`$PREFIX$/IAP/medium`	True if passed by the Authentication provider	Life Science AAI relays the value provided by the Authentication provider.
`$PREFIX$/IAP/high`	True if passed by the Authentication provider	Life Science AAI relays the value provided by the Authentication provider.
`$PREFIX$/IAP/local-enterprise`	Always missing	Not applicable for research infrastructures
`$PREFIX$/ATP/ePA-1m`	Always true	ePA attribute carries the person's affiliation with the infrastructure
`$PREFIX$/ATP/ePA-1d`	Always true	ePA attribute carries the person's affiliation with the infrastructure
`$PREFIX$/profile/cappuccino`	True if /IAP/medium	Compound value
`$PREFIX$/profile/espresso`	True if /IAP/high	Compound value

Single/multi-factor authentication

Life Science AAI supports REFEDS Single factor authentication[35] and multi-factor authentication[36] as follows:

[35] https://refeds.org/profile/sfa
[36] https://refeds.org/profile/mfa

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Value	Implementation in Life Science AAI	Rationale
`https://refeds.org/profile/sfa`	True if passed by the Authentication provider	Life Science AAI is dependent on the authentication quality of the Authentication provider
`https://refeds.org/profile/mfa`	True if passed by the Authentication provider or Life Science AAI step-up authentication	Life Science AAI is dependent on the authentication quality of the Authentication provider. However, Life Science AAI step-up authentication (see next section) can deliver MFA authentication if the user's Authentication provider doesn't provide it.

3.8. Step-up authentication[37]

A user can associate a second authentication factor to their Life Science ID and a Relying service can ask the Life Science AAI to perform a step-up authentication using it. The second authentication factor can for instance be a smartphone app running in the user's phone.

The Step-up authentication service first checks if the user has an Authentication provider that supports REFEDS MFA (for instance, by issuing an authentication request with requested authentication context equals REFEDS MFA). If that fails the Step-up authentication service proceeds to enroll an MFA for the user.

4. Attributes and authorisation

In addition to the identifiers presented in section 2, the Life Science AAI can decorate Life Science IDs with attributes which are useful for the Relying parties to decide the user's permissions in their services. When possible, the attributes will follow the emerging AARC specification on Community identity attributes.

Each attribute is either

- Common, which means they are visible to all Life science research infrastructures or communities (Virtual organisations), or
- Community specific, which means the attribute is visible only to the Relying services of the Virtual organisation (research infrastructure or community) that manages it

[37] The requirement of associating Identity proofing to the registration/activation of the MFA was discussed but currently not included to the requirements

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

4.1. Home Organisation Affiliation(s) of a user

Each user can be affiliated to one or more Home organisations (such as, a university, research institution or private company) and the user's affiliations may change over time. A Relying service wanting to couple user's permissions to their continuing affiliation can observe the Home Organisation Affiliation attribute and their changes.

The syntax and semantic of the attribute follows the eduPersonScopedAffiliation attribute defined in eduPerson schema (version 201310). Following values are recommended for use to the left of the "@" sign:

`Faculty`	The person is a researcher or teacher in their home organization. The exact interpretation is left to the home organization, but the intention is that the primary focus of the person in his/her home organization is in research and/or education. Note. This attribute value is for users in the academic sector.
`Industry-researcher`	The person is a researcher or teacher in their home organization. The exact interpretation is left to the home organization, but the intention is that the primary focus of the person in his/her home organization is in research and/or education. Note. This attribute value is for users in the private sector.
`Member`	`Member` is intended to include `faculty`, `industry-researcher`, `staff`, `student`, and other persons with a full set of basic privileges that go with membership in the home organisation, as defined in eduPerson. In contrast to `faculty`, among other things, this covers positions with managerial and service focus, such as service management or IT support.
`Affiliate`	The `affiliate` value for eduPersonAffiliation indicates that the holder has some definable affiliation to the home organization NOT captured by any of `faculty`, `industry-researcher`, `staff`, `student` and/or member.

In other words, if a person has `faculty` or `industry-researcher` affiliation with a certain organization, they have also the `member` affiliation. However, that does not apply in a reverse order. Furthermore, those persons who do not qualify to member have an affiliation of `affiliate`.

Examples

- `faculty@helsinki.fi`
- `industry-researcher@zeiss.com`
- `member@ebi.ac.uk`

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

To become a holder of the `faculty`, `industry-researcher` or `member` attribute values in Life Science AAI, the user must either

- Perform federated login to the Life Science AAI using their home organisation's credentials, during which the home organization releases the related eduPersonAffiliation or eduPersonScopedAffiliation attribute, or
- Be assigned that identifier by a dedicated person in their home organisation

To become a holder of the `affiliate` attribute value, the user must either

- Use either of the two alternatives above, or
- Demonstrate he/she controls an e-mail address that belongs to the home organisation

The Life Science AAI keeps (caches) the attribute values described above for 12 months and releases them all to the Relying service. The freshness of the attribute values is guaranteed by asking a user to refresh the value every 12 months using the procedure described above.

A relying service can request a user to authenticate using a particular Authentication provider in order to receive fresh affiliation information from the related Home Organisation. The exact mechanism will be specified by the AARC community.

There must be a mechanism to revoke a person's affiliation immediately if needed.

4.2. Home Organisation properties

The Life Science AAI provides attributes[38] describing properties of a user's Home organisation:

- **Home organisation country**, following ISO 3166 codes (or "int" for an international organisation). The attribute is used for authorisation and statistics purposes
- **Home Organisation legal status**, following the definitions by the European Commission. This attribute is used for the EC funding application process and reporting

The properties are described by a mapping table from the Home Organisation Affiliation attribute. The Life Science community will define a management process for the table.

4.3. Researcher status and attestations

As described above, any natural person can register a Life Science ID. To narrow down the user base for Relying services limited to researchers, a user could apply for and receive further researcher qualifications, such as a "bona fide researcher" status[39].

The Life Science AAI has a service that can assign users one or more researcher qualifications based on, for instance,

- Their Home Organisation's ability to deliver `faculty@<home-organisation>` value (described above in section 4.1), or
- Another qualified researcher vouching for them or

[38] An attribute describing the research infrastructures a home organisation is affiliated with was discussed but not included in the requirements.
[39] See Registered access: authorizing data access: https://www.nature.com/articles/s41431-018-0219-y

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Them making an attestation that they commit to a certain community code.

4.4. GA4GH Passports

The Life Science AAI implements a Passport Broker, as defined by the Passport (ver 1.0.1)[40] and AAI (ver 1.0.2)[41] specifications of the Global Alliance for Genomics and Health. New versions of the specifications will be implemented and supported, depending on the resources available.

After authenticating a user the Life Science AAI Broker would gather their visas from several sources

- Data owners (Data access committees and data archives representing them) describing a user's permissions to controlled access data (ControlledAccessGrants visa), e.g. European Genome-Phenome Archive (EGA) and REMS tool that has a REST API for visa retrieval.
- Home organisations, describing a user's roles there (AffliationAndRole visa) as defined in section 4.1 above.
- Its internal sources, such as a person's status as a researcher (ResearcherStatus visa) and their attestation that they commit to a certain community code (AcceptedTermsAndPolices visa) as described in section 4.3. above.
- information on ID linking (LinkedIdentities Visa) as described in section section 3.5. above.

In the future, it may be necessary for LS AAI Broker to link also to other Brokers (e.g. Researcher Authentication Service of NIH in the US) for the exchange and aggregation of visas there.

4.5. Groups

The Life Science AAI has a service for managing users' group memberships and roles in the groups they belong to. Management of groups is done using a web interface.

Each user can belong to one or several groups. This is represented by the user having a "member" role in the group. A group member can have also arbitrary additional roles in the group, such as "secretary" or "chair".

Groups can be one of three types:

1. Secret group (where the group is not shown to anyone and the group creator/manager adds members manually).
2. Private groups (where users can have URL to the registration form for the group, and the group manager can approve or decline membership requests).
3. Public group where users are able to register as for the private group, but will be automatically added to the group without the need for group manager approval.

Each group has one or several managers who are able to

- delegate group manager role to other users and groups
- manage the group's properties (such as name)

[40] https://github.com/ga4gh-duri/ga4gh-duri.github.io/blob/master/researcher_ids/ga4gh_passport_v1.md
[41] https://github.com/ga4gh/data-security/blob/master/AAI/AAIConnectProfile.md

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- invite group members (requires confirmation by the invited user)
- add group members (no confirmation needed by the invited user)
- edit the type of the group (secret, private, or public)
- add other group as group member (members from other group became members of the group as long as other group is member of the group)
- remove group members
- assign and delete additional attributes (roles) for users in the group

The group manager needs to periodically confirm that the group is still active. The members of the group may need to periodically refresh their membership.

Groups have hierarchy i.e. member of a child group is automatically a member of the parent group.

4.6. Dataset authorisation

The Life Science AAI has a workflow service dedicated for the management of users' access rights to resources, especially to sensitive datasets. A user applies for access rights to the datasets by filling in and submitting an electronic application with the necessary attachments. The application is circulated to the individual or body (such as, a Data Access Committee) evaluating the applications and approving or rejecting them or returning them for amendments. If approved, the members of the application receive access rights to the resource applied.

The service has the necessary functionality for reporting and audit trail of the entitlements.

The service has interfaces for

- Bulk import for datasets' metadata from the data archive's catalogue for automated provisioning of the related application circulation workflows
- Launching data access application from an external source, such as the portal of the data archive
- Exporting the entitlements to an external system for access rights enforcement

4.7. Resource capabilities

The Life Science AAI enables Relying services to define capabilities which the service expects the individual users might be authorized to. Capabilities can represent virtually anything that the Relaying services support, for example it can be role, type of access (read-only, read+write), access to particular resource or even combination of aforementioned). It is up to each service to define which capabilities are supported and what are their semantics on the Relaying service. Capabilities will be represented in the Life Science AAI's Identity management component and the delegated managers might assign groups of users to the capability and by that authorize all users from that groups to use the capability.

This enables fine grained access control for Relaying services with capability support. Moreover, it decouples groups from access rights on Relaying services. Without that each service has to maintain a map between group name and the access right. That means that each time a new group should get a permission or each time the group name is changed, this information needs to

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

be communicated to all affected Relaying services and their managers have to put the new configuration in place. Using resource capabilities feature, the capabilities are defined upfront, during registration of a service, and later there is no configuration change needed on the Relaying service side. Based on user managed configuration in the Identity management component, the individual users will or will not be authorized to use defined capabilities when they are accessing services.

On a technical level, resource capabilities are expressed using eduPersonEntitlement attribute as it is defined in AARC recommendation AARC-G027 Guidelines for expressing resource capabilities.

4.8. Other attributes

The Life Science AAI supports adding arbitrary attributes to a Life Science ID, including

- If the Life Science ID is a user ID or a service ID
- User's name
- User's e-mail address (which is confirmed by an e-mail handshake)
- User's ORCID ID (which is recorded using ORCID APIs)
- Other wider researcher identifiers (such as, a researcher ID assigned to users by e-infrastructures) if they emerge
- User's public key (for SSH secure shell access)

4.9. Life Science ID user profile panel

Users can view their Life Science ID, attributes and linked accounts (section 3.5) and manage some of them in a dedicated web page "Life Science ID user profile panel".

User attributes are obtained from an Authentication provider, a Relying service or filled in by the user themselves. The user filled attributes are controlled solely by the user. Depending on the particular attribute the user might or might not have the rights to modify it by himself, but no other role (e.g. group manager) should have rights to modify it without the user's explicit permission. Ability to manipulate mentioned data by other parties creates a security risk and therefore it is strictly forbidden. The only exception from this rule is the Life Sciences AAI operations, which will have the right to modify this data, but this has to be done in accordance with the defined procedures, with explicit agreement from the user, and properly audited.

5. Access control

5.1. Active role selection

Some services expect a user to select the role they are currently acting in and couple the user's permissions to that role. For instance:

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- A user is associated to several projects (represented by the group membership attribute, see section 4.4 Groups) and they need to select the project they are currently active, providing them access to only those resources assigned to the project.
- A user is affiliated to several Home Organisations (represented by the Home Organisation affiliation attribute, see section 4.1) but their access rights are coupled to their continuing affiliation with a particular Home Organisation. The user needs to select the Home Organisation to which they want to couple their access rights.

Active role selection is an additional service which the Relying Service can subscribe. The relying service identifies the attribute(s) whose active value the user needs to select when they log in. The result is then mediated to the Relying Service.

5.2. Access control enforcement during login

A Relying service can subscribe an additional service where the Life Science AAI enforces access control after authenticating the user but before the user's browser is returned to the Relying service. The access control can be based on

- The user's membership in a particular group (see section 4.4),
- The user having sufficient level of identity and authentication assurance (see section 3.7) or
- Any other attribute of the user

If the Life Science AAI learns the user does not pass the criteria, it will (depending on the configuration made for each Relying service separately)

- display "Permission denied" message, or
- display "Permission denied" message and a free text message instructing the users on how to remedy, or
- (if permission is denied due to a missing group membership) display "Permission denied" message and the list of private and public groups whose members have access to the Relying service. The user can select a group which will redirect them to the registration form of the group

5.3. Life Science AAI Test environment

Life Science AAI has a Test environment for the Relying services to test their technical integration. When a new Relying Service is registered to the Life Science AAI, it is first exposed to the Test environment. After completing the tests and committing to the Life Science AAI policies for Relying Services, they are moved to the production use. Transfer to the production environment must not require any configuration updates for the Relying Service.

The Life Science AAI enforces access control of the Test environment (see the previous section). Only users who are members of a dedicated Test user group can access the Relying Services in the Test environment. For other users, the Life Science AAI displays instructions on how to apply for membership in the Test user group (see previous section). Membership in the Test user group expires in 30 days.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

6. Technical interfaces

6.1. Federated login and attribute release

The Life Science AAI provides an Identity/Service provider proxy with three primary interfaces for federated authentication and release of the attributes described in this document

- SAML 2.0, using the SAML2int profile or its successor
- OAuth2, including support to encoding attributes to access tokens as signed JWT
- OpenID Connect, including support to encoding attributes to claims in id-tokens and retrieving them from user-info endpoint.
- GA4GH Passports, as described in section 4.5.

It must be possible to configure what attributes are released to a Relying service for each Relying service separately.

The Life Science AAI pays attention to the smooth integration to the federated login with the Home Organisation credential via eduGAIN. The goal is that common end users do not need to face unnecessary technical hurdles for federated login.

6.2. Attribute retrieval from external sources

The LS AAI can retrieve attributes from external sources using a REST API during the OAuth2 and/or OIDC protocol flow and embed them to the claims and tokens released (see the previous section).

6.3. Credential translation

Relying Services can subscribe to the credential translation service of the Life Science AAI, allowing the users to obtain X.509 certificates based on the login described in the previous section.

6.4. User synchronisation

When needed, the Life Science AAI can synchronise users from external sources. That enables managing group membership within Life Science AAI from the external system. Users that weren't registered in Life Science AAI before, will have to approve the Acceptable usage policy before they can utilize any services provided within Life Science AAI.

The synchronization will be done periodically in a configurable time period depending on a particular use-case and the technical capabilities of connected external sources.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

6.5. Provisioning

Life Science AAI can provision user identities and attributes (such as group memberships) to Relying Services. Provisioning is done either by providing attribute authority or by pushing the data directly to Relying Service. Regardless of the provisioning method, the Relying Service should obtain only data about the users who are entitled to use the service. The provided data should be limited to minimal subset which is actually required by the Relying Service.

7. Logging, statistics and data retention

- The IdP/SP Proxy must collect appropriate logs.
- The AAI must provide anonymised statistics on # of Relying services, # of identities, # of logins (live and historical), # of logins by different Identity Providers to a given Relying service
- The AAI must display a public listing of current relying services both in test (section 5.3) and production environment, including a link to their privacy policy, location and organisation responsible for the service
- The AAI must follow data retention practices. Accounts must be closed if not used for 24 months. Users must be informed of the account closure well in advance.
- All the operations within the Life Science AAI must be recorded in audit logs

8. Information security

The Life Science AAI must be operated following professional information security practices.

The Life Science AAI must follow the security incident response framework described in Sirtfi v1.0[42].

9. Usability

9.1. Ease of use

All services and service components exposed to common end users must be easy and intuitive to use without any particular training or experience on similar services. Help text should be provided where required to enhance the user experience. All navigation options, buttons, and help text must be simple, clear, and concise.

All administrative interfaces (group manager, dataset authorisation, home organisation assignment) and relying service management interfaces must be easy enough to use after studying related online materials (manuals, videos, etc). Such materials should be provided in a centralised location and made accessible to all administrators.

[42] https://refeds.org/sirtfi

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Language should be familiar and non-technical i.e; technical terms and acronyms such as 'VO' should be avoided or explained, if avoiding is difficult.

9.2. Usability expert review

All services and service components will be exposed to a review by a usability expert and their providers are expected to implement reasonable improvements based on the review results.

9.3. Consistency

The Life Science AAI should allow templating determined by the SP that the user is coming from. The templating should be consistent throughout navigation on Life Science AAI pages.

10. Capacity

The Life Science AAI must have sufficient capacity to serve

- 25000 logins a day
- 100000 OpenID Connect introspections a day
- A peak of 500 OIDC requests (introspection or userinfo) simultaneously (i.e. within the timeout of the components)

There should be the potential to increase this capacity to meet increasing demand as the user base and number of relying services grow.

11. Accessibility & Compatibility

The user should be able to access the Life Science AAI regardless of the device (e.g. phone, tablet, PC), the operating system (e.g. Android, Mac OS, Windows, Linux) or the browser used. **Any browser version with >= 0.5% global usage must be supported**. At the time of writing this document this covers:

Evergreen Browsers

- Google Chrome
- Firefox
- Edge
- Opera

Desktop Browsers

- Internet Explorer; version 11 only as long as it will be supported by Microsoft
- Safari

Mobile Browsers

- iOS Safari
- Chrome for Android, Firefox for Android, UC browser for Android, IE Mobile

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

For the supported browsers, those browser versions are supported by Life Science AAI that are supported by the browser vendors.

Cross browser compatibility with other browsers should be on a best-effort basis. The Life Science AAI should also ensure compatibility with any new browser version which obtains greater than 0.5% global browser usage.

Font scaling should be fully supported and should scale the interface along with it such that users with large-font settings can properly read content on the system.

Large mouse pointers should be enabled, and large targets or hotspots provided.

Menus and controls should be accessible from the keyboard and/or voice activation. All content of the system should work with screen reader technology.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.