EOSC-LIFE: Providing an open collaborative space for digital biology in Europe

Deliverable D5.1

ACCESS AND USER MANAGEMENT SYSTEM FOR LIFE SCIENCE – THE BLUEPRINT

WP5 – **User management and access services**

Lead Beneficiary: **MU and Instruct**

WP leader: **Ludek Matyska (MU) and Suan 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 August 2019**
Actual delivery date: **4 September 2019**
H2020-INFRAEOSC-2018-2

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

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable is the 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 will be managed by the life sciences community and operated by the e-infrastructures, such as GEANT, EGI and EUDAT. The requirements of the Life Science AAI were initially developed by CORBEL WP5 and piloted with e-infrastructures in the AARC2 project. This deliverable lays the ground for the production deployment with e-infrastructures during the EOSC-Life project.

This deliverable and its appendices first introduces the technical and non-technical requirements for the e-infrastructure service offering and further clarifies how that relates to and integrates to the AAI service components operated by the life sciences research infrastructures. Relevant external work is then introduced. Finally, the deliverable provides an overview of Life Science AAI's potential relying services from EOSC-Life project and beyond.

PROJECT OBJECTIVES

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 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 have then served as key contributions towards the ongoing dialogue with the e-infrastructures on deployment and operation of the Life Science AAI whose first phase concluded with an official proposal from the e-infrastructures in June 2019. The technical and non-technical requirements are presented in Appendix A and B, respectively.

Some BMS RIs have been operating their RI-wide AAI services for several years, including BBMRI-ERIC AAI (343 users and 8 production services by the end of July 2019) and ELIXIR AAI (2174 users and 50 production services at the end of 2018). 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' forwards 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. DESCRIPTION OF WORK

2.1. LS AAI SERVICE ECOSYSTEM

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

- the 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;

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

- 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.

These two entities are shortly introduced in this section.

2.1.1. 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 (2.1.2 and Appendix B) which are both shortly highlighted here with references to more information in the appendices.

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 were invited to make their own proposal. However, the requirements can be easily met with an architecture that is compatible with the AARC blueprint architecture [AARC19].

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), home research infrastructure (section 4.2) 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. 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 on changes in user accounts even if the user is not directly interacting with them.

Non-technical requirements

The non-technical requirements on the Life Science AAI (Appendix B) 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 (Appendix B, section a);

- how the **monitoring** of the Life Science AAI, its availability, in particular, should be organised and reported (section b);
- 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 (section c);
- 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 (section d);
- 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 (section e);
- considerations on the **future evolution** of the e-infrastructure service offering (section f).

The technical and non-technical requirements have been provided to the e-infrastructures as the basis for the negotiation on operations of the Life Science AAI.

2.1.2. 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.

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.

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

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

(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.

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 118 biobankers registered to represent 311 biobank collections, and overall there are close to 250 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 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

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

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

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 organizations 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 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 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

[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/
[8] https://helpdesk.bbmri-eric.eu/
[9] https://directory.biobankinguk.org/

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.

Resource Entitlement Management System (REMS)

Resource Entitlement Management System (REMS) [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 audit trail.

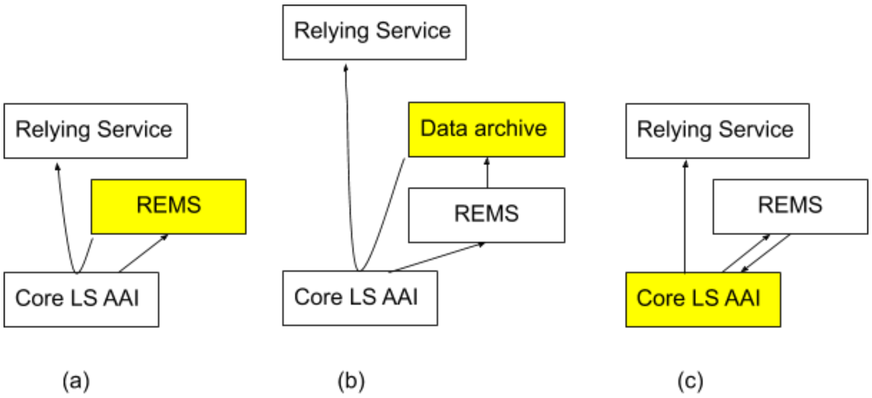


Figure 1: Deployment scenarios for REMS in the Life Science AAI. The yellow components indicate the authoritative source of permissions.

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 federated identity protocols, such as OIDC). This approach has been demonstrated in the Nordic Tryggve project[10]. In scenario (b) REMS is seen as a sub-components 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.

[10] https://neic.no/tryggve

2.1.3. 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[11] 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).

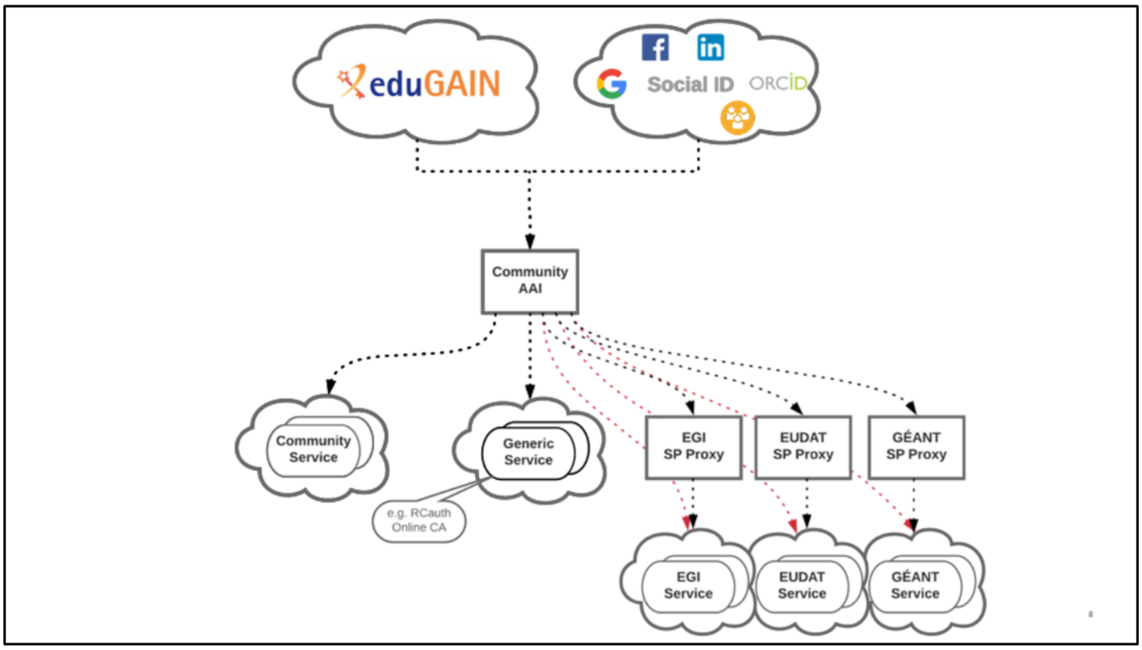


Figure 2: AARC2 Blueprint architecture introduces an approach where Community AAIs (such as Life Science AAI) manage users' identities and roles, group memberships and permissions and e-infrastructure AAIs are Service Provider proxies for e-infrastructure services. Source: AARC2 project.

To facilitate the integration of the AAI services, AARC2 project has introduced an AAI blueprint architecture (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, such as EOSC AAI (for EOSC services) or B2Access (for EUDAT 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.

[11] http://pan-data.eu/Umbrella

2.1.4. RELEVANT EXTERNAL WORK

This section introduces external actors 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.

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]

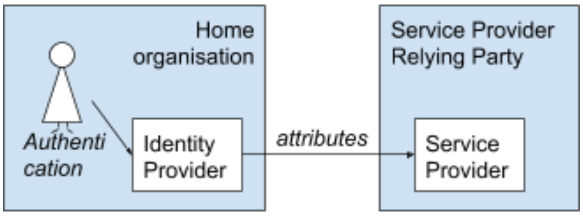


Figure 3: In federated identity management, a service (service provider) delegates user authentication to an identity provider server managed by the user's home organisation. [CORB19b].

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]

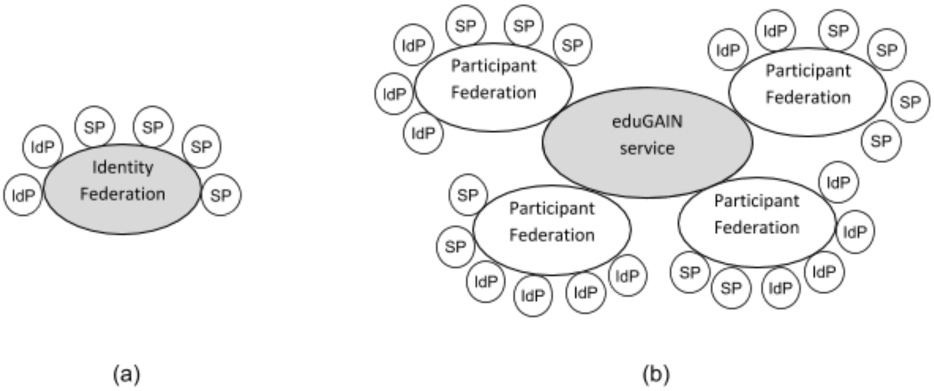


Figure 4: An identity federation (a) connects identity providers (IdP) and service providers (SP). eduGAIN interfederation service (b) further connects identity federations. [CORB19b].

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 (REFEDS[12]) 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[13] services, indicating that the service is operated for the purpose of supporting research and scholarship interaction.
- GEANT Data protection Code of Conduct[14], indicating that the service is following practices derived from the European data protection laws
- SIRTFI[15], indicating that the service follows community practices for information security, in particular those related to management of security incidents

Federated Identity Management for Research (FIM4R)

Federated identity management for research (FIM4R[16]) is a global cross-discipline network of IAM professionals in research infrastructures and services. After its first meeting at CERN in June 2011, FIM4R has had 12 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]

[12] http://www.refeds.org/
[13] https://refeds.org/category/research-and-scholarship
[14] https://wiki.refeds.org/display/CODE
[15] https://refeds.org/sirtfi
[16] https://fim4r.org/about/

Global Alliance for Genomics and Health (GA4GH)

Global Alliance for Genomics and Health[17] 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 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 is developing a specification for Researcher Identity and Access Claims[18], describing four attributes of a researcher:

- AffiliationAndRole claim expresses a user's role in their home organisation.
- AcceptedTermsAndPolicies claim describes 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 claim indicates that this person has been acknowledged to be a researcher, for instance by their home organisation or a peer.
- ControlledAccessGrants claim presents 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.

The Security Work stream is developing a GA4GH AAI 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. Both the Researcher ID claims and AAI specifications are still work-in-progress in the GA4GH and expected to complete the product approval process in October 2019.

2.2. 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.

2.2.1. 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[19]

[17] https://www.ga4gh.org/

[18] *A claim* is an OpenID Connect term for what is called *an attribute* in SAML (and this document). See Appendix 3 for the mapping of SAML and OpenID Connect terms.

[19] http://galaxyproject.org/

[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.

2.2.2. RESEARCH INFRASTRUCTURE ACTIVITIES

This section introduces potential other relying services of the Life Science AAI from the research infrastructures in broad categories. For instance, ELIXIR AAI has currently 62 registered relying services[20].

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 assurance of user authentication and sophisticated management of authorisation. Examples services are Euro-BioImaging Web Portal[21], ELIXIR Intranet[22] and ELIXIR e-Learning Platform[23].

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)[24] and BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort)[25].

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

[20] https://login.elixir-czech.org/services
[21] To be launched in autumn 2019.
[22] https://elixir-europe.org/intranet
[23] https://elixir.mf.uni-lj.si/
[24] https://ega-archive.org/
[25] http://www.bbmri-eric.eu/scientific-collaboration/colorectal-cancer-cohort/

security enhancements are expected. Examples of services are de.NBI cloud[26] [BELL19], EMBL-EBI Embassy[27] which already support ELIXIR AAI and Helix Nebula Scientific Cloud[28] and Google Cloud Platform[29].

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[30] [RASC19].

REFERENCES

AARC19 AARC2 project. Blue-print architecture. https://aarc-project.eu/architecture/

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, 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

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.,

[26] https://www.denbi.de/cloud
[27] https://www.embassycloud.org/
[28] https://www.hnscicloud.eu/
[29] https://cloud.google.com
[30] http://galaxyproject.org/

	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
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
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

DELIVERY AND SCHEDULE

The delivery is delayed: No

ADJUSTMENTS

Adjustments made: None

APPENDICES

APPENDIX 1: TECHNICAL REQUIREMENTS SPECIFICATION V2

The requirements were developed during CORBEL and reported as a CORBEL deliverable [CORB19].

Date	Editor	Change
22 Oct 2018	Mikael Linden	First draft of the updated requirements
30 Jan 2019	Mikael Linden	Elevated to Release Candidate by the BMS AAI telco
10 June 2019	Mikael Linden	Approved remaining changes to append the document to EOSC-Life deliverable D5.1

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. 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.
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.

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 “lifescienceid.org”, 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@lifescienceid.org

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 “lifescienceid.org”, 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@lifescienceid.org

The Life Science identifier and Life Science username “test@lifescienceid.org” 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.

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:

- Unique user identifier (eduPersonUniqueID, SAML subject-id, eduPersonTargetedID, SAML Persistent NameID or SAML pairwise-id)
- 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 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 management panel” (section 4.7) 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”.

A user can unlink an account in the “Life Science ID management 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, https://refeds.org/assurance) 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
`$PREFIX$/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 (https://refeds.org/profile/sfa) and multi-factor authentication (https://refeds.org/profile/mfa) as follows:

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

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 services proceeds to enroll an MFA for the user.

The enrollment of the second authentication factor must qualify at least to RAF `$PREFIX$/IAP/medium`.

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.

Each attribute is either

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

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). If necessary, a new attribute following the eduPersonScopedAffiliation syntax will be defined. 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`

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 freshness of the attribute values is guaranteed by asking them to refresh the value every 12 months using the procedure described above.

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

4.2. User's Research Infrastructures attribute

Universities, research institutions and other organisations may be affiliated with one or more research infrastructures, giving their users access to the research infrastructures' Relying services. User's Research Infrastructures attribute indicates to which research infrastructures the user's Home Organisation is affiliated with.

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[31].

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

4.4. 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).

[31] See Registered access: authorizing data access: https://www.nature.com/articles/s41431-018-0219-y

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 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)
- 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.

4.5. 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

4.6. 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
- The country in which the user's Home Organisation resides (if the user has several Home Organisations, the attribute may be multi-valued). This determines to what services the user has access (if at all - some services are for members only) and under what conditions (some services may be paid in the future for non-members, and discounted for observers, while free for members).
- User's public key (for SSH secure shell access)

4.7. Life Science ID management 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 management 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. Dependent 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 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:

- 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.

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.

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 the they can utilize any services provided within Life Science AAI.

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

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 (https://refeds.org/sirtfi).

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.

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. Life Science AAI must have cross browser compatibility with the following:

Desktop Browsers

- Google Chrome; latest version and the previous five versions. Currently from 67.0 to 72.0
- Firefox; latest version and the previous five versions. Currently from 60.0 to 65.0
- Edge; latest version and the previous three versions. Currently from 38 to 44
- Internet Explorer; latest version and the previous three versions. Currently from v8 to v11
- Safari; latest version and the previous three versions. Currently from v10.1 to v12.0.2

- Opera; latest version and the previous three versions. Currently from v55 to v58

Mobile Browsers

Latest version and versions of the previous two years (Chrome for Android, Firefox for Android, UC browser for Android, IE Mobile, and iOS Safari)

Cross browser compatibility with other browsers should be on a best-effort basis. The Life Science AAI should also ensure compatibility with any new browsers which obtain greater than 1% global browser usage.

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

Menus and controls should be accessible from the keyboard.

APPENDIX 2: NON-TECHNICAL REQUIREMENTS FOR THE LIFE SCIENCE AAI V1

5 Dec 2017	Mikael Linden	Initial draft
28 Aug 2018	Mikael Linden	Added a section on the proposed data controller model
5 Feb 2019	Ludek Matyska	New version, includes component availability categories, precise the purpose of the document
5-12th Feb 2019	AAI dev group	Multiple comments
12 Feb 2019	Ludek Matyska	New clean version
18 Feb 2019	Ludek Matyska	Final version

Introduction

This document defines non-technical – operational, organizational, and data protection related – requirements and specifications of the Life Science (LS) community (represented by the partners of the CORBEL project to be replaced by the EOSC-Life project community) to be fulfilled by the Life Science AAI implementation. The document complements the technical requirements specification and together they represent the assignment for LS AAI.

The Life Science community plans to outsource part of the LS AAI to an external party, preferably a party that consists of large European e-infrastructures' partners. In the proposed model, e-infrastructures (such as EGI, EUDAT and GEANT) are expected to operate the key technical components of the LS AAI, while the service will be owned by the Life Sciences community. This document defines how responsibility would be split between the e-infrastructures and the Life Science community for the duration of the initial phase (at least first two years, starting March 1st, 2019) of the EOSC-Life project. While the collaboration and outsourcing is expected to exist during the whole duration of the EOSC-Life project (4 years) and beyond, the specifications and requirements are expected to evolve. There will be a planned checkpoint after the first two years (this could especially touch, but not be limited to, the GDPR-related specification where we expect evolution of its interpretation).

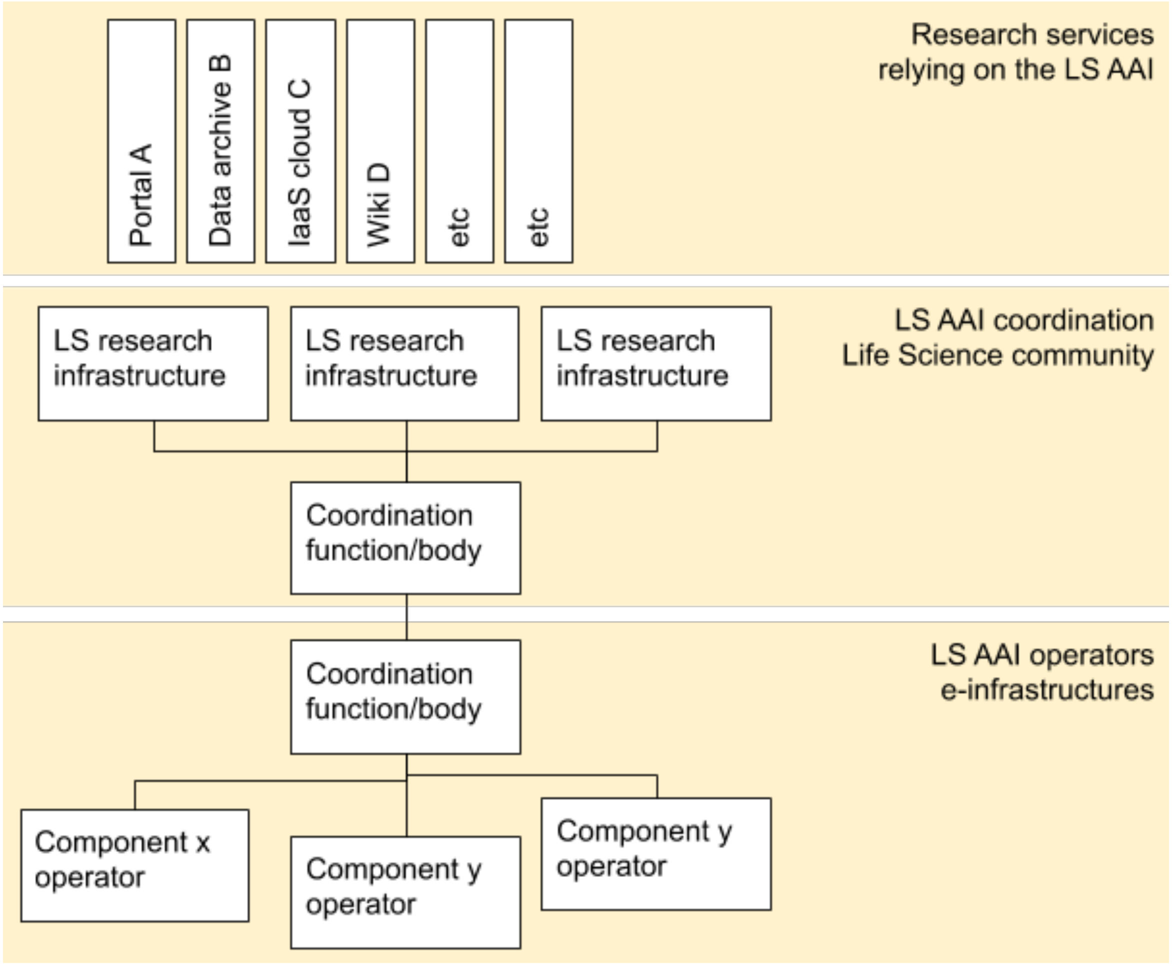
Overview

In the figure below we define three organisational layers to underpin the complex LS AAI (by LS AAI we understand the set of authentication and authorization capabilities described in the Technical Requirements specification):

- **LS AAI operators** are the organisations (e.g. e-infrastructures and/or their partners) running the technical components of the Life Science AAI.
- **LS AAI coordination** is a process through which Life Science community will organise the management and coordination of the LS AAI service. We expect that both the LS community and LS AAI operators will each create a single coordination body (as presented in the picture) that will represent each party during the negotiations and also during the actual operation of the LS AAI. Both such bodies must be authorized by the respective party to take decisions and to represent individual partners of each party, hiding thus the complexity of the LS community on one side and the LS AAI service on the other side.
- **Services relying on the LS AAI** are the customers of the LS AAI; the LS AAI exists to meet their needs for authentication and authorisation of the LS researchers. The services may be managed by the research infrastructures, organisations (such as universities or research institutes) affiliated with the research infrastructures, e-infrastructures or companies (such as, commercial cloud providers). It is assumed that the relying services do not participate directly in the coordination of the LS AAI.

The Figure describes the interaction at the coordination level, but does not exclude the interaction among parties at the technical level – e.g. the AAI operators are expected to directly interact with relying parties (enrolling them to the LS AAI), but the process is coordinated (and controlled) by the LS community.

We assume a formal level of coordination on behalf of both the LS community and the AAI operators. In addition to regular informal communications through several channels, a formal dialogue to include potential agreements between the LS AAI operators and


Portal A
Data archive B
IaaS cloud C
Wiki D
etc
etc
Research services relying on the LS AAI
LS research infrastructure
LS research infrastructure
LS research infrastructure
Coordination function/body
LS AAI coordination Life Science community
Coordination function/body
Component x operator
Component y operator
Component y operator
LS AAI operators e-infrastructures

Component availability categories

Based on the experience gained primarily through running the ELIXIR AAI in a production environment and through the LS AAI pilot operations, the LS community prepared a breakdown of expected components and/or services/functions of the future LS AAI production service associated with the expected availability and reliability category. The categories are:

- Red represents components where a malfunction makes the service essentially non-operational (severely degraded); malfunction stops significant number of users from logging in (they are not able to use the relying services).
- Yellow represents components where a malfunction stops some users from using the service (the impact is clearly constrained – e.g. users served by a particular IdP – and large majority of users is not impacted).

- Green represents components where malfunction leads to a delay in a usability, but does not affect ordinary users (for instance, group or attribute updates cannot be done or will not propagate but that does not impede login).

All the components must have a mechanism that can be used to monitor the authentication flow end-to-end from an end user perspective.

While the categories are set for now, there may be changes after sufficient experience with the LS AAI is gained. As stated in the Introduction, a check is expected after two years in operation.

Category	**Red**	**Yellow**	**Green**
Description	Malfunction stops significant number of users from logging in	Malfunction stops some users from logging in	Malfunction delays people’s work
IdP/SP proxy, its attribute-backend and associated services/modules, such as - IdP discovery - AUP commitment check - attribute release inform dialogue - active role selection - access control enforcement	X		
Hostel IdP	X		
Step-up authentication		X	
APIs for Relying services to read/write attributes		X	
User registration	X		
Other user’s identity management self-service - account linking - attribute self-management - Life Science ID management panel		X	
Other attribute management			X

- Home organisation affiliation - Researcher status			
Service ID management functions			X
User synchronisation and provisioning			X
Group management, dataset authorisation			X
Credential translation			X

Nota bene: Software services implementing the above stated functional components can fall into different categories. The software services that pose online dependency (i.e. their real-time availability is required to implement the given functional component) will consequently require the highest availability category of any components dependent upon that software. The software services that pose only offline dependency (i.e. services that periodically push their output into some functional component) generally have lower availability. For instance, the workflows to manage group memberships of a person, researcher status, or dataset authorisations belong to the green category; but the software service delivering their values for the IdP/SP proxy component assembling them to SAML assertions, OIDC claims or OAuth2 access tokens belongs to the red category as it is an online dependency for IdP/SP proxy functional component. Implementers may, for instance, choose to deploy an intermediate LDAP directory (red) serving the attributes to the IdP/SP proxy (red). The LDAP directory is updated by the back-end identity management services (green).

Requirements

The non-technical requirements are split into the following categories:

a) Operational

b) Monitoring

c) Governance and interaction model

d) GDPR and related data protection issues

e) Financial aspects

f) Further evolution of the LS AAI

Operational requirements

The LS AAI is expected to run around the clock, without interruptions. Therefore, the target for the service hours is 24/7/365 with no planned downtime. Any interruption of the service must be communicated immediately (via an agreed channel, see governance below).

The **availability of the LS AAI** will be measured as a simple ratio of up-time and the wall clock time elapsed in a measured period. Both monthly and yearly values will be provided (see also monitoring below). This parameter should express user's view of the availability of the service, i.e. its full functioning, not just trivial accessibility.

While the **target availability** of all the services is **100%** (i.e. no interruptions of the service), the **acceptable availability numbers** are in the following table (differentiated by the component availability categorisation). The yearly aggregate is the primary one, the monthly are approximate, the reading should be: For 12 month period the yearly acceptance threshold is kept AND no monthly threshold is broken.

Availability Category	Monthly acceptable availability threshold	Yearly acceptable availability threshold
Red	99 %	99.9%
Yellow	95 %	99.5%
Green	90 %	98 %

The yearly numbers represents acceptable cumulative interruption of 8,76 hours, 1 day 19,83 hours, and 7 days 7.2 hours for Red, Yellow, and Green availability category, resp. For the Red and Yellow categories, the monthly numbers roughly allow to have the whole interruption within a single month, while for the Green category a cumulative monthly interruption cannot exceed 3 days in any single month.

The LS AAI operators must setup a manned **service desk** for the services LS AAI operators provide that will be available **Monday to Friday, 9-17 CE(S)T** (excluding mutually agreed public holidays). The LS operators are also expected to establish mechanisms for continuous monitoring and acceptance of service requests (such as an automated helpdesk/request tracking system). The out of service desk hours requests will be served on a best-effort basis.

The response time, defined as the time taken to respond to a user after an incident or a service request is recorded in the service desk, will depend on the availability category (mostly measuring general impact of the interruption) above combined with the actual identification of the relative importance of an incident, problem or change.

The following table provides proposal to be negotiated.

Category	**Description**	**Response Time**
Low priority	Anomalies have been detected, but overall user experience is unchanged	3 Service Desk working days
Medium priority	Service is degraded	1 Service Desk working day
Top priority	Service is interrupted and the incident/request needs to be addressed as soon as possible	4 support hours

There is a weak correspondence with the components availability categories presented in the previous section, i.e. even just a degradation of a service under the red availability category may need the top priority response while an interruption of a service in green category may fall under the medium priority response time. The priority classification of specific cases will be results of mutual agreements between LS AAI operators and LS community.

Monitoring

The LS AAI service and its components will be continuously monitored by LS AAI operators and the monitoring output will be made available to the LS community, e.g. via a dashboard and summaries, the precise monitoring method is to be negotiated. The monitoring must cover whether each individual component is 'live' and whether each individual component is functioning correctly (e.g. a ping to a web service is not sufficient, the ability of the service to provide correct responses must be monitored, too).

The monitoring should also **follow the components availability categorization above**, as different categories are expected to have different levels of service. While the internal structure of the whole LS AAI is the responsibility of the LS AAI operators, the detail shared with the LS community must be consistent with the categorization. A precise definition (semantics) of what is monitored and the complete list will be negotiated based on the proposal from LS AAI operators.

The **aggregate summaries are to be given on a monthly basis**, they will also be used as the primary values to measure the achieved quality of LS AAI and its components. The aggregates must also include information about any downtime and a brief explanation of the cause/remedy.

The LS community can provide its own monitoring tools that will be setup with a support from LS AAI operators where needed.

Governance and interaction model

As presented above, both the LS community and LS AAI operators are both expected to create a single coordination body that will represent each party during the negotiations and during the actual operation of the LS AAI. As the LS AAI operators are providing complex service and the LS community is not directly selecting individual components, the coordination body of the LS AAI operators must be able to take full responsibility of the correct behavior of the LS AAI and proper interaction between individual components.

On top of this requirement, the LS AAI operators must guarantee that for each Red category there is an additional single representative that hides its internal complexity; e.g. if there are several IdP/SP proxies operated by several e-infrastructures' partners, a single entity must represent them in a transparent way and must guarantee that the whole works as expected. The LS community should have no responsibility for the integration and interoperability of components provided by the LS AAI operators.

For the sake of proper communication, the LS community will be also represented by a single body that will be authorized to take decisions related to the LS AAI operation, evaluating the performance etc.

GDPR and related data protection issues

The role of Data Controller for the Personal Data processed for the LS AAI will be taken by an entity (or entities) within the LS community. The LS community considers creation of a Life Science Research Infrastructures consortium (further as LS-Consortium) or a similar entity for this purpose. In such case the LS AAI operators are expected to become data processors and will process the data on the basis of a contract with the Data Controller. During the first phase (at least the first two years), we expect that the

GDPR formalities will be covered through a participation in the EOSC-Life project and eventual additional GDPR related agreements; during that time the following areas are expected to be covered and the exact requirements will be developed before the end of this first phase:

- documented authorization procedures for access to all components of the AAI infrastructure;
- backup procedures to avoid loss of data;
- logging of access and modifications;
- ensuring up-to-date encryption of communication channels;
- provable destruction of data after termination of service;
- dealing with data protection for repairs and service;
- provide documentation or information tool which shows where the data is distributed in the infrastructure provided as a part of the service delivery;
- handling of Data Subject requests for access and portability of their Personal Data.

It is expected that these requirements will be first piloted internally as a part of the EOSC-Life project in order to assess their feasibility, while the final requirements will be covered as a part of the contract between the LS-Consortium and the LS AAI operators.

The whole LS AAI is expected to be compliant with the GEANT Data protection Code of Conduct v2.0 once approved by the authorities,[32] and the LS AAI operators are expected to contribute to the implementation of the compliance to the necessary extent.

Financial aspects

The cost of the LS AAI operation for 4 years starting 1st March 2019 will be provided through the EOSC-Life project. There is a sum of 600 kEuro available within the project budget to be used for the LS AAI operators (this money will not be available after the project end on 28th February 2023). This equates to 150 kEuro on average per annum although the actual spending pattern is negotiable and need not be linear.

In order to create a collaborative relationship between the LS and e-infrastructure communities, we do not plan to sign subcontracting agreements between the project and the LS operators (the e-infrastructures resp. their specific partners); instead, a more partnership-oriented involvement with the project will be sought.

A model of financial stability and sustainability for the long term LS AAI operation, beyond the duration of the EOSC-Life project, will be developed during the first phase of

[32] https://wiki.refeds.org/display/CODE/Data+Protection+Code+of+Conduct+Home

the EOSC-Life project. The LS AAI operators will be included into a working group that will analyse and define a sustainable funding model. Nevertheless, the funding available in the EOSC-Life project should fully cover the operation of the LS AAI for the whole duration of the project.

Further evolution of the LS AAI

The requirements of the LS community for the LS AAI features and functionality will evolve. The collaboration with an LS AAI operator must allow flexibility to enable evolution at the following levels:

1. Standard care and maintenance of the service, i.e. security and similar patches, small upgrades due to the natural evolution of components etc.
2. Larger upgrades required by the LS AAI operators as the results of their development of the LS AAI components, e.g. introduction of new versions, addition of features developed through the LS AAI operators own R&D work, removal of obsolete components etc.
3. Integration of new or modified functions/features, based on LS community requirements developed either on LS Community request by the LS AAI operators or by the LS community itself.

The first level is expected to be covered by the LS AAI operators standard method of operation, under the discretion of the LS AAI operators. These modifications do not require explicit approval of the LS community prior to implementation.

The second level is also to be performed under the discretion of the LS AAI operators, however in these cases an agreement of introduction of new versions, large scale upgrades etc. is communicated to the LS community and an explicit approval is needed before the implementation.

For the third level, the LS community requires a mutual understanding that the whole LS AAI will evolve together with the evolution of the LS community needs. This implies that new features and/or functions will be developed (or existing will be modified) and will need to be (re-)integrated into the LS. It is understood that any such integration and development needs a mutual agreement and a one time cost may be associated with these activities in specific cases; on the other hand the LS AAI operator must not reject or extensively delay such development or integration requests and must be prepared to work with the LS community to enable further evolution of the LS AAI.

APPENDIX 3: MAPPING OF SAML AND OIDC TERMS

Life Science AAI supports both SAML (Security Assertion Markup Language) and OpenID Connect (OIDC) protocols which use slightly different terms for the same concepts. The table below presents a mapping of some key terms used in SAML and OIDC.

Security Assertion Markup Language (SAML) term	**OpenID Connect (OIDC) term**
Attribute	Claim
Identity Provider, IdP	OIDC provider, OP
Service Provider	Relying party