COMMON AUTHENTICATION AND AUTHORISATION SERVICE FOR LIFE SCIENCE RESEARCH

Mikael Linden, ELIXIR-Finland, mikael.linden@csc.fi, https://orcid.org/0000-0002-3634-3756
Petr Holub, BBMRI-ERIC, petr.holub@bbmri-eric.eu, https://orcid.org/0000-0002-5358-616X
Ilkka Lappalainen, ELIXIR-Finland, ilkka.lappalainen@csc.fi, https://orcid.org/0000-0001-5762-893X
Ludek Matyska, ELIXIR-Czech, ludek@ics.muni.cz
Tommi Nyrönen, ELIXIR-Finland, tommi.nyronen@csc.fi, https://orcid.org/0000-0002-5569-5183
Michal Procházka, ELIXIR-Czech, michalp@ics.muni.cz, https://orcid.org/0000-0001-8376-9533
Jonathan Tedds, ELIXIR-Hub, jonathan.tedds@elixir-europe.org, https://orcid.org/0000-0003-2829-4584
Pasi Kankaanpää, Euro-BioImaging, Åbo Akademi University, pkankaan@abo.fi
Philipp Gormanns, INFRAFRONTIER GmbH, philipp.gormanns@infrafrontier.eu, https://orcid.org/0000-0001-9823-1621
Michael Raess, INFRAFRONTIER GmbH, michael.raess@infrafrontier.eu, https://orcid.org/0000-0002-8759-1186
Natalie Haley, Instruct-ERIC, natalie@strubi.ox.ac.uk, https://orcid.org/0000-0003-2511-0743

Keywords

Authentication, authorisation, AAI, research infrastructures, life sciences

Abstract

Research communities and scientific service providers are often required to authenticate researchers and manage their access rights to sensitive data, detection instruments or online computation resources. The research infrastructures in the life science sector have collated a joint set of requirements for a Life Science Authentication and Authorisation Infrastructure (AAI) which is currently in a pilot state. For the first time, several research infrastructures have come together to define requirements on a common AAI and work with the e-Infrastructures to provide a prototype based on these requirements. The Life Science AAI has the potential to become a widely used service that provides both users and research infrastructures with easier, streamlined and compatible authentication functionality. It can also serve as an exemplar for science AAI development in general. This paper presents the requirements of the Life Science AAI and the rationale behind it. An overview of the non-technical challenges in the upcoming deployment phase is also presented.

1. Introduction

Many research infrastructures in the life sciences have a requirement to authenticate researchers who are accessing online services and to manage their access rights for these services. For simple services such as collaborative tools (e.g. wiki), the functional and security requirements are moderate. However, in more complex services the sensitivity of the service (such as controlling access to repositories of data or biological samples provided by human donors), or cost of the service (such as expensive instruments or computing resources) requires careful authentication and fine-grained access control. These are complemented by the risks of a failure in security practices which are imposed on the service providers and their liabilities under the data protection and privacy protection laws.

Researchers access services from different scientific service providers that in turn issue user accounts for authentication. As a consequence, a researcher has to manage a number of user accounts. Furthermore, these accounts are not standardized by their creation and management processes or the information they contain. For example, an account can be created based on self-registration and self-asserted identity information. As researchers move home university or research institution such accounts are not necessarily closed appropriately. For the scientific service providers, poorly managed user accounts impose a security risk. For the researcher, maintaining a number of the independent accounts is a nuisance. For services dealing with data that are subject to data protection laws, the user has the additional burden of an identity verification process to ensure more than self-asserted identity alone - sometime even requiring physical travels to demonstrate their identities using physical ID cards.

Research infrastructures now assume a role as organisations who provide common services for research projects and scientific service providers. One of the potential service areas is Authentication and Authorisation Infrastructure (AAI), which is a service portfolio for authenticating researchers and helping the scientific service providers (called relying services) decide what the end-users are permitted to do in the services. A research infrastructure AAI helps the relying services with their user management burden and provides the researchers with

a possibility to have fewer credentials for accessing more services. When deployed centrally by the research infrastructures, the AAI systems can provide a better service for a lower cost than if done in each service separately.

CORBEL (www.corbel-project.eu) is an initiative of thirteen biological and medical research infrastructures (see Table 1), which together will create a platform for harmonised user access to biological and medical technologies, biological samples and data services, required by cross-interdisciplinary research projects using the platforms and technologies of different life sciences research infrastructures. CORBEL WP5 aims to develop a common access framework that facilitates user access to services and resources across the involved research infrastructures.

Table 1. Biological and medical research infrastructures participating in the CORBEL project.

Research infrastructure	**Focus area**
BBMRI-ERIC	Biobanking and biomolecular resources
EATRIS-ERIC	Translational research
ECRIN-ERIC	Clinical trials
ELIXIR	Curated databases
EMBRC	Marine model organisms
EMPHASIS	Plant phenotyping
ERINHA	Highly pathogenic microorganisms
EU-OPENSCREEN	Screening and medicinal chemistry
EURO-BIOIMAGING	Advanced imaging technologies
INFRAFRONTIER	Functional genomics
INSTRUCT-ERIC	Structural biology
ISBE	Systems biology
MIRRI	Microorganisms

In May 2016, CORBEL WP5 organised a workshop where the participating research infrastructures agreed to work on defining a common AAI for life sciences. The work started in the autumn 2016 by collecting and documenting use cases for the Life Science AAI. In spring 2017 these use cases were translated into a requirements specification which was provided as input for a pilot project to the GÉANT-coordinated AARC2 project (www.aarc-project.eu). This paper focuses on the requirements specification and the rationale behind it.

The rest of this paper is organised as follows. Section 2 presents the requirements that the Life Science AAI has for the identity and authentication of the user, and section 3 continues with the requirements for the management of user attributes and authorisation. Section 4 describes the alternative interfaces the relying services can use to consume the authentication and authorisation services that the Life Science AAI provides to them. Section 5 provides a short overview of the non-technical issues identified for a proposed deployment project. Section 6 presents some related work and section 7 concludes the paper.

2. Requirements for identity and authentication

This section presents the requirements the Life Science AAI has for user identity and authentication. The section also describes how a user can link multiple external identities to their Life Science ID. The work done on identity and authentication assurance is briefly introduced.

2.1. Life Science ID

A (digital) identity is the abstraction of a user in an IT system and consists of their unique identifier(s) and other user attributes. In the Life Science AAI, users are assigned a Life Science ID, which uniquely identifies them and consists of two identifiers: Life Science identifier and Life Science username.

The Life Science identifier is an opaque and non-revocable identifier, following the syntax of eduPersonUniqueID [MACE16] attribute and having the scope "lifescienceid.org". It is assigned to the user by the Life Science AAI, does not change over time and is not normally visible to the user. In contrast, the Life Science username is a human-readable, revocable but non-reassignable identifier that the user can select themselves. It is supposed to be used when the user identifier is presented in the user interface (for instance, in wikis and other collaboration tools)

and follows the syntax of the eduPersonPrincipalName [MACE16] attribute with the scope "lifescienceid.org". In other words, as a user-chosen identifier, the Life Science username can change over time (e.g. if a person's name changes) but the old username is not circulated to other persons. Table 2 below provides examples on the two identifiers.

Table 2. Life Science ID consists of two identifiers.

Identifier	**Syntax [MACE16]**	**Example**
Life Science identifier	eduPersonUniqueID	28c5353b8bb34984a8bd4169ba94c606@lifescienceid.org
Life Science username	eduPersonPrincipalName	mike@lifescienceid.org

For accountability reasons, the Life Science ID represents a single natural person, i.e., there are no shared user accounts. It is also assumed that a person registers only one Life Science ID, which they keep through their scientific career. There are no practical means to prevent a person from registering and using several parallel Life Science IDs[1]. However, having several parallel IDs is believed to be confusing especially for the users themselves as has been demonstrated by e.g. the Instruct and ELIXIR help-desks.

In certain circumstances there is a need to have a service ID that does not belong to any particular person but represents a service. An example of this is a service for replicating data sets between data centers using common protocols (such as, gridFTP) or file/directory ownership for institutional data. The Life Science AAI permits limited use of a service ID, provided that they are associated to one or more dedicated persons (Life Science ID holders) responsible for their use.

2.2. Authentication and account linking

A Life Science user can link their Life Science ID to one or more authentication providers that are external to the Life Science AAI. When a user registers their Life Science ID or when they log in into a relying service, their web browser is first diverted to the external authentication provider for authentication.

Optimally the external authentication provider is the Identity Provider server of the organisation which the researcher is affiliated with (known as their Home organisation). This can be achieved by the eduGAIN interfederation service (www.edugain.org), a GEANT-operated trusted service that mediates information on the endpoints and configurations of the SAML Identity Providers (see section 4.1) in research and education globally. A benefit of using a researcher's Home organisation for login is that they can also deliver authoritative information on the researcher's affiliation (see section 3.1. on the User Home organisation attribute below).

There are also users, often referred to as citizen scientists, who are not affiliated with a Home organisation. It is also possible to belong to a Home organisation that does not provide an Identity Provider service that is participating in eduGAIN or similar identity federation service. To serve also those users, the Life Science AAI provides a Hostel Identity Provider (HIP). The HIP service allows registration for the creation of an account that can be subsequently used in the relying service authentication process. The Life Science AAI also supports authentication based on commercial (such as Google) or community (such as ORCID) authentication providers.

A user can link multiple external authentication providers to their Life Science ID. This is done by first logging in using an existing account that is already linked to the Life Science ID and, subsequently, logging in using the new external authentication provider. These two external accounts are then linked to the user's Life Science ID. After a successful linking process both identities can be used to authenticate on relying services (provided they have sufficient assurance level - see the next section). The Life Science AAI provides a management console for users to view and manage their linked accounts via the external authentication providers.

2.3. Assurance and Step-up authentication

As described in the previous section, users utilise external authentication providers for registering a Life Science ID and authenticating to the relying services. Therefore, the trust framework for the Life Science ID is based on the assurance associated to the identity received from the external authentication provider. The exact assurance model of the Life Science AAI is still in an early phase of evolution.

The X.1252 standard [X.1252] splits assurance into two components. *Identity assurance* is the degree of confidence

[1] There are re-identification attacks where a user can extract information on samples by mounting several consecutive queries to the database [e.g. HOME08, CRAI11]. A user's ability to have several parallel IDs may complicate detection of those attack patterns.

in the identity validation and verification of a user and that they are actually controlling the related credentials. This covers e.g. the identity proofing of new users and how the credentials are delivered to them. *Authentication assurance* is the degree of confidence that the communication partner is the person the identity belongs to. This covers the quality of the authentication, such as, if authentication is based on one factor (password) or several independent factors. The emerging assurance frameworks (such as, NIST SP 800-63-3 [GRAS17] and REFEDS [REFE18]) have adopted this approach by representing the assurance using two (or more) components.

A researcher using their Home organisation account for login has the potential for a high level of identity assurance. Universities and research institutions normally know their users and issue them user accounts and credentials in a reliable way. Unfortunately, there is no well-established method for the Home organisations in the eduGAIN community to signal the user's assurance level to the relying parties. In the future, it is expected that the researcher's Home organisation could make use of the emerging frameworks, such as the REFEDS Assurance Framework [REFE18].

The commercial providers, such as Google, normally rely on their users to self-register an account and self-assert the attributes associated to the identity. Also, community services such as ORCID or the Hostel Identity Provider (see the previous section) rely on the user's self-assertion - although their Home organisation can elevate the assurance over time.

Although some universities are rolling out multi-factor authentication for their users, most Home organisations in eduGAIN still authenticate their researchers with passwords, providing only limited authentication assurance. To facilitate relying services whose security requirements expect multi-factor authentication (MFA), the Life Science AAI needs to be prepared to issue users tokens for MFA.

There are several approaches to implement the MFA including smart cards, public-key infrastructure, USB tokens, etc. Following the proliferation of smartphones, an approach making use of an application running in the researcher's smartphone is gaining popularity. The ELIXIR research infrastructure has investigated an approach where the authentication carried out by the external authentication provider is followed by a subsequent second authentication done by ELIXIR. The step-up authentication makes use of a smartphone app generating one-time passwords following the Time-based OTP standard [RFC6238].

Regardless of how the MFA is implemented, the Life Science AAI needs to define and deploy a procedure to guarantee identity assurance for the MFA service. This covers how the MFA token is associated to the correct Life Science ID, and how and to what extent the user identity is verified. How exactly this is done is still to be decided. In a classic approach a user needs to present their government-issued photo-ID face-to-face at a registration desk, which is challenging in an environment where the users of life sciences research infrastructures are coming from hundreds of different, geographically dispersed Home organisations. Alternative light-weight approaches have also been proposed, such as just verifying the user controls the email address or the ORCID identifier to which the MFA token is associated.

3. Requirements on attributes and authorisation

To serve the needs of the relying services, the Life Science AAI attaches various user attributes to the Life Science IDs, such as the user's name, e-mail address and other contacts.

Some of the attributes focus on assisting the relying services in managing the user's access rights, such as the researcher's Home organisation. This section presents some of the key attributes, focusing in particular to those useful for authorisation.

3.1. User's Home organisation

A researcher's Home organisation is the university, research institution or other organisation with which the researcher is affiliated. In many circumstances, information on the researcher's Home organisation influences directly the services that they can access. The Home organisation can, for instance, be a member of a consortium, project or research infrastructure whose services are for their members exclusively. Furthermore, many services want to revoke a researcher's access rights if they move from their Home organisation. Those services can observe changes in the Home organisation attribute.

The Life Science AAI manages a (multi-valued) attribute for users indicating their current Home organisation(s). Three values, derived from the eduPerson [MACE16] specification, are defined for the attribute: `faculty` indicates the person is a researcher in their Home organisation; `member` indicates other kind of persons with a full set of basic privileges, such as a staff member or a student; `affiliate` indicates the person has some other definable affiliation which does not qualify for being a member. While the exact criteria of the three values are left to the Home organisations to define, these three values are helpful when coarse-grained authorisation is needed.

Optimally, the Home organisation attribute values are retrieved from the Home organisation any time the user logs in using their Home organisation Identity Provider (see section 2.2. on authentication and account linking above). To serve Home organisations that do not have a compatible Identity Provider server, a dedicated person in the Home organisation can be given rights to elevate users to this affiliation. However, setting the attribute values programmatically is preferred, because the value expires after a defined period, after which it needs to be set again. The `affiliate` value can be also given automatically if the person demonstrates they control an e-mail account within the organisation's DNS domain.

In some circumstances, a researcher's affiliation with a research infrastructure may give them permissions to access certain resources. Life Science AAI defines an attribute indicating a person's affiliation with a research infrastructure. How a researcher becomes affiliated with a research infrastructure is left to each research infrastructure to decide.

3.2. User groups

A researcher's permissions to access resources can often be represented using the *group* concept. Researchers belong to research groups that can be collectively granted a resource quota, such as certain CPU or storage capacity, in a computing environment. Also many kind of other projects and activity can be abstracted as a collection of persons working together and requiring access to some services and data.

In the Life Science AAI, users can belong to one or several groups. Each group can have one or several managers who can add or invite new users to the group, remove users from the group and manage other group properties. Groups can have smaller subgroups and a user's group memberships can be expressed to relying services as a multi-valued hierarchical attribute.

3.3. Tiered access to data

Reuse of research data for secondary purposes is being discussed widely in the research sector. When a research project makes an investment to gather data for its research goals, how and under which conditions can the data be made available to other projects?

The Global Alliance for Genomics and Health (GA4GH, www.ga4gh.org) is a policy-framing and technical standards-setting organisation, seeking to enable responsible genomic data sharing within a human rights framework. The GA4GH's approach is referred to as a tiered access to sensitive human data where non-public data belongs to either a controlled or registered access tier.

3.3.1. Controlled access tier

Controlled access tier is the classical data access model, referred to as an access control matrix in computer security terminology. A researcher applies for access rights to the datasets of interest for themselves and their research group members. The application, supplemented with a research plan and the applicants' commitments to the dataset's terms of use, is submitted to the dataset owner (typically represented by a Data Access Committee or DAC) for review. When approved by the DAC, the researcher can download the dataset or access it in a computing environment where a copy already exists.

The terms of data use and exact practices to review the applications are decided by the dataset owners. The Life Science AAI supports the DACs work by providing an electronic tool that manages the application process [LIND13]. The tool also provides the necessary audit and reporting functionality for the dataset owners to demonstrate compliance with their legal obligations, such as the General data protection regulation (GDPR).

3.3.2. Registered access tier

The downside of controlled access is that the process is relatively work-intensive both for the researcher and the DAC. For a researcher the dialogue with the DAC means a long lead time before they can actually start accessing the data. While the application process is generally justified for the more sensitive datasets, a more light-weight process called *registered access* has been proposed for accessing datasets with limited sensitivity [DYKE16].

For registered access, the person needs to demonstrate they are *bona fide* researchers. The exact process for this is still under discussion in the community; for the initial implementation it can be done using one of the following alternatives:

1. The person's Home organisation confirms they are a researcher.
2. A person who satisfies (1) vouches for their status as a bona fide researcher.
3. The person demonstrates they have publications in recognised scientific journals.

The qualifications are further supplemented by attestations a person has to make (for instance, "I refrain from trying to re-identify individuals from the data") in order to qualify as a bona fide researcher, as proposed by [DYKE16]. In practice, the authenticated Life Science user needs to, e.g. click a button to indicate their commitment to the attestations, which is recorded for audit trail.

Once the Life Science user has passed the process, the Life Science AAI adds an extra attribute to their ID indicating their status as a bona fide researcher. The attribute can then be released to relying services which want to enforce user access rights based on the bona fide role. In computer security terminology, this matches the popular Role-based access control (RBAC) paradigm.

3.4. Active role selection

In the requirements of the Life Science AAI, a design choice has been made that a user is supposed to have only one Life Science ID, despite possibly being affiliated with several Home organisations, projects, groups etc simultaneously. These are indicated in the user's Life Science ID as multi-valued attributes; for instance the Home organisation attribute can indicate `faculty` status in a university and `member` status in a research hospital.

In some cases the user is authorised to access services or data only within a particular role; for example access to a given dataset is linked to a Home organisation. Once the user departs from their Home organisation access to the data must also be terminated. Thus, user permissions to access data must be coupled to a particular Home Organisation attribute value, and each time these permissions are used, the environment enforcing access rights must check that the Home organisation affiliation still justifies the data access. This approach is typical for controlled access datasets (see section 3.3.1. on controlled access above).

Resource access policies may also require dynamic segregation of duties. For instance, a user has been granted permission to access dataset A when working as part of project X, and dataset B when working for project Y. However, the user must not be able to access both of these sensitive datasets simultaneously in the provided computing environment in a way that might allow them to correlate these two datasets, potentially risking a breach of the sample donor's privacy. To enforce dynamic segregation of duties, the user needs to indicate at the beginning of the session which of the two projects they are going to work on, and the computing environment will enforce their access rights accordingly.

The Life Science AAI has a service component for active role selection. The relying services can subscribe a service from the Life Science AAI, where, after authenticating the user but before returning the user to the relying service, the user is presented with a list of their roles (the values of the multi-valued attributes) and they need to select the one they are going to use during the current session. Only the selected value is then released to the relying service, which can use it for access control enforcement.

4. Integration of relying services

The previous sections described how the life science users are authenticated and how the Life Science AAI manages their attributes for authorisation. This section briefly describes the technical interfaces for relying services.

There can be a large variety of relying services using the Life Science AAI for authenticating and authorising users. Examples of simple services are collaborative tools like wikis, intranets and mailing list management systems. More complex services include, for instance, research instruments (such as microscopes) used for producing research data, active data management services, archives for storing and preserving the data, data transfer services (to move the data to the environment where they will be processed) and computing environments (such as private clouds or computing clusters) where the researchers run their pipelines and workflow systems for processing and analysing the data.

4.1. Web based services

The Life Science AAI provides two protocols for integration to relying services on the world-wide web: Security Assertion Mark-up Language 2.0 (SAML) and OpenID Connect. Since its publication in 2005, SAML has a strong deployment base and the relying service can select from a large variety of commercial or open source implementations. In SAML, the Life Science AAI acts as a single SAML Identity Provider, to which the relying service redirects the user for authentication and which then releases the user attributes (see section 3 above) to the relying service. All communication is done using the user's web browser; integrity of the SAML messaging is guaranteed by unique reference numbers (nonces) and XML signatures while confidentiality is assured by XML and transport layer encryption.

OpenID Connect, published in 2014, together with its underlying OAuth2 protocol, makes use of newer

technologies (such as, JSON and REST) and can span beyond the limits of the SAML protocol, for instance to support 3-tier scenarios (another service acting on behalf of the user) and non-browser based scenarios (such as, access from mobile apps or command line), which have recently become very popular among life science users. It is also widely deployed by (e.g. commercial) service providers across the internet. In OpenID Connect, the Life Science AAI exposes an OpenID Connect Identity Provider, to which the relying services can integrate for requesting user authentication and attributes.

4.2. Certificates

Although the flexibility of OpenID Connect provides new opportunities for non-browser based scenarios, there are still services (especially those found in grid computing, such as gridFTP) that require users to have a private key and an associated public key certificate. To cater to users who have limited understanding of certificates, new approaches have been proposed where the key pairs are never exposed to the user but stored in a portal where the user can use them seamlessly. Users can log in to the portal using their Life Science ID to access the services where the certificates are required.

4.3. Provisioning and deprovisioning

SAML and OpenID Connect as described above require activity from the user; the user needs to use their browser to make the attribute release take place from the Life Science AAI to the relying service. On the other hand, there are services where the user does not log in separately (for instance, when a user becomes a group member and a mailing list service subscribes them automatically to an associated mailing list) or where a change in a user's attributes needs to be reflected promptly (for instance, deleting a group/project in the Life Science AAI triggers the shutting down of all virtual machines allocated to it in an IaaS cloud).

Relying services can subscribe to the provisioning/deprovisioning service of the Life Science AAI, which provides a push or pull mechanism for relying services to create (provision) new users or remove (deprovision) expired users automatically in the service, based on the criteria configured for the service.

5. Non-technical considerations

The previous sections introduced the requirements of the Life Science AAI as captured in the requirements specification developed in CORBEL WP5. A pilot based on these requirements started in November 2017 with the e-infrastructures (EGI, EUDAT, GEANT) to study their applicability in practice.

The research infrastructures in the Life Science community have applied for funding for deploying the Life Science AAI as a production service. There are several open issues that have been identified but whose solution has not yet been resolved until the deployment process starts. This section provides a short overview of some of these key issues.

5.1. Policies for use

The Life Science AAI intends to serve user authentication and access control for the life science services. Future projects must define policy for accepting Life Science AAI relying services, the terms and conditions of the AAI services and the approval procedure that should be applied for these services. For instance, could commercial services and public cloud providers rely on the Life Science AAI although their customers include also users outside of the life sciences?

It is assumed that the users of the Life Science AAI must commit to certain Acceptable Usage Policy (AUP) as part of the Life Science ID registration process. The AUP would constitute the baseline for all use and could be supplemented by the specific terms of use of the individual relying services. The AUP will be developed during the deployment phase together with other infrastructures and as part of the AARC2 project [AARC18].

5.2. Service management and sustainability

Life Science AAI is expected to become a critical component for the relying services. Therefore, the sustainability of the service needs to be carefully planned, including how the operations and development of the service is organised and funded after the end of the deployment phase. This requires also developing a management structure, including the necessary bodies and voting procedures for decision making.

5.3. Service operations and data controller model

Operating an AAI is not the core competency for the life science research infrastructures, who would like to see e-infrastructures operating the Life Science AAI for the life science community. E-infrastructures, such as EGI, EUDAT and GEANT, have been working on AAI for more than a decade and serving the research communities

even longer. It remains to be seen in the deployment phase, how and under what conditions the e-infrastructures can operate the Life Science AAI.

Related to the operational model, a model required by the data protection laws (including the General data protection regulation) must be designed for the Life Science AAI personal data processing. Who will assume the role as the data controller of the Life Science AAI? Will the e-infrastructures become potential data processors, operating the Life Science AAI on behalf of the Life Science community? Identifying the controller makes it possible to start working on the responsibilities that follow, including the purpose and legal grounds of processing, organisational and technical measures to protect the data, etc.

6. Related work

The work on the Life Science AAI has been inspired by and leveraged previous and on-going work done in user authentication and authorisation by a number of individual research infrastructures, such as ELIXIR (www.elixir-europe.org), INSTRUCT-ERIC (www.structuralbiology.eu) and BBMRI-ERIC (www.bbmri-eric.eu). However, the authors believe that this is the first time several research infrastructures have proposed a specification for a common AAI.

The research and e-infrastructures have been working together in the AARC/AARC2 project to align their approaches and solutions for AAI. One of the key results of the project has been a Blueprint Architecture (BPA) for an infrastructure AAI [AARC17]. The requirements presented in this paper can be implemented with an architecture following the BPA, and since November 2017, the AARC2 project has run a pilot project together with the e-infrastructures to implement the requirements presented in this paper.

7. Conclusions

In this paper we have presented the key requirements of the Life Science AAI as a common service for life science service providers to authenticate researchers and assist the relying service providers in deciding what the researchers are permitted to do in the service. We have focused on the requirements on identifying and authenticating the researchers and managing their attributes for authorisation in the relying services. We have also presented a short overview of the non-technical considerations of deployment.

An architecture based on the requirements presented in this paper is currently being piloted with the e-infrastructures through the AARC2 project and funding has been applied for in order to deploy the Life Science AAI into production. Experience gathered with preceding research infrastructure AAIs helps towards the eventual goal of having a common Life Science AAI for all of the 13 research infrastructures belonging to the ESFRI cluster of health and food. The work has been carried out together and in a dialogue with the other infrastructures participating in the AARC2 project.

We believe that research infrastructures are optimal entities for managing an AAI for the researchers because research infrastructures (unlike research projects) are permanent structures aimed at providing services for the scientific service providers and research collaborations with whom they have direct connections. However, operating an AAI is not the core business of research infrastructures, leaving the door open for research infrastructures to partner with e-infrastructures, who have a long history of providing critical infrastructure for the research sector.

We believe this work also inspires research infrastructures beyond life sciences, as much of this work is not specific to life sciences, and is directly applicable in other disciplines. We hope that this paper further fuels the discussion about whether all research infrastructures actually need to have an AAI of their own, or if the successor of the Life Science AAI could be a wider “Research AAI”, serving also cross-discipline research.

Acknowledgements

The authors wish to acknowledge the ELIXIR EXCELERATE project for funding the development of the ELIXIR AAI that has inspired the work on the Life Science AAI, the CORBEL project for funding the development of the requirements specification for the Life Science AAI and the AARC2 project for funding the pilot project on the Life Science AAI. The projects receive funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements No 676559, 654248, and 730941.

References

AARC17 AARC2 project. AARC Blueprint Architecture. April 2017.

AARC18 AARC2 project. NA3.3 e-Researcher centric policies. Referenced 11 May 2018: https://wiki.geant.org/x/PIArBQ

CRAI11 Craig, D.W., et al., Assessing and managing risk when sharing aggregate genetic variant data. Nature Reviews Genetics, 2011. 12: p. 730.

DYKE16 Dyke, S., Kirby, E., Shabani, M., Thorogood, A., Kato, K., Bartha M. Registered access: a 'Triple-A' approach. European Journal of Human Genetics volume 24, pages 1676–1680 (2016) Available in: https://www.nature.com/articles/ejhg2016115

GRAS17 Grassi, P., Carcia, M., Fenton, J. Digital Identity Guidelines. NIST Special Publication 800-63-3. June 2017.

HOME08 Homer, N., et al., Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet, 2008. 4(8): p. e1000167.

LIND13 Linden, M., Nyrönen, T., Lappalainen, I. Resource Entitlement Management System. Selected papers of the TNC2013 conference.

MACE16 Internet2 Middleware Architecture Committee for Education, Directory Working Group (MACE-Dir). eduPerson Object Class Specification (201602). March 2016. http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201602.html

REFE18 REFEDS Assurance Framework. Draft 2nd May 2018.

RFC6238 Internet Engineering Task Force. TOTP: Time-Based One-Time Password Algorithm. Request for Comments 6238. May 2011.

X.1252 X.1252 ITU-T X.1252 SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY. Cyberspace security – Identity management. Baseline identity management terms and definitions. 04/2010

Vitae

Dr Mikael Linden has worked on Authentication and Authorisation projects at CSC - the Finnish IT Center for Science since 2002. He has consulted several research infrastructures on their AAI needs and is currently the co-chair of the AAI task in ELIXIR, the European research infrastructure for biological data. He received his doctoral degree in information security from Tampere University of Technology in 2009.

Assoc. Prof. Dr. Petr Holub has been working at Masaryk University. Since 2015, he has become Senior IT and Data Protection Manager with BBMRI-ERIC, European Research Infrastructure Consortium for Biobanking and BioMolecular Resources. He is also Chief Information Technology Officer for BBMRI-ERIC Common Service IT. He has background both in sciences and computer science. He was head of the Department of Communication Technologies at Institute of Computers Science, Masaryk University, as well as architect of advanced multimedia and collaborative systems of the Czech national e-infrastructure at CESNET. His research in computer networks and multimedia processing has led to more than 50 research papers in established computer science, bioinformatics and medical informatics journals and ranked conferences. He has received Best Open-Source Software Award by ACM Multimedia SIG. He is a co-founder of Comprimato company, which commercializes research results in acceleration of compression algorithms.

Dr. Ilkka Lappalainen leads the development of CSC research data services, acts as a deputy head of ELIXIR Finland and coordinates the CSC activities within the GA4GH work streams. After receiving his PhD in biochemistry he has worked for over 15 years in understanding human genomics and inherited disorders, first as a researcher in Cambridge, manager of the EMBL-EBI genetic variation archive services including the European Genome-phenome Archive (EGA) and establishing ELIXIR programs supporting human sensitive data management at the ELIXIR-Hub.

Ludek Matyska is a full professor at Masaryk University (MU) as well as a senior researcher at CESNET. Since 2013, he serves as a director of the Institute of Computer Science at Masaryk University. He is also the director of CERIT Scientific Cloud (CERIT-SC), a national research infrastructure dedicated to the promotion and building of national cloud and integrated grid-cloud infrastructures. He works for CESNET since 1998, serving as principal co-investigator of its research programs. His research interests cover security in large scale distributed systems, e-infrastructure (network, computing and data) architecture and policy and generally in cloud and grid systems. He authored or co-authored more than 100 papers and conference contributions.

Dr. Tommi Nyrönen Adj. Prof. is the Head of Node ELIXIR Finland and executive of the ELIXIR Europe Compute platform. The focus of Finnish participation in ELIXIR is in the unique expertise and data resources in genomic research, and provision of IT services and training for analysis of high-quality biological and biomedical information. Nyrönen is also an author of about 50 scientific articles, several patents, and software specifically in structure-based drug discovery and was a co-founder of a bioinformatics start-up FBD Ltd. in 2000. He is adjunct professor in computational drug discovery at the University of Helsinki, and a past by-fellow in bioinformatics at the Churchill college, University of Cambridge.

Dr Michal Prochazka works at CESNET and Masaryk University (Czech Republic) where he got the Ph.D. from applied informatics. He has been involved in IT security and AAI area for more than ten years; he participated in many national and international projects focusing on federated identity and AAI. Now he is responsible for designing and building AAI for two major projects from the life science research community (ELIXIR and BBMRI). He is also part of the Life Science AAI activity run by the AARC2 project.

Dr. Jonathan Tedds is the ELIXIR Compute Platform Coordinator based at the ELIXIR Hub and drives the implementation of the Platform's technical strategy and supports the coordination of the Platform. He is working with partners in ELIXIR Nodes to ensure integration of the compute resources run by ELIXIR Nodes into an effective portfolio across ELIXIR. He is also an Honorary Research Fellow at the University of Leicester where he has developed integrated database solutions for biomedical research informatics through the £2m+ UK funded BRISSKit platform working with academic, National Health Service and industry partners to develop implementations in e.g. the UK 100,000 Genomes Programme, cardiovascular and respiratory disease. He is Editor-in-Chief of Open Health Data journal.

Dr Pasi Kankaanpää is the Administrative Director of Turku BioImaging and Project Manager of the development of the Euro-BioImaging Web Portal since its initiation in 2015. He is actively involved in setting up the Euro-BioImaging ERIC and its various policies, including user and data access, and he coordinates the development of Finnish image data management especially in the Euro-BioImaging and Global BioImaging context. He has been an active participant in the Life Science AAI community since 2016. He has a strong background in cell biology, microscopy and bioimage informatics, and he has previously coordinated the development of the BioImageXD software package, as well as coordinated a large open access core facility (Turku Cell Imaging Core).

Philipp Gormanns is a Senior Developer at the INFRAFRONTIER GmbH in Munich. He is currently involved in the re-engineering of the INFRAFRONTIER IT-components. This task includes development of a UX-improved web-portal, a user-management and AAI solution, data security improvement and human disease integration. He is involved in the Life Science AAI community since its beginnings. In addition to his developer skills he has a strong background in bioinformatic data analysis focused in the fields of human-mouse data integration and disease pathway analysis.

Dr. Michael Raess is a biologist and research manager. He is Head of General Management of the INFRAFRONTIER GmbH, the management and coordination unit of the INFRAFRONTIER Research Infrastructure. INFRAFRONTIER is the European Research Infrastructure for the development, phenotyping, archiving, and distribution of model mammalian genomes. Michael Raess is heading work package 5 in the CORBEL project, which brings together 13 biomedical research infrastructures to develop common user access solutions across research infrastructures. AAI is one focus activity in CORBEL WP5.

Dr Natalie Haley has a masters degree in Physics from the University of Oxford and a doctoral degree in the Life Sciences Interface DTC also from the University of Oxford. She joined the Instruct-ERIC hub in 2017 as part of the ARIA team. ARIA is the online access management software developed and run by Instruct-ERIC. Natalie has been heavily involved in the CORBEL project, including co-leading CORBEL WP5 (user access) on behalf of Instruct-ERIC.