Federated user ID management: a single identity giving access to numerous bioinformatics services

Life Science AAI attempts to make logging in and access to available services as simple as possible.

ELIXIR has created easy-to-use and secure user ID management, enabling access to numerous data collections and services. The service is being developed and expanded in cooperation with other research infrastructures, so that researchers can also have access to biological samples offered by biobanks or compiled from test animals.

Data resources and research instruments needed by researchers are available in numerous research infrastructures. Although international cooperation within research has intensified during the past decade, researchers must nevertheless deal with various administrational processes during the course of their work. Access to material created by research often requires user identification and access permissions. If each data collection requires its own password and user ID, managing them all becomes too cumbersome for individual users. There must be a way out of the password jungle without compromising the data security of the service, or the user's own rights covered by the General Data Protection Regulation. The ELIXIR research infrastructure has the objective of ensuring easy use of data collections without compromising data security.

ELIXIR's AAI service (Authentication and Authorisation Infrastructure) enables electronic user identification and access rights management. Access to data is always granted by the data owner, but AAI will make access to data quicker, since the use policy is clear and straightforward.

The solution is a federated user ID management, which is simple and easy to use. A single identification with their home organisation's login also provide researchers with secure and reliable access to closely protected data collections.

Federations allow researchers to access their home organisations' user IDs. They provide them with access to services outside their own organisations. The idea behind federations is to manage how user IDs are transferred during login across organisational boundaries. Various user rights of different levels can be coupled with the identity in question, to ensure that the user can access the correct resources for their legitimate reasons.

Federated user ID management is by no means a new invention. It has been used successfully by organisations such as the Haka identity federation of the Finnish higher education institutions. The Haka federation

National Institute of Health and Welfare THL biobank is part of BBMRI and linked to ELIXIR Finland. THL uses ELIXIR AAI REMS to manage access application to biobank samples from the Finnish population-level cohorts, and datasets created from the samples. THL was the first sensitive data controller in Europe to federate data access authorisations electronically in collaboration with ELIXIR. Electronic data access entitlements coupled with the reliable identification of users is part of the national strategy of Finland to comply with the General data protection regulation.. Photo: Shutterstock

consists of more than 300 services and has more than 300,000 users.

The EU-funded eduGAIN service, which combines various federations, was established in 2004. In April 2011 it became a permanent service that combines research and education identity federations around the world. eduGAIN brings together more than 50 federations, consisting of 5,000 organisations. It is open to all of the world's academic federations, enabling reliable user login among federation members.

Service available since 2016

The ELIXIR AAI was launched in November 2016. It is part of ELIXIR's Compute Platform, together with cloud and data transfer services. In late 2018, the ELIXIR AAI service had 2,174 users and an average of 3,200 log-ins a month.

In late 2018, researchers who had logged into the ELIXIR AAI service were able to log into another 50 services connected to the ELIXIR infrastructure. Another 44 services were tested, some of which were offered by other major European research infrastructures. One of these was the EGI (European Grid Infrastructure), which was tested to see if it could access the fedCloud computing service. The number of services is rising all the time.

> **"In late 2018, the ELIXIR AAI service had 2,174 users and an average of 3,200 log-ins a month."**

The ELIXIR AAI service was developed by the Finnish and Czech nodes of the ELIXIR infrastructure. The service not only provides authentication services, but also permissions granted by the material owners. In Finland, the National Institute for Health and Welfare (THL) was the first to test the process, based on ELIXIR AAI's federated authentication and permission management, using sensitive material from their biobank samples. THL's biobank is part of the BBMRI infrastructure and access to the material is an example of cooperation between two European research infrastructures aimed at making researchers' lives easier.

The idea is that users register one ELIXIR identity and continue to use it throughout their careers. All they have to do is update their contact and personal details if these change. It is not practical to maintain more than one ELIXIR identity. An ELIXIR identity does not have a password. During registration, all you need is a connection to an academic or commercial user account that is used for logging in.

ELIXIR AAI already accepts Google, LinkedIn or Orcid as part of identification. Through Orcid, researchers obtain a digital identity that enables them to distinguish themselves from colleagues by the same name. ELIXIR AAI supports also 721 institutional logins via eduGAIN.

Different projects aim for common user control

Administrational burden should be reduced whenever possible. That is why ELIXIR uses federated user ID management. It is efficient, safe, reliable and easy to use.

ELIXIR COMPUTE PLATFORM

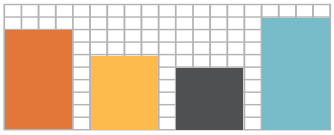
Identity and access management.
Get ELIXIR ID for software, storage and compute.

User data

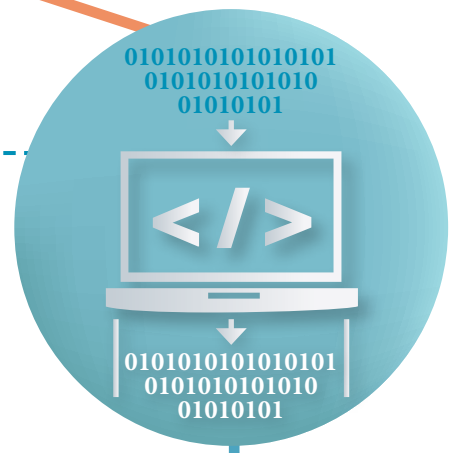
Data resources from ELIXIR Nodes

Human genome data by permission of Data Access Committee

Data transfer and availability.



Data from ENA, UniProt, TaraOceans, Ocean Sampling Day, Biosample DB, EURISCO, EGA



ELIXIR hybrid cloud.
Data and compute services across Europe.

Container orchestration.
The execution of software workflow loads.

Scientific software environment.

elixir

If an applicant has been accepted, ELIXIR AAI uses the OAuth2 protocol to deliver access rights to other reliable services within ELIXIR. Authorisation is performed with the REMS software. Picture: Shutterstock

The challenge for federations lies in the fact that there is no commonly agreed definition for various levels of assurance for identities and authentication. Privacy legislation makes some institutions wary of sharing their researchers' personal data with other jurisdictions.

The requirements for user authentication and the management of user authorisation are tighter when dealing with controlled and sensitive data collections. Users' access rights may have to be categorised, for example. ELIXIR experts are cooperating with other research infrastructures in the EOSC-Life project, which is assessing various user cases in the biological sciences sector to create a common and extensive federated identity management service. This service is called Life Science AAI, which utilises the eduGAIN federation for identification.

Due to the increased need for federated access between research infrastructures, many projects are helping to create common user management. The AARC (Authentication and Authorisation for Research and Collaboration) project was launched in May 2015. The project's second phase (AARC2) was launched in May 2017 and ended in April 2019. The project piloted integrated authentication and authorisation between organisations.

The objective is that each new user will only register a single user ID that follows them throughout their career, even if they change jobs and connections. Because universities and research institutions have connections to several research infrastructures, researchers will have automatic access to them via their own organisation's account. The objective is to create a reference model to manage not only identity (registration, identity proofing) and authentication (logging in), but also other aspects such as researcher status.

Common standards

Cooperation with parties such as Federated Identity Management for Research Collaboration (FIM4R), on the other hand, is aiming for the creation of common standards in order to meet the needs of various research communities. Another key partner is GA4GH.

GA4GH (Global Alliance for Genomics and Health) is an international alliance founded in 2013 consisting of more than 500 bioindustry, healthcare and IT organisations, with the objective of creating standards for genomic data distributed for research use. ELIXIR and GA4GH decided to start a partnership in November 2017. The agreement gives the ELIXIR infrastructure a chance to contribute to the creation of international standards. The agreement is related to the project, the purpose of which is to make the data standards available for clinical patient work by 2022. Now work can be done with over 1,000 organisations not only on standards, but also common principles on how data is processed and distributed.

The challenge is to define the criteria that an organisation must fulfil to become a reliable partner in a global alliance. Registered access ensures the categorisation of various users. It also enables data reuse, but naturally only if consent has been obtained and users adhere to their ethical commitments.

The Global Alliance for Genomics and Health has created three options for accessing human data. These are:

No need to control access Registered access based on the user's role as a researcher Controlled access based on the user's specified access permit.

"Because data is private, federations are controlled by strict data security and regulation, such as the GDPR. ELIXIR's member organisations adhere to European policies in terms of data protection. However, because

research is global, the EU wants to share research data with Canada and the US," says Tommi Nyrönen, Head of Node of Finland's ELIXIR Center. Finland's ELIXIR Center CSC has been building the ELIXIR AAI service alongside the Czech ELIXIR Centre.

"This is why we must manage user information within European and, say, alongside North American organisations. We must have common agreements on how data can be transferred for research purposes in accordance with regulations. Parties responsible for data need sufficient information on users who are requesting access. Only when the user's identity and potential home organisation and status as a researcher have been ascertained, can the data access application be processed. We must also have a mechanism to stop and cancel access to data quickly if it is used for the wrong purposes. This can be done, for example, with a policy and technology specified by ELIXIR AAI.

How to access the service

ELIXIR AAI is a service which researchers can use to request access to sensitive data collections. Users can substantiate their researcher status. To register your researcher status and personal identity in the ELIXIR infrastructure, you must first log in to your home organisation, which will submit your up-to-date user details during the log-in process. The registration may contain additional information, such as the category "bioindustry researcher".

The person in charge of the research project fills in the application form on behalf of the other project participants, and accepts the data collection licence terms. An electronic application form is sent to the Data Access Committee, chosen by the Data Manager, that supervises data access rights. Access is either granted or rejected on the basis of the information on the application form.

If the service requires multi-factor authentication, the user is redirected to a speedy and more efficient identification service, which performs an extra step-up authentication by means of another security factor. Step-up authentication is based on a Time-based One-Time Password (TOTP) and a smartphone application that is registered in the ELIXIR AAI service. Once a user has registered, the TOTP application provides a six-character one-time password, which the user must enter in their browser.

The smartphone application is connected to the correct ELIXIR identity using a text message sent to the phone.

Data owners can cancel or review a user's access rights.Organisations that trust ELIXIR can register their own data collections in the authorisation infrastructure and specify the application forms and the related processes.

ELIXIR Beacon

Effective data security is based on a risk analysis of the requirements of the materials and the nature of the service. For example, researchers can be provided with access to data collections, which have a limited impact on privacy issues, through a lighter application process. This means that all researchers have to do is to prove that they are bona fide researchers and adhere to the general commitments concerning registration.

When they had completed an application, researchers could access all data collections and services available for bona fide researchers, with no extra effort. The ELIXIR Beacon service is an example of such an access process. The Beacon protocol defines an open standard. A website that offers such a service is called a beacon. Beacon is a search engine for finding the location, anywhere in the world, of genome material that contains, say, an interesting change in nucleotide such as in a gene sequence that codes protein, with cytosine (C) changing into guanine (G).

"This change may alter the structure of the protein generated by the gene. Sometimes, such changes are harmless, but they may also lead to illness. The connection between genetic modification and rare diseases is being actively studied, and results could be achieved faster by making material available through Beacon," says Tommi Nyrönen.

The standard and technology have been developed by GA4GH's member organisations. Data searches can be performed on the same principles in ELIXIR and the Beacon Network. Searches are federated and the number of data collections is continuously growing.

Ari Turunen

MORE INFORMATION:

Registered access: authorizing data access
European Journal of Human Genetics (26,2018)
https://www.nature.com/articles/s41431-018-0219-y

Common ELIXIR Service for Researcher Authentication and Authorisation
F1000Reserarch (7, 2018)
https://f1000research.com/articles/7-1199/v1.

CSC – IT Center for Science
is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.
http://www.csc.fi
https://research.csc.fi/cloud-computing

ELIXIR
builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.
http://www.elixir-finland.org
http://www.elixir-europe.org

ELIXIR FINLAND
Tel. +358 9 457 2821s ● e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org

ELIXIR HEAD OFFICE
EMBL-European Bioinformatics Institute
www.elixir-europe.org