B2ACCESS for restricted data

The AlpArray case

Results from the exercise by ETH, GFZ, INGV, LMU, ODC and RESIF to distribute federated restricted data

Summary

Considering the distributed nature of the AlpArray project (AlpArray Seismic Network (AASN), virtual network code “\_AlpArray” includes a temporary network Z3 with >250 broadband stations whose waveforms are archived at across 6 EIDA nodes, as well as all broadband stations from permanent networks in the region already available from all EIDA), we selected the AASN to test the AAI workflow designed and implemented for EIDA. During this exercise, all data centres archiving waveforms took part, GFZ was also hosting the “EIDA Authentication Service”, the B2ACCESS instance was hosted at Jülich, and ETH was in charge of the general management of the user group and the invitation of users to register into the AlpArray group by means of the graphical frontend of B2ACCESS. The final test provided a deeper understanding of the details needed to be included in the documentation (for users and operators), as well as the technical adjustments to operate in a decentralized way.

# Objective

The purpose of the document is to describe in detail all the activities performed during this final test of delivering restricted data from the Alparray experiment to the users. This test encompasses a completely distributed setup, as Alparray is the most complex setup for a seismic network in EIDA, not only from the point of view of the data archival, but also the management tasks.

From its conception, the “EIDA Authentication Service” was designed to make use of the discovery service provided by eduGAIN (GEANT initiative), so that a user could login at EIDA by authenticating at his/her home institution, not with EIDA. This was developed and tested during 2017 by GFZ and the results were successful *from the point of view of the authentication*, but the amount of information (attributes) received after the user login were insufficient to apply any authorization policy.

During the second semester of 2017, GFZ worked, in the context of the EUDAT project, in the modification of the “EIDA Authentication Service” in order to couple it to a B2ACCESS instance, working as a proxy to the eduGAIN service. Two main benefits were expected from this change, as compared to using eduGAIN directly:

* Users without an “eduGAINed” account (from an institution not taking part of the program), could link their identity to an account in B2ACCESS.
* B2ACCESS could manage extra attributes apart from the ones provided by eduGAIN. These attributes could be managed by anyone with proper permission (e.g. data centre operator, network PI), allowing a redistribution of the administrative tasks if needed.

The whole workflow including access control list, definition of groups, management of its members, authentication/authorization and data request, was tested by some members of GFZ. The exercise was successful as a first step and, in a second test, GFZ and ETH used accounts and group attributes both local to a B2ACCESS instance for federated authorisation to give a user access to data distributed over multiple web service endpoints. The services and their roles involved in the exercise were as follows:

**B2ACCESS**

To federate authentication of different backends (including eduGAIN) as well as local accounts.

To manage group adherence.

**EIDA Authentication Service**

to offer a web service which checks authentication with B2ACCESS, retrieves group adherence information, and issues certified tokens of limited validity.

**Auth method extension of FDSN Dataselect:**

* to receive & check the token, create time-limited *queryauth* accounts (to be run by each datacentre as an extension of its FDSN Dataselect)

# Summary of the functionalities tested

* Deployment and configuration of EAS
* Link to the B2ACCESS instance at Jülich
* Creation of an EPOS group (and subgroups) to manage Access Control Lists of different data sets
* Creation of new accounts for users locally to B2ACCESS
* Creation of new accounts from user’s home institution authentication via eduGAIN
* Granting permissions to operators/PIs to include users in different groups
* Check proper setup of SeisComP3 servers at data centres to offer a non-standard extension of the Dataselect WS (“auth” method)
* Requesting a token from the EIDA Authentication System
* Access control administration in SeisComP3
* Request of data from one or many data centres using fdsnws\_fetch

In the next sections, we describe all the tasks we performed. In the last part of the document, we present the results of this exercise and the lessons learned from it.

# Deployment and configuration of the EIDA Authentication System

A simple web page was designed[[1]](#footnote-1) to provide a way to request a token (expiring maximum in one month time) to be later presented to the EIDA services. This web page will redirect the user to a B2ACCESS service and, after a successful login, will provide him/her a digitally signed human-readable file with all the attributes present in the user profile.

The web page also includes the possibility to upload a token to check the validity of its signature. This can be useful in case of problems using the token with an EIDA service.

All information needed for a user requesting a token is also present in the web page in the form of Questions and Answers.

# Link to the B2ACCESS instance at Jülich

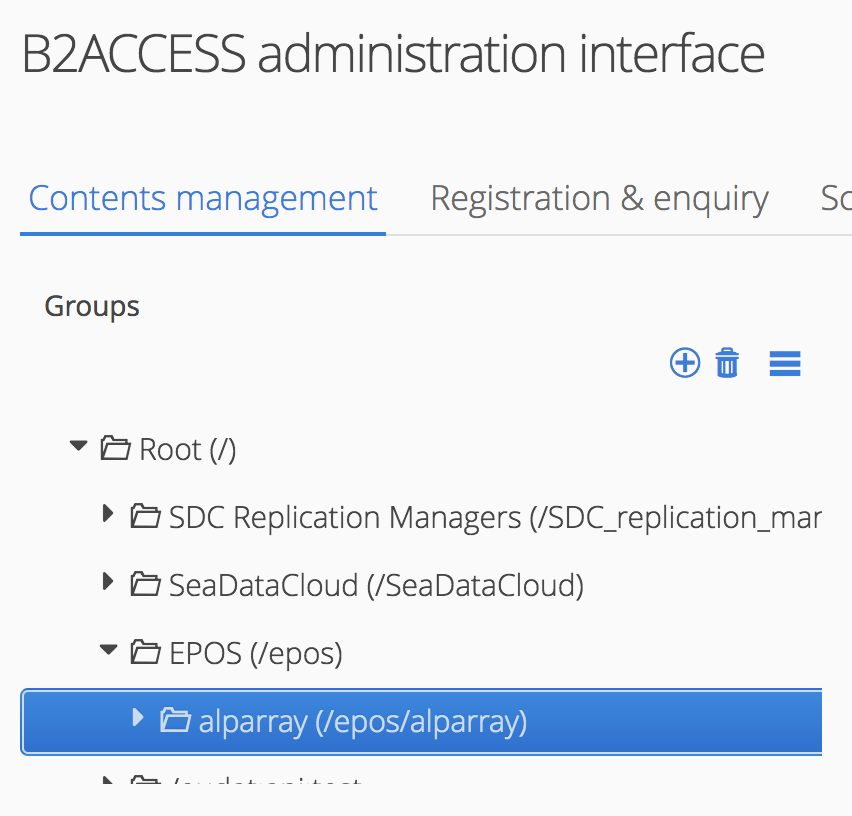
The EIDA Authentication System (EAS) was configured to trust the B2ACCESS instance running in production at the *Forschungszentrum Jülich*, by including the metadata from B2ACCESS in our Shibboleth configuration. From the other side, the B2ACCESS operators at Jülich included our Service ID as a recognized Service Provider. Due to this reason, the service cannot so easily change the URL where it runs. This would also imply a change in the remote configuration of a service running in production.

# The EPOS root and the creation of subgroups for the access management

The B2ACCESS service at Jülich is provided to different communities or groups of users, where EIDA/EPOS is only one of them. In order to keep the organization and the management of the attributes decoupled from the other actors, a group (/EPOS) was created in which all the EPOS related activities and information should take place.

In this EPOS group a small number of trusted persons will have permissions to perform administrative tasks, like create subgroups, include/exclude users in different groups, etc. But the most important action is the possibility to grant permissions to another user who can be manager of a subgroup.

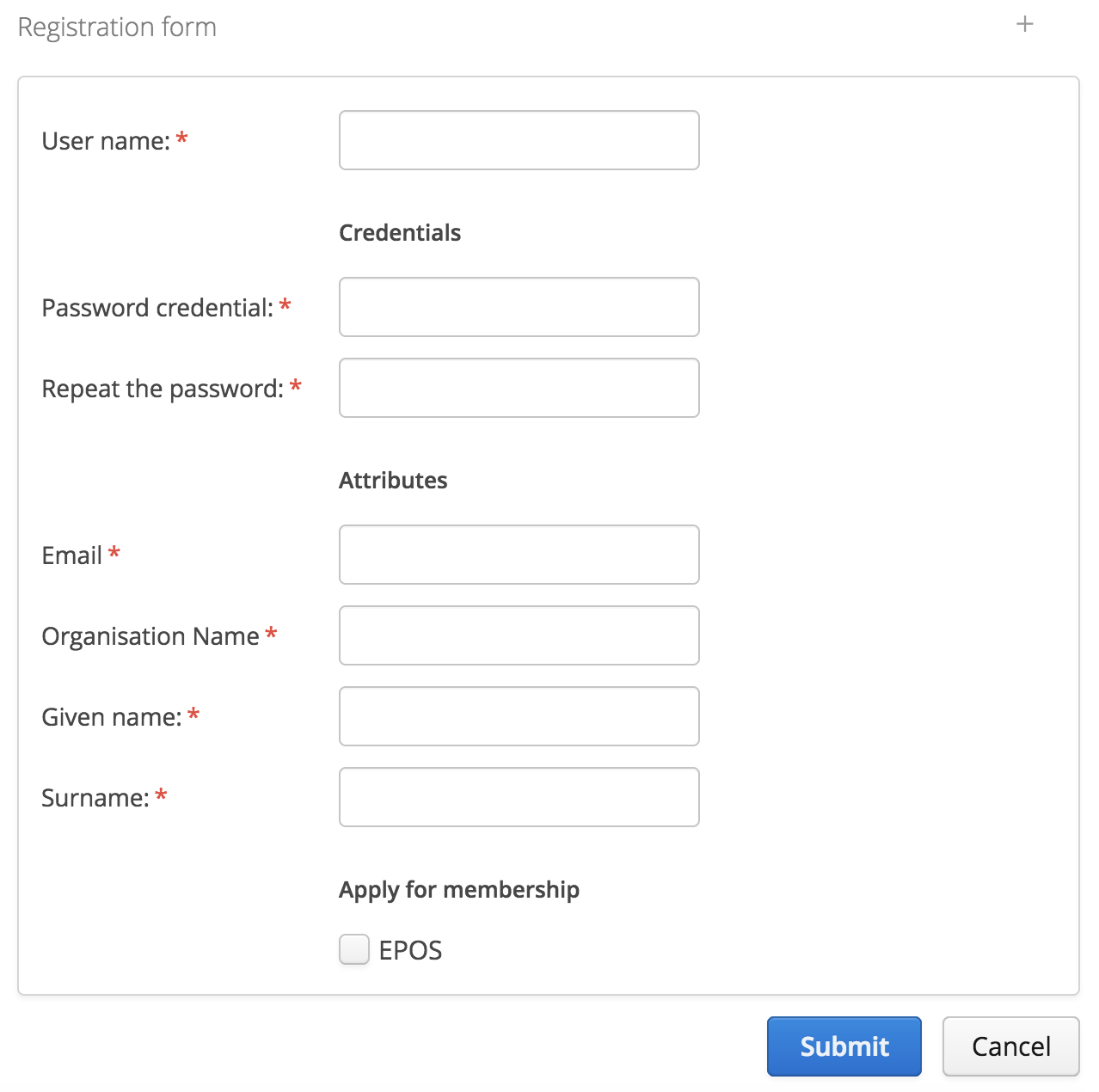
For instance, for the Alparray test, we created a subgroup called “alparray” in which a user from ETH, formally in charge of the administration of the network, was given permissions to manage the users belonging to the “alparray” (sub)group. In this way, the management of the Access Control List can be decentralized and performed by the network operators or someone from the hosting data centre. All options are feasible and can be in each case different.



Groups and subgroups organization for the Alparray test within B2ACCESS.

# Creation of new accounts for users locally to B2ACCESS

Any user can access the B2ACCESS portal[[2]](#footnote-2) and register to have an account in the system. On the upper right corner of the main page, there is a link for this purpose (“*Register a new account*”). There are different technical ways to create this account, being the “*Create B2ACCESS account (username only)*” the recommended one due to its simplicity.



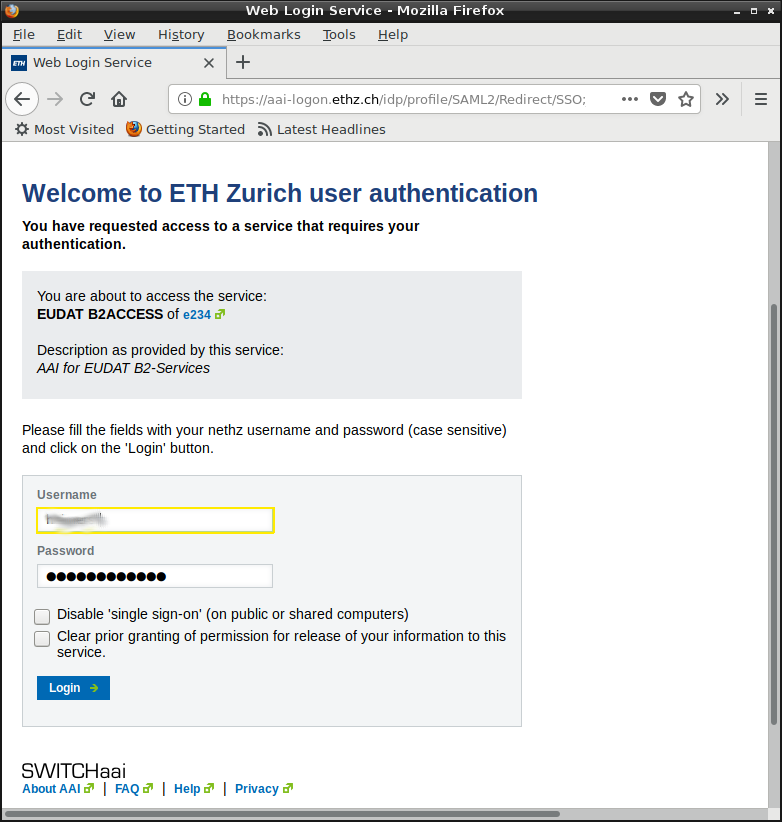
Registration dorm including a checkbox to register as an EPOS user.

The user will need to fill a form with basic contact details (e.g. name, email). In this form, **there is a checkbox available to request being part of EPOS. It is mandatory to check it.** When the form is submitted, the email address will be validated with a message to the user containing a link for the validation.

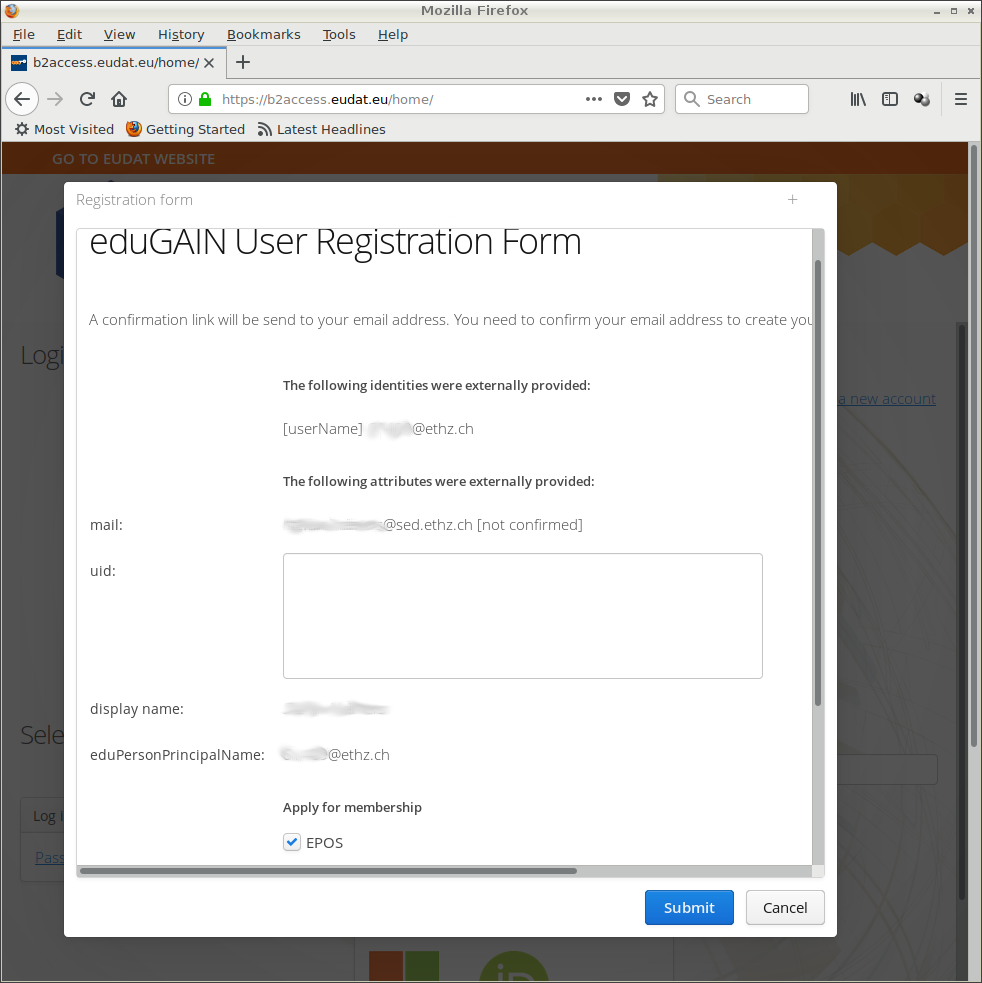
# Creation of new accounts from user’s home institution authentication via eduGAIN

If the user belongs to an institution which is part of the eduGAIN initiative the previous point should be skipped. In order to know whether this is the case, a list of institutions is included in the block “*Login with your institutional ID*” and, due to the length of the list, a search field is also provided to allow for a quick search.

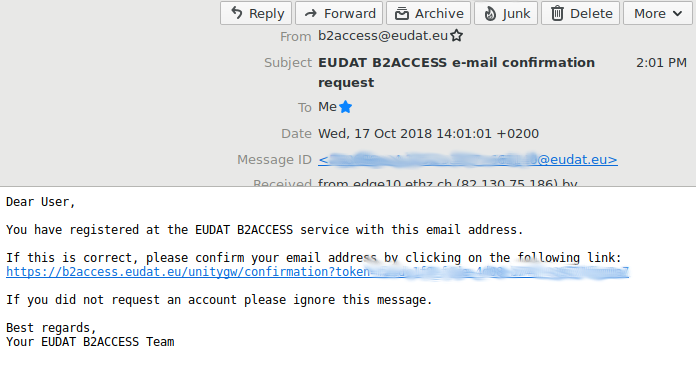
When the user finds its home institution, this can be selected and the user will be redirected to log in there with a message explaining that the login will be used by B2ACCESS for authentication purposes.



Once the user is successfully authenticated at their institution, it is sent back to B2ACCESS to finish the registration, which is usually just the confirmation of the data provided by the home institution. But before sending the information, **there is a checkbox available to request being part of EPOS. It is mandatory to check it.**

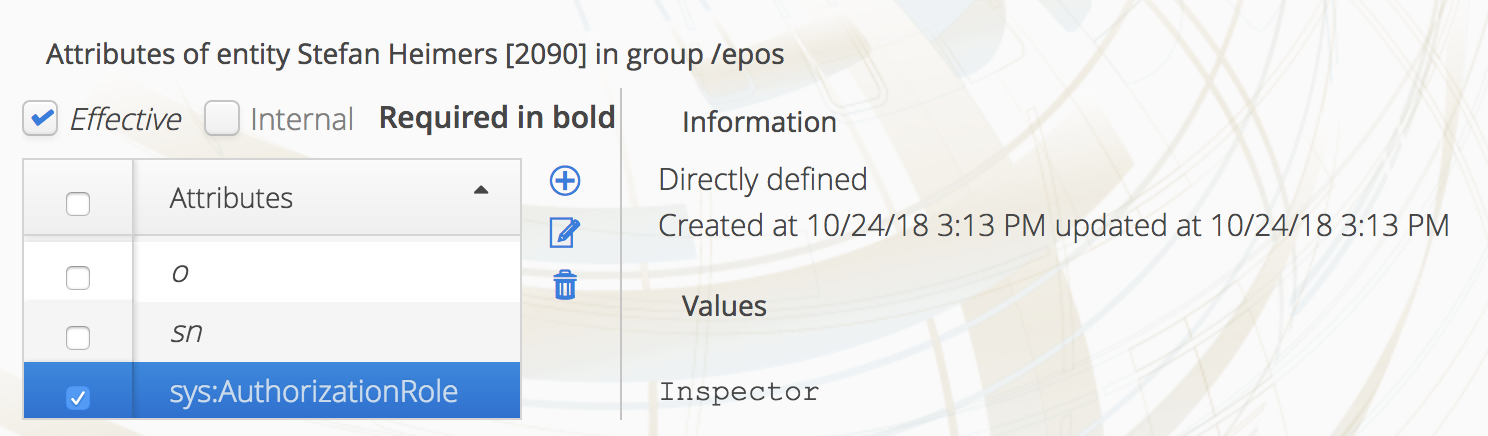


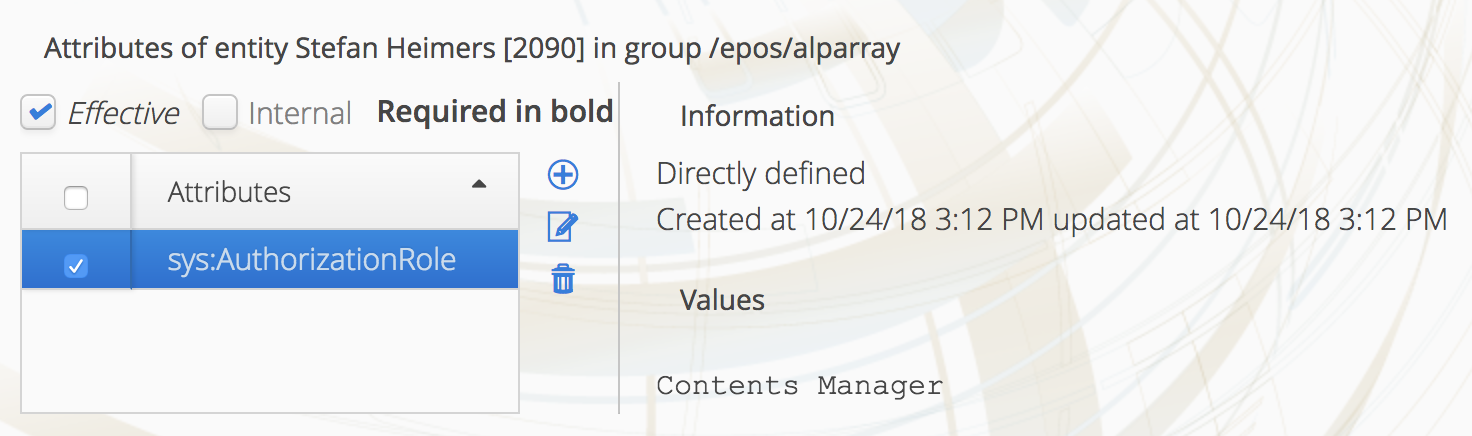
After sending the information a message will be sent to the email address informed to finalize the registration process.



# Grant permissions to PIs to (ex)include users in different groups

For each of the restricted datasets a subgroup can be created where all the users with access to the data can be included. For each of these subgroups a user (or more than one) can be responsible of keeping the Access Control List updated. The user in charge should be given the role “*Content Manager*”, by means of the “*sys:AuthorizationRole*” attribute, in the subgroup created for the dataset. However, as the responsible user should be able to see/list all EPOS users to select the ones to include, also the role “*Inspector*” should be given in the /EPOS folder.





Once the attributes are properly set the user in charge of giving access should be able to select the users from the */EPOS* group and drag-drop them to the corresponding subgroup.

# SeisComP3 and the non-standard “auth” method

The standard way to access restricted data according to the FDSN-WS specifications is using the HTTP Basic Digest authentication. And as this is a standard, it is available at all EIDA data centres. However, this authentication method does not fit well with the construction of a federation. Otherwise, it would be needed to synchronize sensitive data (e.g. usernames/passwords from ACLs) between nodes.

The extra non-standard “auth” method works as a translator between the unified EIDA token and the local username/password at each data centre. The benefit is that these usernames/passwords are local and short-lived. Therefore, they don’t need to be synchronized.

The “auth” method is implemented in SeisComP3 as an extension of the Dataselect method and offered in the same port. This is done on purpose, because SeisComP3 is not meant to be exposed directly, but behind an Apache or Nginx instance. This has to be configured so that *the “auth” method is accessed only through the secure channel HTTPS*, while all other methods are accessed from either port.

# Requesting a token from the EIDA Authentication System

On the user side, once the user has registered, a token to access data (or services) from EIDA can be requested at the following production server URL: <https://geofon.gfz-potsdam.de/eas>

The user has to select how long the token will be valid (e.g. 1 day, 1 week, 1 month) and click on the button to request it. The first time, the user will be redirected to the B2ACCESS login page (from there it could be the home institution login page, in the hypothetical case of an eduGAIN approach), where the user has to login successfully. After the successful login, the token will be automatically downloaded and all the attributes related to him will be included. For the case of the Alparray exercise, the most important attribute is the one called “*memberof*”. In this attribute, it will be specified to which groups the user belongs to (or to which restricted data has access).

Before B2ACCESS releases the attributes to EAS, a detail of the attributes would be present to the user to confirm that these can be sent to EAS. The granularity level even allows that *a user generates a token with only some of the groups. This could be very important in case that a user should provide a token to a third to act on his/her behalf and not to have full permissions on other restricted datasets*.

# Access control administration in SeisComP3[[3]](#footnote-3)

On the SeisComP3 side, the access control list was previously maintained as a list of email addresses, which look like

[mail1@domain.com](mailto:mail1@domain.com), [mail2@domain.org](mailto:mail2@domain.org), [mail3@nowhere.net](mailto:mail3@nowhere.net)

In addition to this, some other possibilities to assign permissions have been introduced in the latest version of SC3. For instance, the access can be given to members of a particular group of B2ACCESS by means of the “group” attribute.

group:/EPOS/alparray

The formal notation expects first a schema or prefix and then a (sub)group (with full path notation), both separated by a “:”. It should be noted that Arclink will ignore any entry in the access control list containing a “:” character, as it will be interpreted as the new notation to be used for the EIDA web services.

Currently, only two prefixes ("mail:" and "group:") are supported. Apart from “group”, which has been already explained, "mail:" can be used to add additional e-mail addresses that are hidden from Arclink.

This means, that an operator could be in charge of maintaining the “alparray” subgroup in B2ACCESS and automatically all data centres could grant data access to all users in this group without the need to maintain a user list, just by adding the proper entry (e.g. “group:/EPOS/alparray”).[[4]](#footnote-4)

Detailed instructions for the deployment of the latest SC3 including these new features are provided at the following URL: <https://dev.knmi.nl/projects/eida/wiki/FDSN_-_dataselect_-_auth> . Read carefully that page! There, you will also find technical details about a normal workflow for data requests and which client tools can be used.

# Users: Testing the whole workflow

Once a user profile has been setup and is included in the *alparray* group in B2ACCESS interface, you will need to follow the following steps to check that the whole workflow is valid:

* Request a token from <http://geofon.gfz-potsdam.de/eas> .
* Check with a normal text editor that the attributes in the token look reasonable. Do not modify it!
* Option 1: Use the *fdsnws\_fetch* client to request data from Z3 stations archived at any data centre. At the moment of this test almost all data centres (see next section) include all the updates to support this workflow. Details about how to install this tool and the syntax to make a request can be seen in <https://dev.knmi.nl/projects/eida/wiki/FDSN_-_dataselect_-_auth>
* Option 2: clients provided by *Obspy* that support the automatic internal routing and token for EIDA (i.e. *RoutingClient*).
* Option 3: Use http://eida.gfz-potsdam.de/webdc3/ and fdsnws authentication, loading token
* If some other tool has to be used, this should be taken into account on the client-side. Routing Service should be contacted to see where data is archived, and the “auth” method transforming a token into a short-live user/password **at each data centre** must be called.

# Status of all EIDA nodes

Regarding the non-standard “auth” method, **all data centres are supporting it and tests have been successful**. One minor but necessary resource is that INGV has not still been able to deploy a real certificate for its secure web server. However, the code has been tested on port 80 and it is working perfectly, but it will not be useful for clients expecting the service to be answering at the HTTPS port.

The authorization based on username was tested also and all data centres are supporting it. For the data centres not hosting restricted data, the queryauth method was tested requesting open data with the user/password provided by the “auth” method.

In the case of the more advanced feature of authorizing users based on attributes other than the email, **RESIF is not supporting it because the IRIS Dataselect code is used. A workaround for this is being evaluated.**

# Results

The shared exercise between all nodes archiving Alparray data validated successfully the complete workflow from the creation of group attributes in B2ACCESS (as a decentralized responsibility of data centres) to their usage by the EIDA Authentication Service, their transport with tokens, and their interpretation by the SeisComP3 implementation of the *auth* extension of FDSN Dataselect.

The feedback from ETH members in the role of B2ACCESS group manager, as well as user who requests data, helped to early detect technical issues and solve them properly. The documentation has been also improved and will be uploaded to the EAS web page to provide the users much clearer instructions about how to retrieve data and make use of EIDA services by means of tokens provided by EAS.

1. <https://geofon.gfz-potsdam.de/eas> [↑](#footnote-ref-1)
2. <https://b2access.eudat.eu/> [↑](#footnote-ref-2)
3. if EIDA and FDSN services are provided by different software implementations (outside SC3), comparable features need to be made available [↑](#footnote-ref-3)
4. As a security measurement, an upgrade is highly recommended to prevent anybody with username “group:/EPOS/alparray” getting Alparray data. [↑](#footnote-ref-4)