Federated authentication: overview and status

J. Quinteros, with contributions from: Rob Casey, Jerry Carter, the EIDA Technical Committee, and many others

Why do we authenticate?

- The first reason is always **because there are restricted data**.
- However, this is important for authorization, not authentication.

- Authentication checks who a person is.
- Authorization checks which specific resources a user has access to.

- To understand usage patterns is fundamental to improve services.
- It could also improve the quality of our statistics.
- It allows our funders to understand the impact of our data and services.

FDSN current standard

- Only the dataselect web service has the option to be used with authentication. EventWS and StationWS don‘t have authentication.
- In dataselect there is a „queryauth“ method to authenticate and submit a data request.
- We have 2 methods to get data: query and queryauth.
- HTTP Digest Authentication (RFC 2617) should be requested from the client.
- Authentication credentials are data center specific.

Limitations

- EventWS and StationWS don‘t understand what a user is.
- To have 2 methods to request data could be confusing.
- Digest Authentication (RFC 2617) and different credentials per data centre seriously limit the scalability of a federation.
- Data centres must keep a list of usernames and passwords.

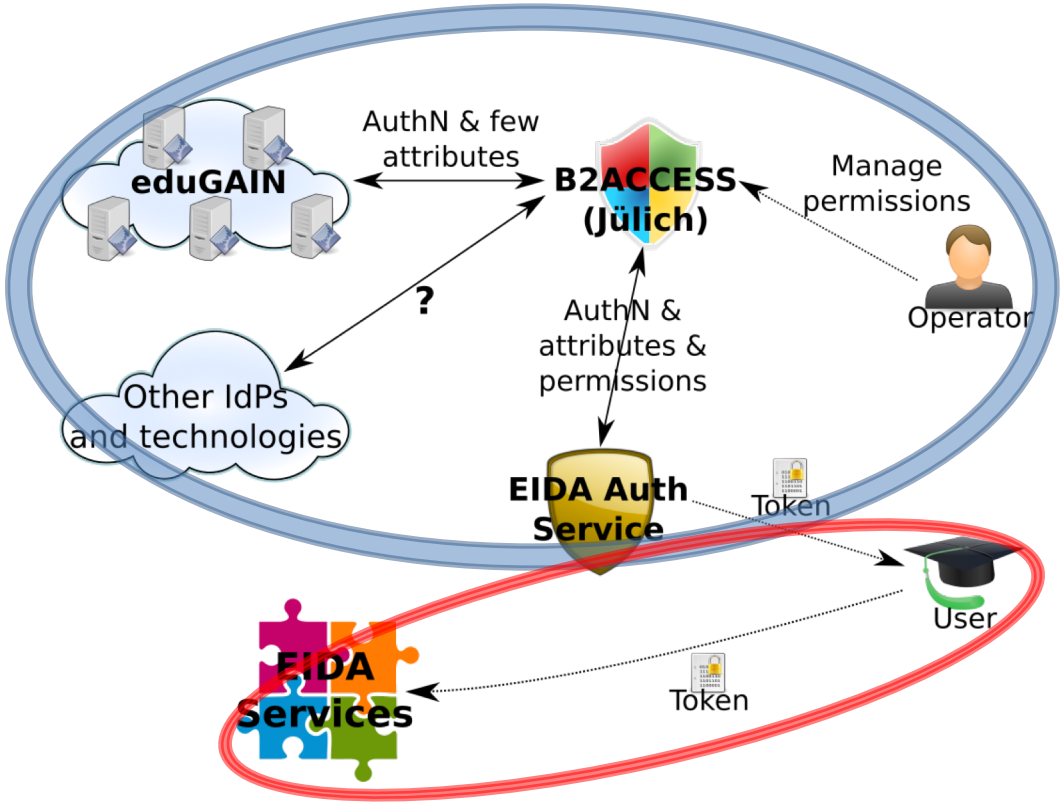
Federated authentication at ORFEUS

Challenges:

- Services supporting open/embargoed data.
- Thousands of users/year around the world. Most of them unknown.
- New regulations on privacy (GDPR).
- Avoid the need to manage sensitive data at the data centre.
- Foster user authentication for open data (better statistics).
- Better understanding of how data is being used.
- How to properly manage a user database?

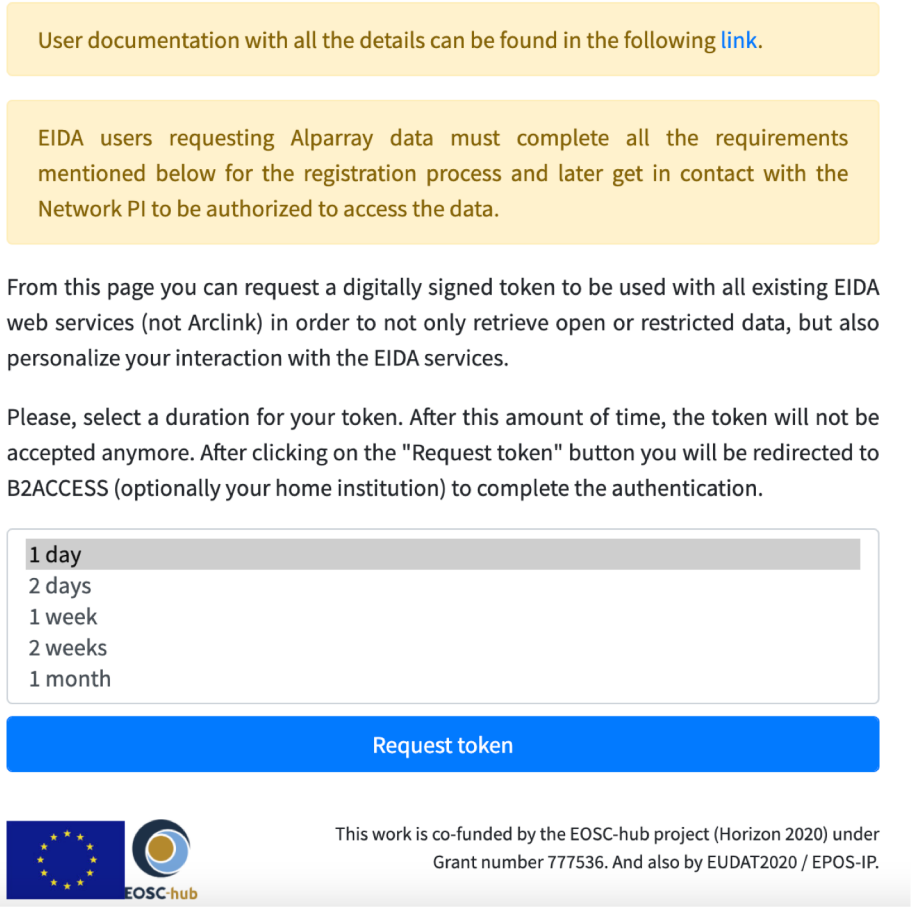
Federated authentication at ORFEUS

- **Approach**: Complete decoupling from **user login** and **service provisioning**.
- User receives a token and presents it to the service providing data.
- The eduGAIN initiative (>8000 institutions) allows users to log in at their home institutions. We don't store any user data!



Get a token easily

- As a user, you can get a token from our web page.
- You will be redirected to your institution.
- Log in there.
- You will receive the token as soon as you authenticate.
- This needs to be stored where different tools (or you) can find it.



EIDA token digitally signed

-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA1

{"valid_until": "2019-05-08T10:21:26.269027Z", "cn": "Javier Quinteros", "memberof": "/epos/alparray;/epos;/", "sn": "Quinteros", "issued": "2019-04-08T10:21:26.269034Z", "mail": "javier@gfz-potsdam.de", "givenName": "Javier", "expiration": "1m"}
-----BEGIN PGP SIGNATURE-----
Version: GnuPG v1

iQEcBAEBAgAGBQJcqyCmAAoJEEFpzp0AlwdXaBQIAL9I7lUriWaoWMDPAnUTLUVEN8XjVN3Kfxa
bokCY3jEfI6hRVLmEO8ofaV9iHktrXqvPaC6Ygp3w6raJi9mtmsS1O61FcUcBS1vG1LdWzcpM==8ke
x
-----END PGP SIGNATURE-----

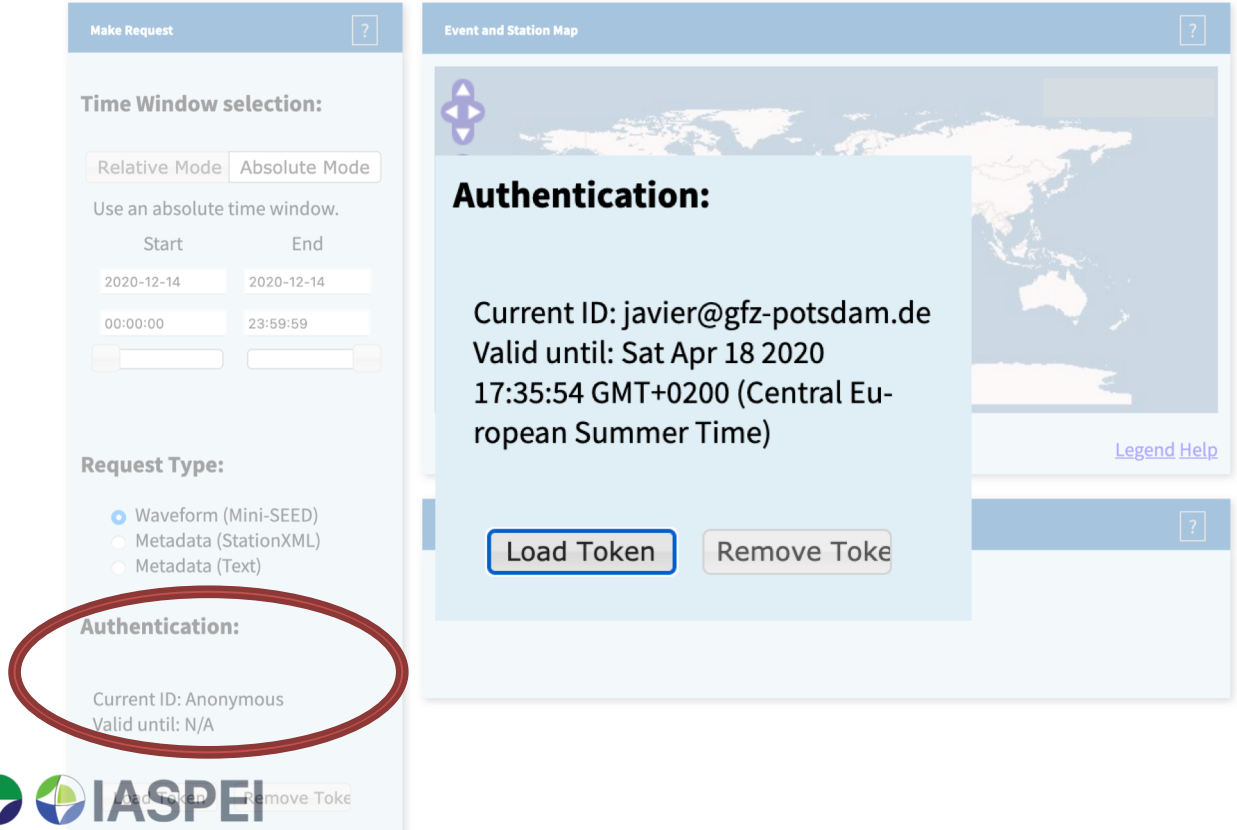
Use of EIDA token with Obspy

```
>>> rsClient = RoutingClient("eida-routing", credentials={'EIDA_TOKEN':
        '/Users/javier/.eidatoken'})

>>> st = rsClient.get_waveforms(network="Z3", channel="HHZ",
        starttime=UTCDateTime(2016, 3, 1), endtime=UTCDateTime(2016, 3, 1,
0, 2, 0))

165 Trace(s) in Stream:
Z3.A022A..HHZ | 2016-03-01 - 2016-03-01T00:02:01 | 100.0 Hz, 12396 samples
(163 other traces)...
Z3.A216A.00.HHZ | 2016-03-01 - 2016-03-01T00:02:00 | 100.0 Hz, 12001 samples
```

Also for WebDC3 (GUI)



Problems with the current standard

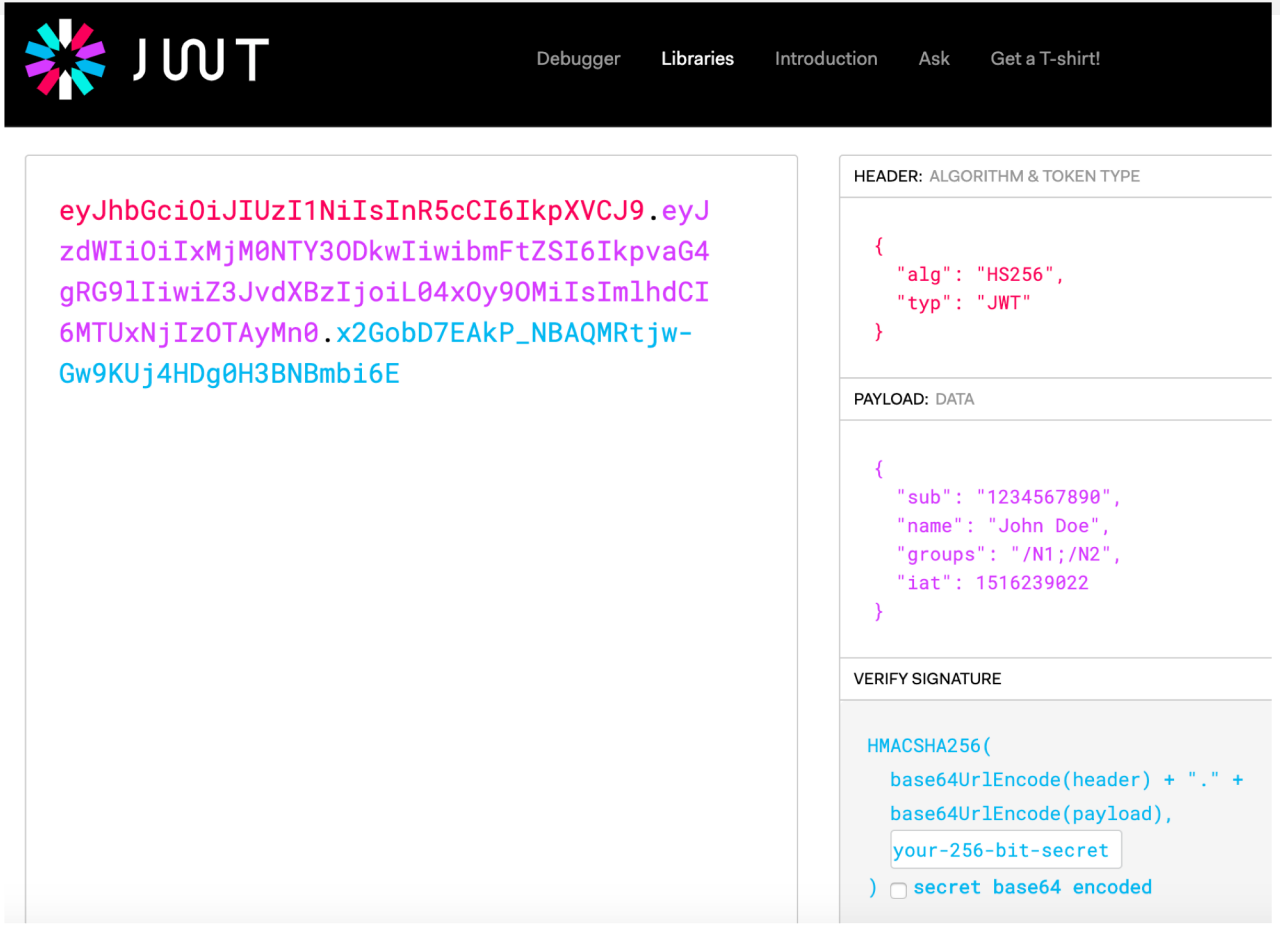
- Our current specification forces us to use digest authentication (username and password).
- To avoid this, we use of a token, which is a „certificate“ to recognize the user without username and password.
- To respect the standard we create temporary username/passwords, which are transparent for the user, and use them with „queryauth“.
- Only clients take care of these steps (e.g. obspy, fdsnwsscripts, pyrocko).
- The client has to do these **two steps to request data**.
- The data centre has no reliable statistics from the user perspective.

On-going activities on the IRIS side

- IRIS (EarthScope) implemented a similar AAI system.
- The basic, philosophical approach is basically the same.
- They implemented their token as a standard JSON Web Token (JWT).
- This one of the formats that everyone is using all the time. Even if we don‘t see them.

JSON Web Token (JWT)

- Tokens can be signed.
- They cannot be edited.
- A token consumer can verify the issuer of the token and that hasn‘t been changed.
- They can also be encrypted (not interesting for us).



Improvements for the community – A roadmap?

- Adopt JWT as our standard format for tokens.
  - IRIS has done it. EIDA could do it transparently soon.

- Discuss a minimum number of fields we must have in the token, taking into account GDPR. Any data centres can always add more fields, if needed.
  - Some are: „issued at“, „expiration“, „email“, „name“, „groups“ (for the future authorization)

- Use the HTTP header to transmit a token to the query method, if needed.

```
Authorization: Bearer <token>
```

Improvements for the community – A roadmap?

- Get rid of the queryauth method. (Finally!)
- Provide a simple, minimalistic way to issue tokens by small data centres.
  - Small data centres do not have the capacity to manage complex IT solutions. Implementation and operational procedures should be very simple.
- Discuss how to trust tokens from other data centres. It could be as easy as exchange a data centre key.
  - Data centres to agree on the algorithm to use and how to exchange the key.

Conclusion

- We have a unique oportunity to make a step forward and simplify our data provision system.
  - Adapt a standard as **JWT** and **avoid the 2 steps**.
  - Simplify in **a unique „query“ method** to request data.
  - **All services** could support authentication to improve statistics and analysis.
  - **No more passwords** for our users.
- This, plus the addition of the FDSN Data Centre Registry, would allow the user to **detach completely from where the data is hosted** (a unique, global seismological data centre).