AAI: Experiences with Making a Service Available for Users Worldwide

Jozef Mišutka, Ondřej Košarko, Amir Kamran, Michal Josífko, Pavel Straňák
LINDAT/CLARIN dev team

Institute of Formal and Applied Linguistics
Charles University in Prague, Czech Republic
The ultimate objective of CLARIN ERIC is to advance research in humanities and social sciences by giving researchers unified single sign-on (SSO) access to a platform which integrates language-based resources and advanced tools at a European level.
LINDAT/CLARIN AAI requirements

• unified SSO
• users are verified (researchers)
• possibility to identify users
• possibility of a project wide authorisation

Repository:
• valid email address
LINDAT/CLARIN AAI

- **shibboleth** federated identity solution
- Security Assertion Markup Language (SAML)
- trust between Service Providers (SPs) and Identity Providers (IdP)
- umbrella model
Web based login

• visit http://lindat.mff.cuni.cz/repository
Web based login

• select your IdP
Web based login

- redirected to your IdP’s login page which should be known to you
Web based login

- enter your credentials and you will be redirected back to our service

http://www.lindat.cz
How to achieve

• **install** shibboleth (or SimpleSAMLphp)
• **join federations** and get metadata feeds of trusted IdPs
  – national federation
  – inter-federation e.g., Service Provider Federation by CLARIN, eduGAIN
  – individually e.g., project wide IdP for “homeless” users
  – setup harvesting of the trusted IdP metadata feeds
• **allow the user to select one of the trusted IdPs** using DiscoJuice library
Consistency between components

- **harvest** metadata feed of IdPs that your SP trusts
- IdPs must trust your SP
- the user should be able to **select** from the list of trusted IdPs
Reality – what can happen

- **IdPs do not trust your SP**
  - only the technical trust is granted
  - either technical problem on their side or they explicitly filter your SP out, opt-out policies

- **IdPs do not send ANY attributes**
  - information that a user exists

- **IdPs send attributes but not** the ones which can be used for identification

- **IdPs send only one of the “id” attributes** (eppn, eptid, persistent-id?)

http://www.lindat.cz
Reality

• users are not from R&E
  – hopefully, they are at least properly tagged
• scoped attributes are filtered
  – do not specify shibmd:Scope
Persuade IdPs that you play nicely

- Data Protection Code of Conduct
- SP is responsible

```xml
<saml:Attribute
    xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion"
    Name="http://macedir.org/entity-category"
    NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:uri">
    <saml:AttributeValue>
        http://www.geant.net/uri/dataprotection-code-of-conduct/v1
    </saml:AttributeValue>
</saml:Attribute>

<mdui:PrivacyStatementURL xml:lang="en">
    https://lindat.mff.cuni.cz/privacypolicy.html
</mdui:PrivacyStatementURL>
```

http://www.lindat.cz
Monitor

- successful / failed login attempts
Monitor

- IdP QA
Monitor

- attributes statistics
Monitor

• unified IdP discovery
Few statistics

• 45 IdPs or national IdPs
  – SPF: 29 (10)
  – eduGAIN: 34 (15)
  – eduID: 14
• eppn: 36 (10)
• targeted id: 30 (4)
• email: 33
• no attributes: 5
Conclusion

- **state-of-the-art** best results for R&E
  - SAML + national federation + SPF + eduGAIN
  - DP-CoC
  - optionally create your own list of trusted IdPs
Thank you