

# CLARIN

Common Language Resources and Technology Infrastructure



# Federation for eHumanities & eSocialSciences

## CLARIN AAI

---

**Dieter Van Uytvanck**

**Max Planck Institute for Psycholinguistics**

**Dieter.VanUytvanck@mpi.nl**

**AAI workshop, Köln**

**2013-10-18**

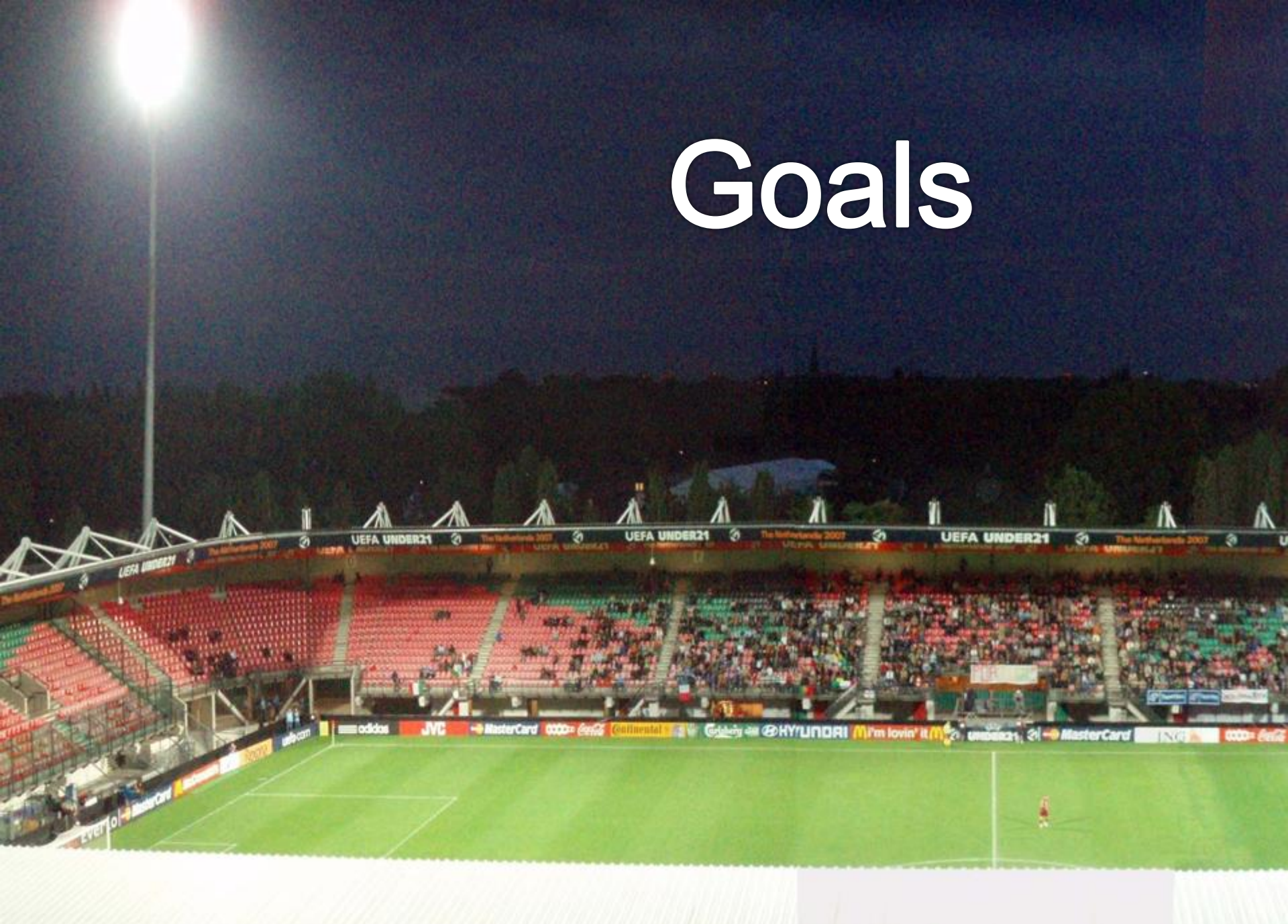
# CLARIN?

---



- **Common Language Resources and Technology Infrastructure**
- aims at providing easy and sustainable access for scholars in the humanities and social sciences
  - to **digital language data** (in written, spoken, video or multimodal form)
  - to **advanced tools** to discover, explore, exploit, annotate, analyse or combine them
- independent of where they are located: a shared **distributed** infrastructure
- More information: [www.clarin.eu](http://www.clarin.eu)

# Goals



# Goals of CLARIN

---



- Making resources and tools available
  - to a broader audience
  - in an easy-to-use and transparent way
- What and how?
  - Data: corpora, lexica, annotations, ...
    - Via your internet browser
    - Or on your local computer
  - Services & software
    - Via a web applications
    - Download and install
    - Via web services (for programmers)

# Goals of CLARIN

---

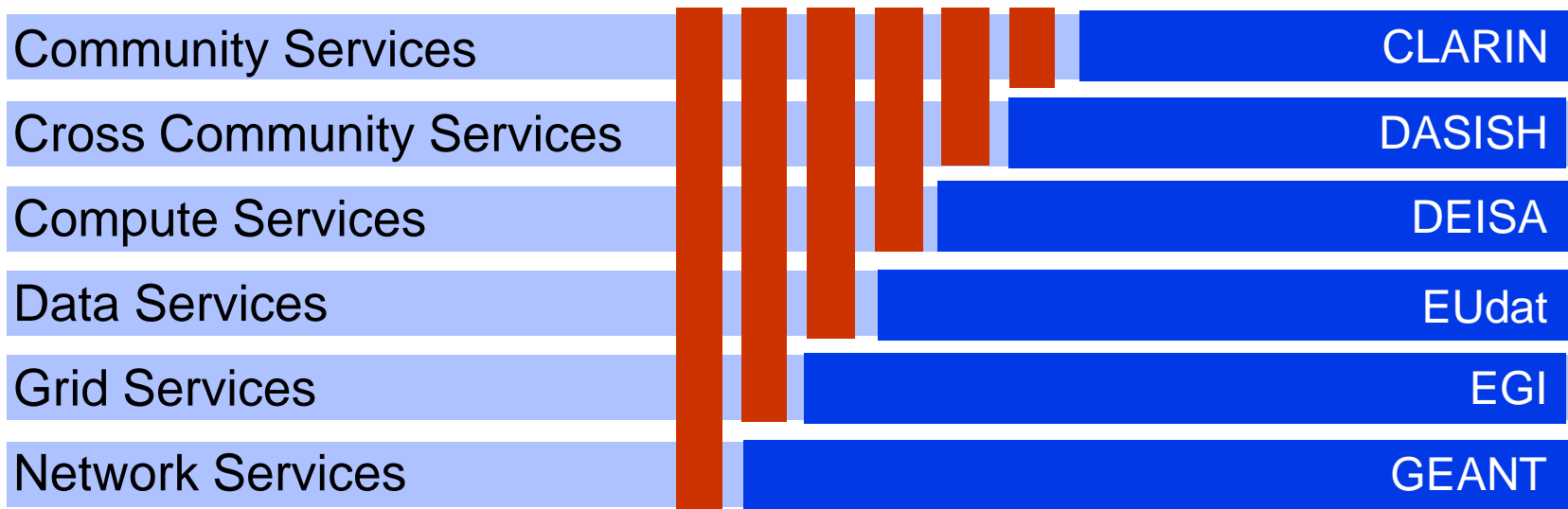


- Enhance interoperability
  - Standards & protocols
  - Semantic interoperability (~ isoCAT)
- Digital sustainability
  - Long-term availability of:
    - Data
    - Services
  - Openness (easy to find resources and to access / share them)



# Context: an ecosystem of infrastructures

---



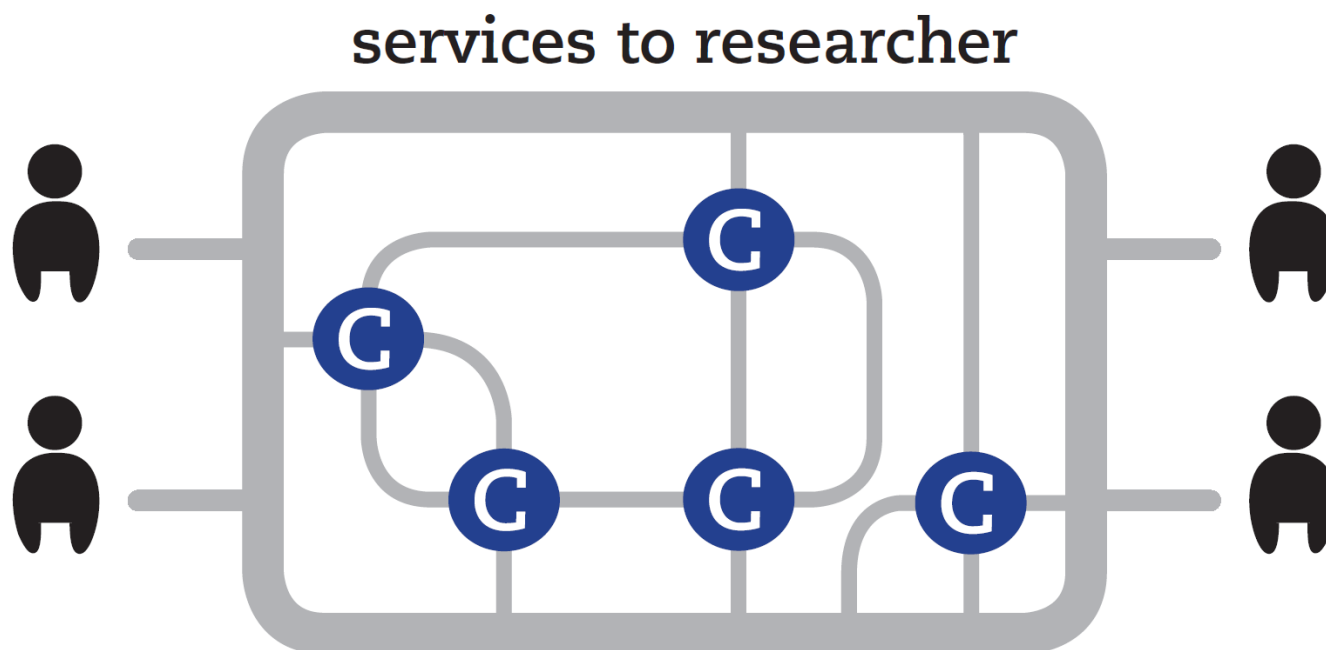
# Centres



# CLARIN centres



- Not one monolithic setup
- But using a well-working paradigm (think of the Internet): distributed architecture





# Centres in CLARIN

---



- Each one with its own specialties
- Extendible (with other institutes/countries)
- Not ad-hoc but within a clear vision
- More information:
  - <http://www.clarin.eu/node/3812>

A large, ancient stone pillar, likely from the Indus Valley Civilization, is the central focus. It is composed of several stacked sections. The upper part features a prominent relief of a seated female figure with a large, ornate headdress, possibly a goddess or a royal figure. To her left, another figure is partially visible. Below these figures are decorative friezes with repeating patterns. The pillar is set in a museum gallery with a white wall and a red wall to the left. A white shelving unit with various artifacts is visible on the right.

# Infrastructure Pillars

# Some important pillars

---



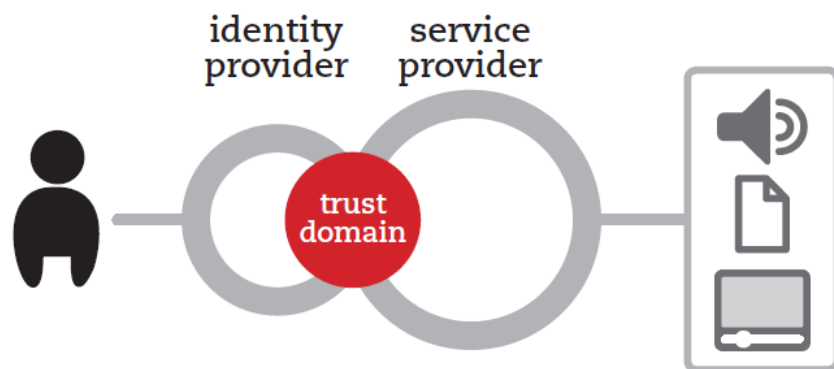
- Sustainable repositories
- Persistent Identifiers
- Web Services (and chaining of ...)
- Federated Content Search
- Registries (for centres, controlled vocabularies, virtual collections, ...)
- Monitoring
- And, of course...

# Federated Identity

---



- Use your own (institutional) login to access password-protected resources
- Requirement for B centers
- Available Services:
  - Service Provider Federation
  - CLARIN Identity Provider
  - Easy-to-use discovery service





I would like to use a CLARIN service...



1. wants to access

7. uses

6. redirects to resource for authorization check



Service Provider

5. User enters credentials

3. User selects IdP

2. redirects to

4. redirects to

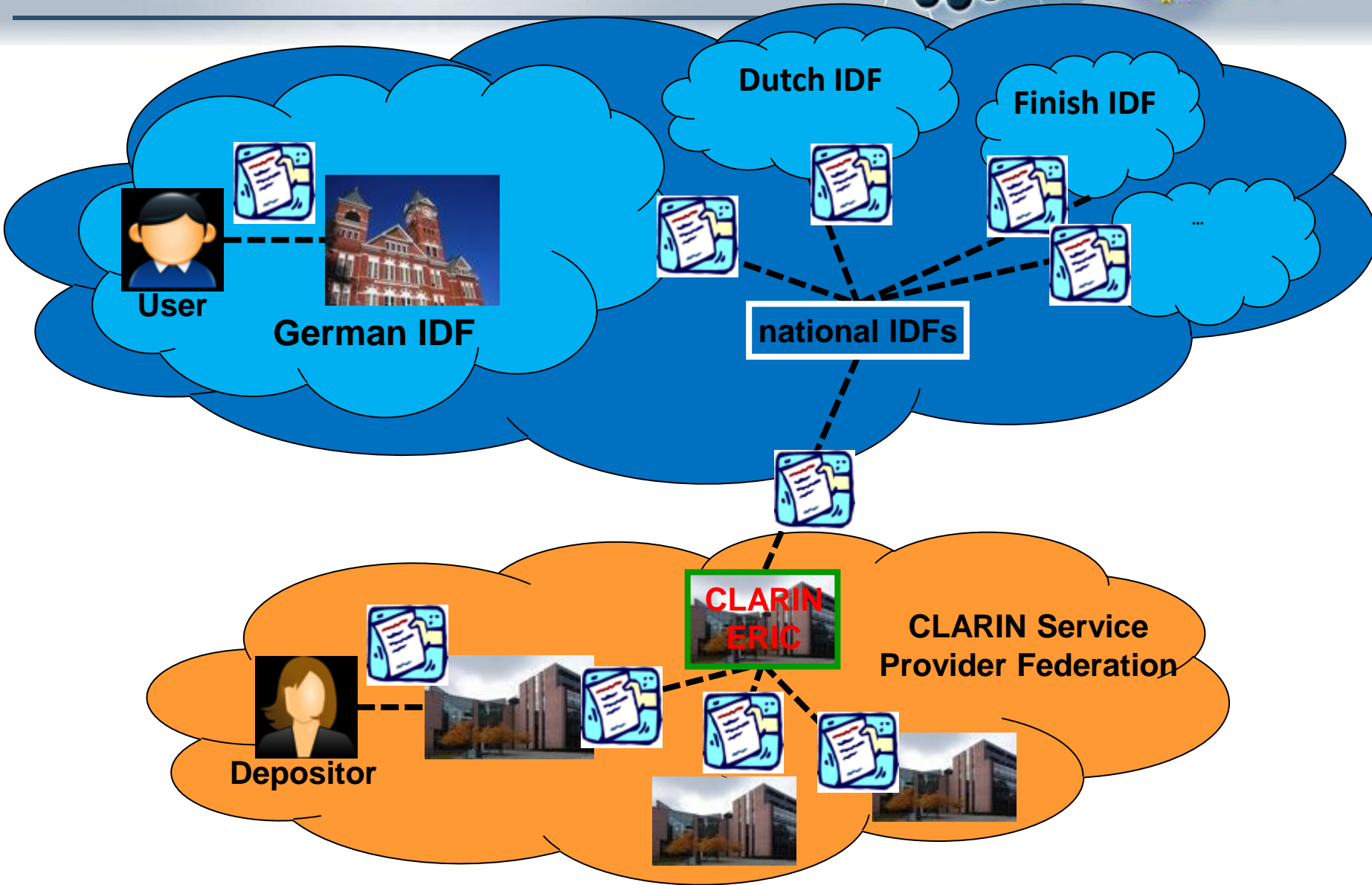


Identity Provider



DiscoJuice  
version 1.0

Discovery Service



# Pillars: Federated Identity

---



- CLARIN's Service Provider Federation (SPF)
  - currently:
    - 11 SPs (number is growing)
    - about 200 IdPs
  - agreement: CLARIN ERIC becomes responsible legal entity
  - further extension of the SPF
    - Connect more CLARIN services to it
    - Connect to more national identity federations (currently 6)
  - details: [www.clarin.eu/spf](http://www.clarin.eu/spf)

SP	agreement	SP status	connection SURFconext	connection DFN	connection HAKA	connection KALMAR
MPI (lux17)	signed	OK	OK	OK	OK	OK
MPI (catalog)	signed	OK	OK	OK	OK	OK
MPI (corpus1)	signed	OK	OK			
INL	signed	OK	OK			
IDS	signed	OK	OK	OK	OK	
BBAW	signed	OK	OK	OK	OK	
CSC	signed	certificate expired				
UTU	signed	OK	OK	OK	OK	OK
UFAL	signed	OK		OK		
ATILF	signed	OK				
ICLTT	signed					
Meertens	signed	OK	OK			
Huygens	signed	OK	OK			



# Current connections of SPF



Country/Organisation	National Identity Federation	Connected to CLARIN SPs
Austria	<a href="http://www.aco.net/federation.html">http://www.aco.net/federation.html</a>	planned
Bulgaria	<b>none</b>	
Czech Republic	<a href="http://eduID.cz">http://eduID.cz</a>	planned
Denmark	<a href="http://www.wayf.dk/">http://www.wayf.dk/</a>	yes (via Kalmar Union)
Dutch Language Union	<a href="http://surffederatie.nl/">http://surffederatie.nl/</a> + <a href="http://federation.belnet.be">http://federation.belnet.be</a>	planned
Estonia	<a href="http://taat.edu.ee">http://taat.edu.ee</a>	planned
Finland	<a href="http://www.csc.fi/english/institutions/haka">http://www.csc.fi/english/institutions/haka</a>	yes
Germany	<a href="http://www.aai.dfn.de/">http://www.aai.dfn.de/</a>	yes
Netherlands	<a href="http://surffederatie.nl/">http://surffederatie.nl/</a>	yes
Norway	Feide	yes (via Kalmar Union)
Poland	<b>none</b>	

# Pillars: Federated Identity

---



- At the same time we are exploring the options to make use of eduGAIN
- Pro:
  - you only need to join your national Identity Federation
  - no need for SPF (less administration)
  - access to non-EU user base

# Pillars: Federated Identity

---



- Con:
  - Opt-in: only institutions that agreed to use eduGAIN can use Services
  - Status at **April 2013** (test with <http://lux17.mpi.nl/ds/aaid/?shhaaDo=II>):
    - CZ: of all 21 IdPs only 3 have joined eduGAIN, so we would need to convince 18 organisations to do so
    - BE: only Belnet office in IdP list
    - CH: 15% have signed the necessary legal document for interfederation support since March
    - IT: The answer to your question, with this IDP login works only if attribute release policy is manually configured. This means that someone has previously asked to release the needed attributes + agrees to IDEM policies
    - EL: no attributes for the IdP tested (<https://login.athena-innovation.gr/idp/shibboleth>)
  - At this moment, not an alternative for the SPF

# Open issues

---



- Missing attributes (mainly DE):  
<http://www.clarin.eu/node/3578>
  - CLARIN IdP (but: scalability?)
  - call for action CLARIN-D and DARIAH-DE:  
<http://www.clarin.eu/node/3500>
- Trust delegation: allow a web service or application to access a shibbolized Service Provider
  - Hopefully SAML/OAuth2 bridge can solve this
    - NDG Oauth service -  
[https://github.com/wvengen/ndg\\_oauth/tree/master/ndg\\_oauth\\_server](https://github.com/wvengen/ndg_oauth/tree/master/ndg_oauth_server)
    - Contrail



# Next steps

---



- See if we can smoothen the opt-in process by signing a Code of Conduct
- Look at the experiences of CZ and NL with the opt-in process
- Continue to extend the Service Provider Federation as it currently seems to be the most certain way to reach our users
- Mutual recognition of the “homeless” IdPs of DARIAH/CLARIN
- Redundant setup of CLARIN IdP and Discovery Service