

Open Science Lunch

All about secure data archiving CLARINO



Juliane Tiemann & Paul Meurer
University of Bergen Library

What is CLARINO?



CLARINO is the Norwegian node in *CLARIN – The Common Language Resources and Technology Infrastructure in Europe*

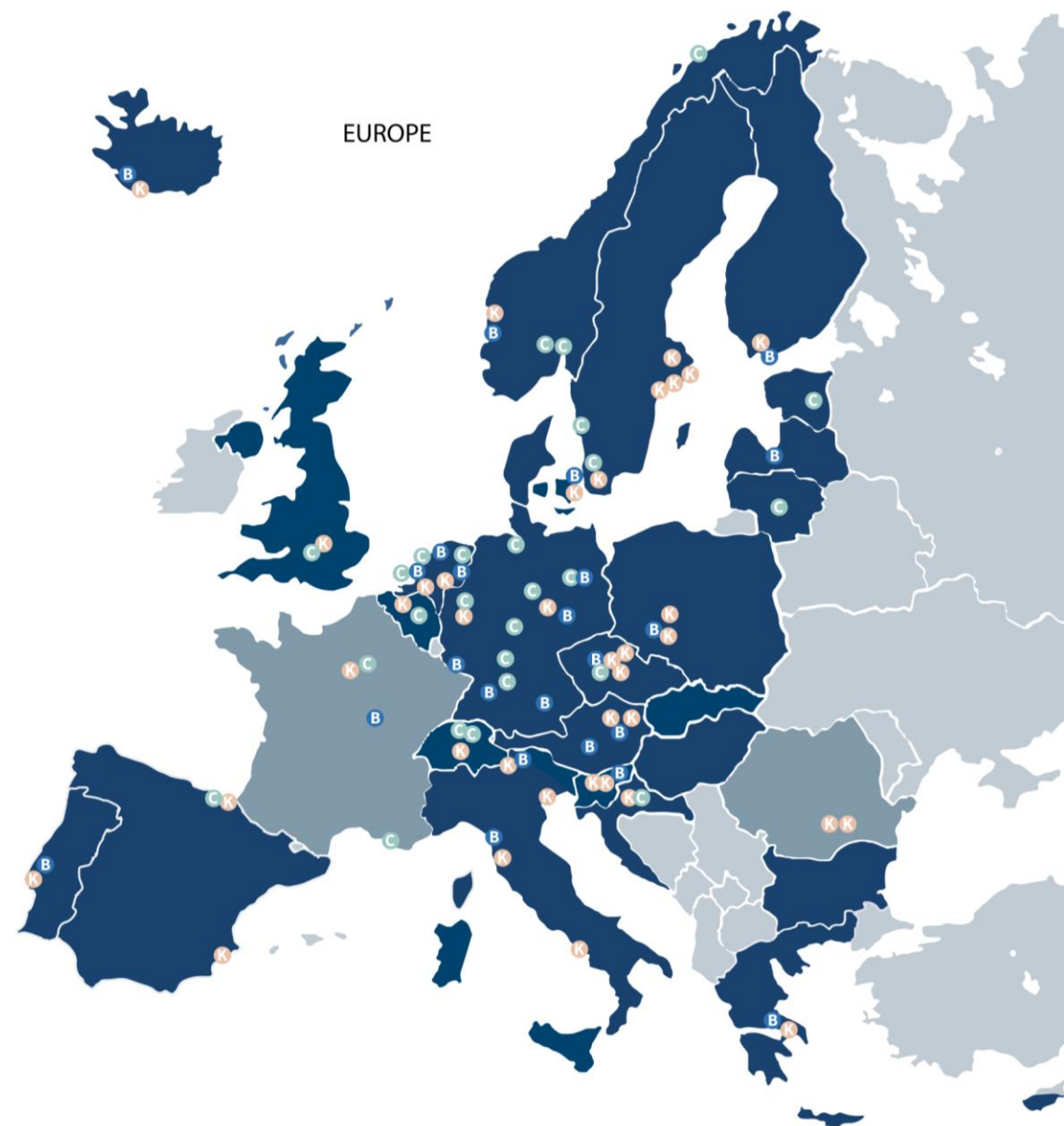
- a networked federation of centres: language data repositories, service centres and knowledge centres
- relevant for humanities, social sciences, information science, media studies etc.

National Consortia

- majority of operations, services and centres of the CLARIN infrastructure is provided and funded by CLARIN ERIC members/observers
- all members/observers create and provide access to digital language data collections, digital tools and expertise



- ERIC members
- Observers
- Countries with participating centres
- B Centre Providing Data
- C Centre Providing Metadata
- K Knowledge Centre



Data and knowledge centres in Norway



1. CLARINO Bergen Centre, UiB (B-centre; national coordinator)
2. Treebanking (Knowledge centre; operated jointly by INESS (Bergen) and LINDAT (Prague))
3. Text Laboratory, UiO
4. TROLLing (DataverseNO), UiT
5. Språkbanken, National Library

Earlier: Norwegian Centre for Research Data (Knowledge centre; merged into Sikt)

CLARINO offers



- *easy and sustainable access* to digital language data (in written, spoken, or multimodal form)
- *advanced tools* to discover, explore, exploit, annotate, analyze or combine such datasets
- help with *managing, preparing, storing, accessing or analyzing* research data related to language
- clear *licenses* (mostly open data, if possible), but restricted if needed (due to copyright or personal data)

Corpuscle and INESS



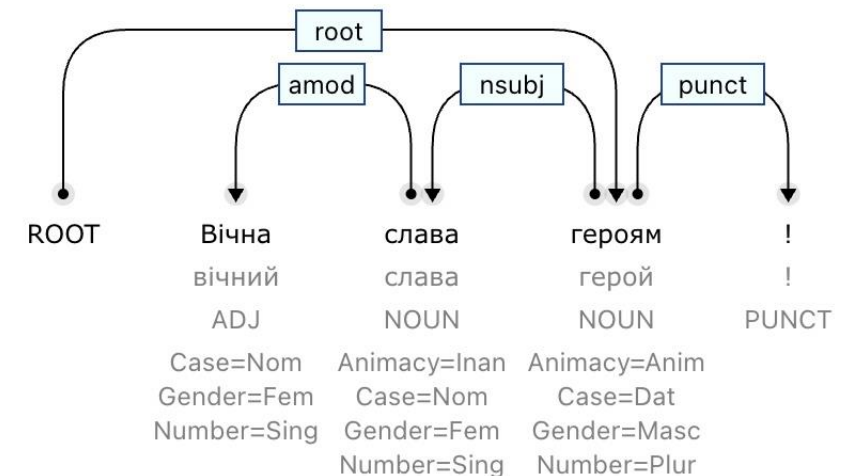
Content: searchable grammatically annotated texts in the form of *corpora* and *treebanks*

Aims: long-time *storage*; *accessibility* for the research community and the general public

Access to many of the texts is *restricted*
(Copyright, privacy)

Three types of *licenses*: **PUB, ACA, RES.**

- subtypes that encode additional licensing conditions and terms



Licenses



Access to any resource requires accepting the *license terms*

- **PUB:** Resource accessible for *all users* (including unidentified guests)
- **ACA:** Resource accessible for *academic users*. Access is granted *automatically*
- **RES:** Resource accessible for *individual users* according to the conditions stated in the license. Access is granted *manually*. Some resources require a *research plan*

Ressursen NSPC/Nor er lisensiert under følgende vilkår:

CLARIN_ACA   

BY NORED

Klikk på lenken for å se lisensvilkårene.

Når du har akseptert lisensvilkårene vil du få tilgang til denne ressursen.

Aksepter

Ressursen ASK Hovedkorpus er lisensiert under følgende vilkår:

CLARIN_RES-PRIV      

PERM PLAN BY PRIV NORED

Klikk på lenken for å se lisensvilkårene.

Denne ressursen har en restriktiv lisens. Please contact [Paul Meurer](#) to apply for access to the resource.

Credentials management



paul.meurer@uni.no ×

ID: paul.meurer_uni.no@clarin.eu

Email: paul.meurer@uni.no

Entitlement: <http://www.clarin.eu/entitlement/academic>

Credentials:

Family	Credentials	Resources	Types	Languages	Collections	Delete
treebank	treebank-view-all	(all)	(all)	(all)	(all)	
corpus		(all)	(all)	(all)	(all)	

☐ Show all

- ☐ administrator
- ☒ corpus-view
- ☐ corpus-annotate
- ☐ documentation-edit
- ☒ search-query-save
- ☐ search-query-save-public

Technical solution



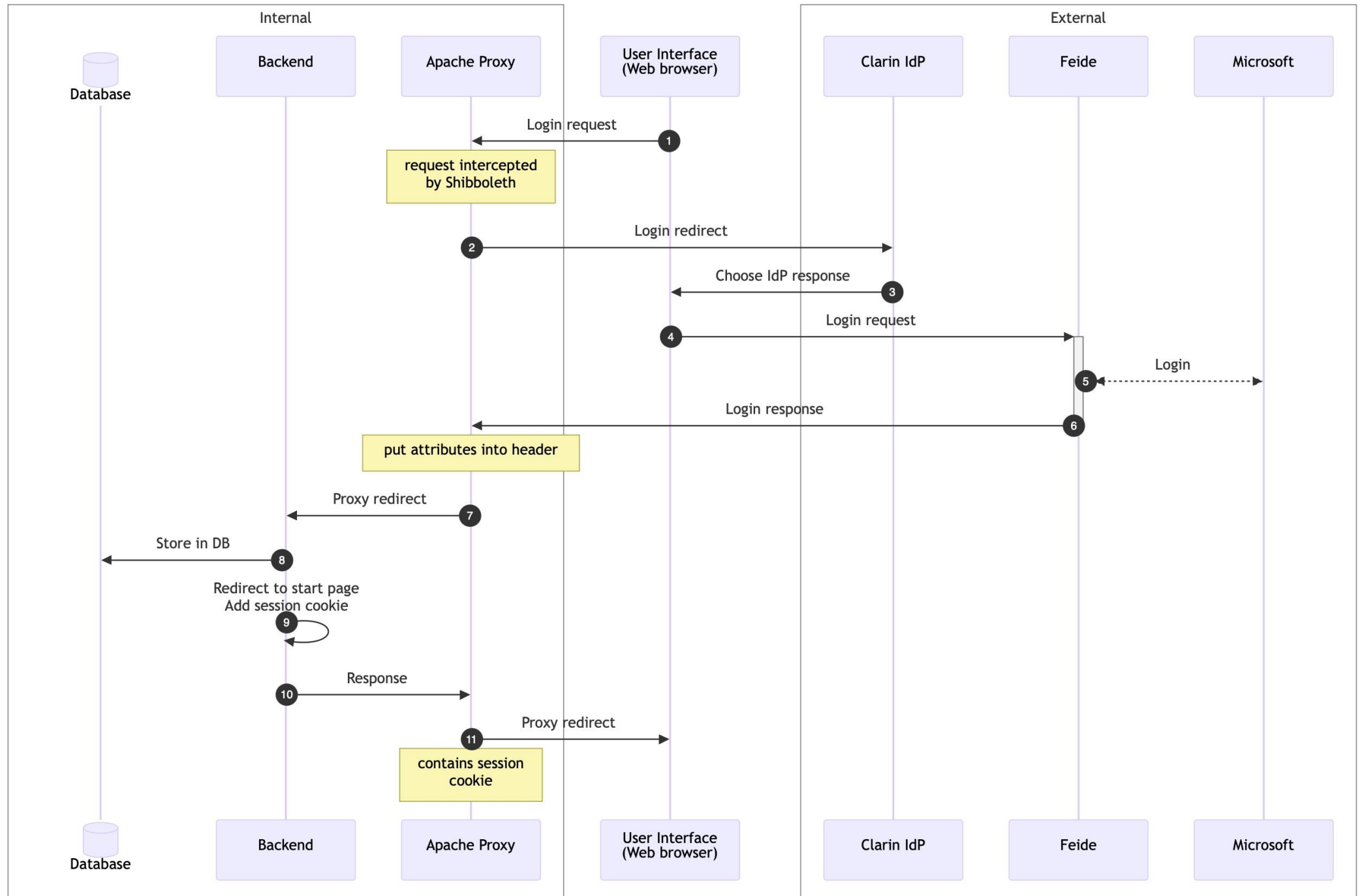
Login via CLARIN Service Provider Federation

- Feide, eduGain, and other federations, CLARIN's own IdP
- Users not in eduGain can apply for a *CLARIN account*; they are asked to describe their motivation

Software:

- Backend and API web server: Common Lisp
- Frontend: vue.js
- Web Proxy: Apache
- Authentication and authorization: SAML2 protocol, Shibboleth module in Apache

Login request



Normal request

