

CLARINO CONSORTIUM MEETING 2013

Koenraad De Smedt

University of Bergen, Norway

Solstrand, September 12-13, 2013

MEETING GOALS

- Get acquainted (new co-workers, Advisory Council)
- Present overview of status
- Receive feedback from AC
- Adjust plans
- Establish procedures
- Strengthen coordination and communication

PRESENTATION ROUND

Gisle Andersen

Volker Boehlke

Torbjørn Breivik

Koenraad De Smedt

Björn Gambäck

Ciprian

Gerstenberger

Anje Gjesdal

Øyvind Gjesdal

Kristin Hagen

Lars Hellan

Knut Hofland

Lars Johnsen

Rune Knudsen

Marita Kristiansen

Renate Krokås

Emanuele Lapponi

Leif Longva

Gyri Losnegaard

Bente Maegaard

Talan Memmott

Paul Meurer

Johan Myking

Anders Nøklestad

Jan Odijk

Oddrun Ohren

Vemund Olstad

Christian-Emil Ore

Alois Pichler

Øystein Reigem

Scott Rettberg

Eiríkur Rögnvaldsson

Victoria Rosén

Eike Schnall

Erik Velldal

Freddy Wetjen

CLARIN / CLARINO

CLARINO (Common Language Resources and Tools Infrastructure Norway) defines the Norwegian participation in the construction (implementation) phase of CLARIN.

The overall objective is to give researchers improved access to language data and related tools.

PROJECT FACTS

CLARINO runs from April 12, 2012 until March 28, 2019.

It is funded for 5 years by the Norwegian Research Council under grant 208375/F50, and in part by the consortium members.

- University of Bergen (UiB: LLE and UB) coordinator
- University of Oslo (UiO: ILN and IFI)
- Norwegian School of Economics (NHH)
- University of Tromsø (UiT)
- Norwegian University of Science and Technology (NTNU)
- Uni Computing, a division in Uni Research AS

Cooperation partners are the National Library of Norway and Uninett Sigma AS.

CLARINO AND CLARIN ERIC

The ESFRI initiative CLARIN became an ERIC on February 29, 2012.

The CLARINO consortium became an observer in CLARIN ERIC from February 1, 2013.

An observer agreement was signed on September 4 by the CLARINO coordinator and the CLARIN ERIC director.

USERS THROUGHOUT CLARIN WILL BE ABLE TO:

- connect to Europe-wide facilities through single sign-on;
- transparently search Europe-wide catalogues of documented data and tools;
- be licensed to download and use data under clear conditions;
- search and analyze data with powerful tool chains, visualizations and transparent use of HPC;
- reference data persistently and upload own data which becomes long term archived.

WHAT NEEDS TO BE DONE?

Minimally:

- Document our resources with CMDI metadata which can be harvested
- Make sure that all our resources are properly negotiated and licensed
- Authenticate our users through AAI federations
- Give all our resources proper PIDs
- Provide data analysis tools (corpus search, workflows, etc)
- Establish certified centers

CLARINO PROJECT PLAN

- WP1** Centres Setup
- WP2** National Registry and Long-Term Archiving
- WP3** Trusted AAI
- WP4** Electronic Editions Platform with IDP
- WP5** Glossa Integration
- WP6** Corpuscle Integration
- WP7** Terminology Integration
- WP8** LAP
- WP9** Tool Adaptation
- WP10** Data and Metadata Adaptation
- WP11** Management and Dissemination
- WP12** Use Cases, Evaluation and Delivery

MAIN INITIAL TECHNICAL CENTERS

- The National Library of Norway: A + B center
 1. National cataloguing and long term preservation
 2. Access to its own catalog of LRs (Språkbanken)
- The University of Bergen: two B centers
 1. Treebanking services (INESS) and Corpuscle (WP6)
 2. Repository of humanities resources and terminology resources and services (WP7, in cooperation with the NHH).
- The University of Oslo: three B centers
 1. The Text laboratory will provide LRT access and corpus services (WP5)
 2. EDD will provide access to digital editions for the Humanities (WP4)
 3. IFI will provide a Language Analysis Portal with workflows and HPC access (WP8)

CENTER REQUIREMENTS (B)

1. Server certificates
2. Federated identity management (Feide, CLARIN idp, eduGain)
3. Assessment (Data seal of approval) of well-documented repository
4. Metadata (CMDI, IsoCat, OAI-PMH)
5. PID (handles)
6. (optionally) support for Federated Content Search

No requirements to adopt specific software (DSpace, Fedora Commons, etc.)

LICENSING OF EACH RESOURCE AND TOOL

- CLARIN-compatible agreements must be signed by IPR holder and CLARINO
- Define which license holds between CLARINO center and end user (often CC-by)
- CLARINO centers must announce their CLARIN Terms of Service

USE CASES AND TARGET AUDIENCE

- Which user groups, also outside of linguistics and language studies, benefit from access to the infrastructure?
- Which use cases and demonstrators can exemplify this potential?
- Which user communities can we involve in further work?

OBSERVER STATUS IN CLARIN ERIC

This regulates some rights and obligations of both parties.

CLARINO cooperates in:

- CLARIN Legal Issues Committee (CLIC)
- Center Assessment Committee
- Metadata (CMDI) and data registry coordination
- Mobility actions
- National Coordinators' Forum (NCF)
- Workshops and Annual meeting (Prague 2013)