

**Title** Minutes SCCTC 8 September 2020  
**Version** 1  
**Author(s)** LS  
**Date** 2020-09-08  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2020-1734



## Participants

Martin Matthiesen (Chair, FI), Jan Hajič (CZ), Pavel Straňák (CZ), Thomas Eckart (DE), Lene Offersgaard (DK), Krista Liin (EE), Roberts Dargis (LV), Riccardo Del Gratta (IT), Daan Broeder (NL), Luis Gomes (PT), Leif-Jöran Olsson (SE), Brian MacWhinney (USA), Dieter Van Uytvanck (CLARIN ERIC), Linda Stokman (CLARIN ERIC, minutes)

Excused: Matej Ďurčo (AT), Vanja Štefanec (HR)

## 0 Action points

#	Action	By whom	By when
1	Prepare a broader discussion on the FAIR Data Maturity Model within the competence of the PID Task force this week and send out an invitation for this. (see questions Martin M. sent to SCCTC by email on 27 April 2020)	Daan & Martin M.	Oct 2020

## 1 Agenda

1. Agenda: the agenda is approved with the following additions to agenda item 5 (Any Other Business): 'Items to be highlighted during the Annual Conference' and 'Organizing upcoming Taskforce meetings'
2. Approval [minutes last meeting](#) (30.6.2020) & action point status
3. Centre assessments
4. Status update per country/member (please provide a short bullet-wise summary in the Google Doc): [https://docs.google.com/document/d/1YRYqFuyLMOVEO8Gb6Fg4IYVYsid9CVS\\_oJiHXRgnHrec/edit](https://docs.google.com/document/d/1YRYqFuyLMOVEO8Gb6Fg4IYVYsid9CVS_oJiHXRgnHrec/edit)
5. AOB
  - a) Items to be highlighted during the Annual Conference
  - b) Organizing upcoming Taskforce meetings

## 2 Approval minutes last meeting & action points

- a) The meeting minutes of 30 June 2020 ([CE-2020-1696](#)) are approved.

## 3 Centre assessments

Currently there is no report, because there is still an issue that Lene needs to discuss with the Assessment Committee and Dieter. The report will be available at the next Centre Committee meeting.

## 4 Status update per country/member

### Austria

- No report

### Bulgaria

- Bulgarian CHILDES corpus was uploaded to the [TalkBank](https://chilides.talkbank.org/access/Slavic/Bulgarian/LabLing.html) system:  
<https://chilides.talkbank.org/access/Slavic/Bulgarian/LabLing.html>

### Croatia

- The contract for the selected second FTE (Daša Farkaš) is prepared and she will start working for HR-CLARIN from 2020-10-05. This will fill up the Croatian CLARIN coordinator's team and we'll finally be able to participate in the full capacity.

### Czech Republic

- Continued work on datasets (discourse in treebank, semantic classes, etc.) and TEITOK environment for annotation and search of TEI documents, also minor web changes, to simplify access by Digital Humanities (and any other, in fact) Users
- SSHOC work continued, inc. a visit from Goettingen, active participation at the SSHOC conference (Sept. 9-10)
- Colleagues from the UFAL Institute (<https://ufal.mff.cuni.cz>) with co-authors from Oxford University and Google Brain published a paper in Nature Communication about the latest cs-en MT which rivals professional translators in adequacy (<https://www.nature.com/articles/s41467-020-18073-9>, MT service already running at the usual <https://lindat.mff.cuni.cz/services/translation/> ("CUBBITT" models))

### Cyprus

- No report

### Denmark

- Deployed a Dockerised version of Språkbanken's KORP, using a 2020 build of the software replacing our old fork.
- In the middle of further refining this dockerised KORP for use in another project
- Continued development of an adaptive UI for rendering of TEI documents

### Estonia

- Preparing to upgrade our KORP
- Collaboration with literary scientists – the seminars were a success, writing papers about the correspondence use case.
- User involvement with non-scientists: 3 seminars to entrepreneurs introducing practical usage of language resources for content generation
- Some new big datasets available: Estonian National Corpus 2019 - [sentence aligned version](#) and [.vrt version](#).

### Finland

- Funding application sent for providing [speech data](#) to research and business.
- Lat.csc.fi to be [shut down](#), 25k PIDs need to be reassigned.
- Ongoing:
  - Server consolidation (next: metashare.csc.fi)

- Korp.csc.fi update

#### France

- No report

#### Germany

- Continued work on AAI integration in FCS
- Continued work on Fedora Commons repository migration / awareness at Lyrasis/Duraspace

#### Greece

- No report

#### Hungary

- No report

#### Iceland

- A number of additional deliverables of the National Language Technology Programme have been added to the CLARIN-IS repository.
- CLARIN-IS participates in preparatory discussions for submitting a large joint proposal to the new infrastructure roadmap for Iceland.
- Together with Almennarómur - Icelandic Centre for Language Technology and the SÍM Language Technology Consortium, CLARIN-IS is planning a conference on Language Technology which will be held online, probably in February.

#### Italy

- ILC4CLARIN
  - Imported 394 items in the repository for the ALIM (Archive of medieval Latin documents). Details on <https://github.com/cnr-ilc/alim2clarin-dspace>
  - Working on others items (different class of documents) to load.

#### Latvia

- Nothing to report

#### Lithuania

- September 22-23: virtual conference “Human Language Technologies - the Baltic Perspective ([Baltic HLT 2020](#))”, organized by the Centre of Computational Linguistics and CLARIN-LT centre at Vytautas Magnus University, Kaunas.

#### The Netherlands

- The Final report of the CLARIAH-CORE project has been approved by our funder, NWO.
- An OpenSoNaR tutorial is being prepared, to be held on Oct 9, 14:00-17:00. The tutorial will be on-line and in Dutch.  
(<https://www.clariah.nl/evenementen/agenda/opensonar-tutorial> )

## Norway

- Continued working on plans for the CLARINO+ work packages, including dissemination and outreach of CLARINO tools and services

## Poland

- Launch of a new services on CLARIN-PL:
  - ComCorp - a tool for comparing language characteristics of corporas - <http://ws.clarin-pl.eu/compcorp.shtml>
  - spelling correction tool - <http://ws.clarin-pl.eu/speller.shtml>
  - punctuation correction tool - <http://ws.clarin-pl.eu/punctuator.shtml>
- User Involvement: preparation for publication of the 1500 corpus of the most important Polish political speeches from 1918-2018.

## Portugal

- Continued work on integrating new tools into the workbench and fixing bugs
- Beta testing of file processing services

## Slovenia

- Second resubmission of CLARIN.SI application for CTS, taking care of 2 remaining comments
- Upgrade of public CLARIN.SI pages (in progress)
- New submissions to repository, e.g. annotation models for standard and non-standard Croatian and Serbian
- As one of the organiser, preparing the (virtual) conference Language Technologies and Digital Humanities (September 24-25, 2020), <http://www.sdjt.si/jtdh-2020/en/>

## South Africa

- Deployed development version of Taalportaal from IVDNT to support localisation to isiXhosa and isiZulu
- Web platform improvements

## Sweden

- Continued work on AAI integration in FCS
- New Swe-Clarín website to be rolled out before consortium meeting
- Responded to SSHOC call for CTS process support, new resources for publication on the way. Received DataCite org account
- Finally some progress with our OpenShift 4 deployment as well as supplementary data import into repo.

## United Kingdom

- Funding proposals are in preparation for full membership of CLARIN ERIC and for research infrastructure services in the UK.

## TalkBank, CMU (USA, third party)

- All TalkBank data were moved to the CMU Cloud and Cloud Plus last year.
- We are exploring having a mirror site at the MPI-Nijmegen for GDPR issues.
- In May, we applied for Core Trust Seal. No reply yet

- We fixed our problems with OAI harvesting for VLO etc.
- We have created a comprehensive and fast database search engine called TalkBankDB
- We have implemented a system for Collaborative Commentary on browsable we transcripts linked to media.
- We organized a dementia-recognition challenge for Interspeech
- We improved our Grand Rounds pages for tutorials on language disorders.
- We are unable to get Shibboleth running.

## 5 AOB

### a) Items to be highlighted during the CLARIN2020

Martin M. suggests highlighting Task Forces during the Annual Conference, especially the Task Forces that have been most active. For the less active Task Force like AAI, it can be made known that AAI is not a high frequency Task Force. Martin M. also suggests making Task Force mailing lists available for people outside the Task Forces, so positive topics can be shared more broadly. This way relevant players can be drawn out to exchange information. Above information will be put on a poster Martin M. is preparing.

### b) Organizing upcoming Task Force meetings

Taskforce meetings will be organized separately from the CLARIN Annual Conference. Martin M. will liaise with the Task Force coordinators to make sure a decision on a date will be made soon so it can be announced/ distributed at the Annual Conference.

**Next meeting:** A doodle for the next virtual meeting will be sent out shortly.