

**Title** Minutes SCCTC 14 October 2020  
**Version** 1  
**Author(s)** LS  
**Date** 2020-10-14  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2020-1758



## Participants

Martin Matthiesen (Chair, FI), Matej Ďurčo (AT), Felix Helfer (DE), Lene Offersgaard (DK), Simon Gray (DK), Krista Liin (EE), Vanja Štefanec (HR), Roberts Dargis (LV), Tomasz Naskręć (PL), Martin Hennelly (SA), Leif-Jöran Olsson (SE), Cyprian Laskowski (SI), Brian MacWhinney (USA), Dieter Van Uytvanck (CLARIN ERIC), Linda Stokman (CLARIN ERIC, minutes)

Excused: Daan Broeder (NL)

## 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
2	Complete all assessment validity dates in the centre registry	Linda (Office)	Dec 2020
3	Take care of CNRI/EPIC Handle issue	Dieter	Dec 2020

## 1 Agenda

1. Agenda: the agenda is approved.
2. Approval [minutes last meeting](#) (8.9.2020) & action point status
3. Centre assessments
4. The Vienna Repo Solution: Presentation by Matej and discussion
5. Status update per country/member (please provide a short bullet-wise summary in the Google Doc): <https://docs.google.com/document/d/1F1j2FnX9ogAfu4OUMGjwVNVhQPM300XCVtIoMu1H-3U/edit#heading=h.gjdgxs>
6. AOB

## 2 Approval minutes last meeting & action points

- a) The meeting minutes of 8 September 2020 ([CE-2020-1734](#)) are approved.

## 3 Centre assessments

Possible candidates for re-assessment in the next assessment round of November:

- AT: ACDH-ARCHE

- USA: CMU Talkbank

The deadline for the next assessment round is set for 16 November 2020, if more time is needed please let Lene know.

*AP2: Complete all assessment validity dates in the centre registry. LS (Office)*

#### 4 The Vienna Repo Solution: Presentation by Matej and discussion

ARCHE was running on Fedora 4 since 2017, this version was considered unfit for production even by its creators. There were major issues with performance on ingest and read, massive workarounds around the deficiencies of Fedora 4 lot of php-code around the Fedora-blackbox, the specific requirements were not covered by any of the known solutions and all known solutions have Java component for which ARCHE did not have sufficient in-house competencies. Matej held a presentation about ARCHE's own custom-made solution for its repository. [See the slides \(PDF\)](#)

#### 5 Status update per country/member

##### Austria

- Preparing the CTS and CLARIN-B reassessment

##### Bulgaria

- Nothing to report

##### Croatia

- Second FTE at HR-CLARIN consortium coordinator (FFZG) started her employment from 2020-10-05 and we are now fully equipped for further actions.
- The domain for HR-CLARIN (clarin.hr) has been parked and we are preparing the web site of the HR-CLARIN consortium.

##### 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

- Setup of the international advisory board

- Ongoing organisation of the CLARIN-CY general assembly in December 2020
- EuroMed2020 conference on digital Heritage organisation (almost 600 registrations up to now)
- National and Univ. aggregators agreements, Europeana
- Negotiations with the deputy ministry for research and innovation for a sustainable financial support.
- Ongoing consultations with IIF-group of experts for standards on 3D holistic documentation

### 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
- Server consolidation of repository - still dealing with smaller related issues

### Estonia

- Continuing to upgrade KORP to v9

### Finland

- Nothing (new) to report, efforts reported last month ongoing.

### France

- ORTOLANG is now officially a C-centre. A Core-Trust-Seal application has been sent and the centre is waiting for the evaluation. If it is positive, it will apply for B-centre status.
- Both COCOON and ORTOLANG (French C-centres) are now interfaced with the content search web tool. The coverage of the content search will be improved during the next months.
- Finishing the blog posts for the "Tour de CLARIN" for France with (a lot of) help from Darja and Jakob

### Germany

- Ongoing communication with Lyris concerning new Fedora version, preparations for migration tests.

### Greece

- Preparation and uploading of new resources by three partners' repositories.
- Contacts with two Universities candidates for joining the national network.
- Design and development of a new backend, frontend and user management module for the CLARIN:EL infrastructure (to be available by the end of the year).

### Hungary

- No report

## Iceland

- Almost 30 deliverables of the Icelandic National Language Technology Programme - tools, databases, language descriptions - were added to the CLARIN-IS repository in September.
- Iceland will be featured in the third volume of Tour de CLARIN. We have finished writing four blog posts and an Icelandic researcher has been interviewed.

## 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 other items (different class of documents) to load.
  - Repository updated to release 2020.01

## Latvia

- State research programme “Resources for Digital Humanities” accepted, starts on October 1, IMCS UL (CLARIN-LV) is partner involved in creation and maintenance of LRs
- Active participation in CLARIN conference, including presentations at PhD session and CLARIN for teachers Presentation at ELG workshop of Baltic HLT 2020 conference *“Latvian Language Technology Landscape in 2020: Some Important Achievements after White Papers”*

## Lithuania

- September 22-23, Baltic HLT 2020 conference organized virtually; [proceedings](#) published
- New resources added to CLARIN-LT repository

## The Netherlands

- Successful webinar on OpenSONAR corpus application use on Oct 9 <https://lotschool.nl/opensonar-tutorial/>, in the ‘Week of the Dutch Language’. More than 40 registrations, more than 30 participants. Most participants from NL and B but also from Sweden, Canada, Germany and South Africa. Unfortunately, despite earlier testing, many authentication problems for non-Dutch participants. Fortunately CLARIN helped out by fast track approving new CLARIN IdP accounts, but we still had to temporarily take the authorisation off completely. That happens when one makes backwards-incompatible upgrades (which should be forbidden, imho [JO]). The webinar has been recorded and will be made publicly available soon, as will the presentations, exercises and solutions to the exercises.
- Interesting ‘Datastory’ on The ‘Spanish’ Flu Epidemic in The Netherlands 1918-19. See <https://stories.datalegend.net/spanishFluNetherlands/>

## Norway

- No report

## Poland

- Nothing to report

## Portugal

- New resources added to the repository
- Federate content search: up and running, after technical glitch in eric central hub eventually repaired
- Implementation and testing of file processing (continued)
- Implementation and testing of web services API (continued)
- Newer, more streamlined docker stack configuration and deployment routines

## Slovenia

- Received CTS / B-centre (re)certificatioion
- Virtual conference "[Language Technologies and Digital Humanities](#)" (September 24-25, 2020) successfully concluded, proceedings and video recording published
- Working on ParlaMint 1.0, soon to be published.

## South Africa

- Moodle instance has gone to production for delivery of course content related to academic writing
- Grammar Portal, from IvDNT, development version is complete and content is under development with intention to push to Production in 2021
- Implementation of CLARIN IdP was missed as part of CLARIN B center and is now under implementation

## Sweden

- Swe-Clarín Consortium (virtual) meeting performed. All but 1 partner participated. CLARIN Annual conference attendance high. Positive experience.
- Deployment of new Swe-Clarín webpages postponed due to university email outage since 2020-09-18. Reassessment preparations.
- Resource release preparations progressing slowly.
- FCS-AAI some further slow progress too.

## United Kingdom

- Tour de CLARIN interviews and blog posts completed.

## TalkBank, CMU (USA, third party)

- Last week, we finally received a reply from CTS on our application from May for recertification. Their requirements are becoming increasingly detailed, so it may take some time to get this done.
- We have been doing a lot of software development for FCS-related applications, collaborative commentary, and other database systems.
- There is also a large increase now in interest from universities and companies in work on speech processing for people with language disorders and TalkBank data in this area is being heavily used.

## 6 AOB

### • **Shibboleth Proxy**

Matej: Concerning the Shibboleth proxy, currently every time we want to register a service on our side as it is a separate service provider and we have to propagate the meta data for each of them separately and we have 20 services. This is difficult to

manage and I understood that there was an idea of having a kind of CLARIN or Research Infrastructure as a proxy that all older IDP would have contact which as a trusted man in the middle which is kind of trusted man in the middle.

- Dieter: this is still on the roadmap, because it relates to having a bridge to Open ID but it is not yet in place.
- Matej: My vision would be that we have many services, and I would like to have all the users logging into the them through single sign on for people, but it's basically not feasible to do that, if I have to register each service separatley.
- Dieter: that is something you could already circumvent by having a proxy SP on your side. That's what we're also doing on a central level, for that configuration we can readily share all the details on how to do that because that really saves you a lot. You only need to register one Shibboleth sp and everything is proxies behind that on your site.

Dieter will have a chat with Matej and point him to the available documentation.

- **Updates in the pipeline about persistent identifiers,**

Dieter: We got an answer on a request from the IDs to Epic whether they could migrate their handle prefix from CRI to epic and apparently but that is no longer possible. This is quite a drastic change in policy. So we will try to find out details about it and to document it better because it means that one of the original promises that we got by using the handle system they named this mobility with a prefix is no longer in effect. So this is something we need to Get into discussion with the epic to find out the details about this.

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