

**Title** Minutes SCCTC 7 December 2017  
**Version** 1  
**Author(s)** LS  
**Date** 2017-12-13  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2017-1149



## Participants

Dirk Goldhahn (DE), Jan Hajič (CZ), Krista Liin (EE), Martin Mathiesen (FI), Harris Papageorgiou (GR), Lene Offersgaard (DK), Riccardo Del Gratta (IT), Marcin Pol (PL), João Silva (PT), Luís Gomes (PT), Leif-Jöran Olsson (SE), Jan Jona Javoršek (SI), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

Excused: Matej Ďurčo (AT), Kiril Simov (BG), Nicolas Larousse (FR), Christophe Parisse (FR), Andrius Utkla (LT), Robert Dargis (LV), Daan Broeder (NL), Freddy Wetjen (NO), Martin Wynne (UK)

## 0 Action points

#	Action	By whom	By when
1	Set up small author task force to summarize proposals made for archiving software (agenda point 4)	Lene	Next SCCTC meeting
2	Submission of evaluation documents and web pages must be accompanied by a short summary in English. Add this requirement to the B-centre criteria	Lene	Next SCCTC meeting
3	Propose new centre labels as a more human-readable alternative to the letters	Dieter/Karolina	a.s.a.p.

## 1 Agenda

The following agenda was approved:

1. Agenda
2. Approval minutes last meeting & action point status ([CE-2017-1100](#))
3. Update from the assessment committee (Lene)
4. Update on the transition from DSA to CTS (Dieter, [CE-2017-1144](#))
5. Draft Centre Committee Workplan 2018, version 2 ([CE-2017-1099](#))
6. Report from the [DELAD](#) workshop (Martin Matthiessen)
7. Reports from the task forces: (task force coordinators) -> [see Google Doc](#)
  - CMDI
  - Metadata curation
  - Federated Content Search
  - Persistent Identifiers
  - AAI

8. Status update per country/member (all) -> please add a snippet for your country to [the reporting Google Doc](#)
9. Any other business

## 2 Approval minutes and action points

The minutes of the f2F Budapest SCCTC meeting of 18 September 2017 are approved.

## 3 Update from the assessment committee

### Summary

*Lene (DK)*: The deadline for the 10<sup>th</sup> round was in October. The assessment committee did the individual assessments in November. The summaries are being written, once written they will be send to the centres that applied, so they can give their comments. Then the report can go to the SCCTC for approval.

### Proposal

Lene suggest to send the report to the SCCTC by e-mail and give the committee one week to reply and inform if they have any objections instead of waiting for comments till the next Centre Committee meeting.

Lene can send the report of the summaries to the individual countries out today or at least next week so they can comment. We can then have it for approval in the first week of January.

**This proposal has been accepted by the SCCTC.**

### Updated checklist for next round

*Lene (DK)*: There will be an update of the checklist for the next round, it will be minor things. The updated checklist should be available in January.

### General comments/questions

*Dirk (DE)*: 8 German centres will be expiring end of June 2018, there will also be two new centres (Cologne and Frankfurt), when would be the best time for assessment?

*Lene (DK)*: Next assessment round will be March 28 (not a fixed date)

*Dieter (CLARIN ERIC)*: but we will try to have it by default unless someone strongly objects and we should hear that in January 2018 latest.

*Dirk (DE)*: Will discuss this with the centres, should probably be possible

*Martin (FI)*: We have submitted our CoreTrustSeal (CTS) request 6 weeks ago but haven't heard back yet, unsure about the timetable from CTS side (what questions can we expect), so unsure if can we submit our re-assessment in March.

*Dieter (CLARIN ERIC)*: You can always mention in your submission for B-centre re-assessment that you initiated the CTS, this is valid enough for granting the B-centre status, because we know the timetable is unreliable on the CTS side.

*Marcin (PL)*: we will expire in March, should we prepare for march or for the round after that?

*Dieter (CLARIN ERIC):* We have two rounds a year, one in the spring (March) and one in the fall (September), If you would like to go for an early re-assessment in 2018, March would be the one to go for. Please do not hold back because of the slow responses by DSA/CTS, these are two different processes.

*Marcin (PL):* We will prepare for March 2018

#### **4 Update on the transition from DSA to CTS**

*Dieter (CLARIN ERIC):* The official answer from the CTS board on our remarks can be found in document CE-2017-1144. They understand that the communication was not optional. There was a bit of confusion and a mistake in some of their calculations that they would accept 7 out of 10 applications but they confirmed that it will be 8 out of 10, which means that the two new centres will go through the official procedure in January instead.

About the Fee structure: quantum discount for CLARIN ERIC will be discussed in the CTS Board meeting this afternoon. DVU will have contact next Tuesday. We aim to have a clear outcome in December.

*Marcin (PL):* they will not proceed our application even though it is almost done.

*Dieter (CLARIN ERIC)* You have submitted after the first of November, so it is in line with the agreement that we have, that the assessment that are not pre-announced will be taking place in the CTS context. We will inform you soon about the fees issues and if CLARIN ERIC can take care of the fees for you.

#### **5 Draft Centre Committee Workplan 2018, version 2 (CE-2017-1099)**

*Dieter (CLARIN ERIC):* This document is a follow up of what has been put together (ingredients) for the Centre Committee Workplan that was discussed during the Budapest meeting. Dieter elaborates on the document.

##### **Comments/requests/ ideas**

*Martin (FI):* What does starting to work on the EOSC-Hub mean?

*Dieter (CLARIN ERIC):* We have 3 tasks in this EOSC-Hub project. What we will do is continue working on the VLO and also looking at metadata sources from outside, analyse where there are other relevant sources of information and make a specific plan.

Will be send to SCCTC for approval, final version by the end of this year or early January

#### **6 Report from the DELAD workshop (Martin Matthiesen)**

*Martin Mathiesen (FI):* This workshop took place in Cork, Ireland in October. Dieter and FIN-CLARIN (myself and Mietta Lennes) were their together with researchers from the DELAD initiative. The initiative is about sharing speech data from Corpora of disordered speech (CSD), they are gathering sensitive data and would like to be able to share it. Researcher presented their current research on the sensitive data sets they have gathered. Mietta and I presented our access control solution in Finland and Dieter presented CLARIN as a whole. CLARIN was received quit positively. It was seen as a good way to share data and metadata, However, there is still quite a bit of work ahead, because of the sensitivity of the CSD data. A lay-out approach is definitely needed because of the sensitivity of the data.

Data protection regulation says that you should only share data that is really needed for your kind of research. Researcher who don't need raw data who can also work with anonymous data would need to be able to get access to that data. What is needed is a kind of dynamization of data sets so that sub datasets can be applied, to my knowledge no Centre is able to do it. We could do this by hand by just packaging things, but the complexity is going to be quit high quit quickly.

We had a look at Talkbank and from a European perspective it was a bit shocking, they had highly sensitive data available just by clicking on the download link. It is mentioned for research only and anonymize properly but anybody can access it.

General problem with data came up: There have not yet been any data breached that courts have dealt with, so committees have pretty strict restrictions on everything, which makes it very hard to share data. This is partially why the DELAD initiative was started to overcome this. If Talkbank can make research on this possible we should be able to in the EU. We are in discussion with Helsinki to share CSD data. Its serves as a practical approach to see how far we get with this.

*Lene (DK)*: If we cannot anonymize the data there is a strict demand for login. The way Talkbank downloads cannot be done in EU. Talkbank is under US law which we cannot copy here in the EU. Login is very important because of the regulations, but we are not yet able to log everything.

*Martin (FI)*: CMU Talkbank very useful, however we do not want to put data out there as downloads like Talkbank, on the other hand it makes it possible to make serious researchers do serious research. I am not yet aware if data in Talkbank is used in serious research, but even though I do not agree with the way Talkbank shares data, I do think that in EU it should be possible for researcher to do serious research with this data.

*Harris (GR)*: are there any better alternatives, for sharing this private data? Instead of keeping the data somewhere where the data owner things it safe to keep and curate and instead downloading the data, allowing data work flows and any kind of research can run on the siet where the data exist , so instead of bringing data to the technology, bring the technology to the data. We should see if this is possible.

*Martin (FI)*: We have a service that makes it possible. You have to trust the researcher that after he downloaded and encrypted the data he will delete it after processing.

*Dieter (CLARIN ERIC)*: URL one of this was brought up during the workshop and is a bit inline work suggestion Harris: <https://www.beat-eu.org/platform/>

*Martin (FI)*: CLARIN has a nice role to play here. CSD can be archived at local CLARIN centres all over Europe whereas they can be made findable through a central portal via their (harvested) metadata. CLARIN precisely offers the standards, best practices and services which are needed for this.

*Dieter (CLARIN ERIC)*: This workshop originates from a workshop type II proposal, a written report in the form of a blog on the CLARIN website soon. There will be a follow up meeting where things can be discussed more.

## **7 Reports from the task forces: (task force coordinators)**

Dieter (CLARIN ERIC) suggests to have a rotating system for the taskforce reports per meeting, so we can go into more detail. We can also invite the taskforce coordinators who can give us more in-depth information.

SCCTC agrees to use this new rotating system from now on.

#### **CDMI**

- (no report this time)

#### **Metadata Curation**

- (no report this time)

#### **FCS**

- Continued work with centres to get further FCS 2.0 endpoints online.
- Refactoring work on Aggregator and updated dependencies and continued planning of syntactic search low hanging features
- Outreach presentation will be held at NCN final event Copenhagen next week

#### **PIDs**

- (no report this time)

#### **AAI**

- André attending GÉANTTrust and Identity Meeting 7./8.12. in Amsterdam

## **8 Status update per country/member (all)**

#### **Austria**

- No report

#### **Bulgaria**

- No report

#### **Czech Republic**

- Negotiation with Ministry of Education on the merge LINDAT/CLARIN and DARIAH CZ, still unclear when the merged center will start and when the addition of DARIAH CZ (financially) takes place
- Work on associated Ministry of Culture project for oral history workbench connected to a repository
- Preparation of new H2020 projects - meetings, conf. calls & travel
- Continuing maintenance work on the DSpace repository

#### **Denmark**

- Further development of tools and data centre
- Working with tasks in Parthenos
- Korp installed and now used in a workshop
- Partner in a so-called datasprint for Danish Literature from 1800 century

## Dutch Language Union

- No report

## Estonia

- Submitted B centre and DSA applications, currently waiting for feedback.
- Adding new corpora to our KORP - the 270M token corpus is almost out, as well as several smaller ones.
- Work on the FCS endpoint will resume after the end of semester.
- Working on adding Estonian webservice to WebLicht - currently working on the OAI-PMH endpoint.
- User involvement (at the Estonian digital humanities conference and on smaller level).

## Finland

### Done

- korp.csc.fi updated
- Major corpus update (urn:nbn:fi:lb-2017091901)
- DELAD Workshop attended

### In Progress

- Korp-cooperation
- FCS still not implemented.

## France

- No report

## Germany

- The centres are busy with their DSA/CTS assessment procedures (>= 5 centres have already submitted)
- EPIC v1 discontinuation: last German centres are in contact with DWDG to finalize migration (unproblematic)

## Greece

- No report

## Hungary

- No report

## Italy

- Edited the DSA and sent back to the reviewer. Waiting for next response
- waiting for B-assessment response;
- Continuing working on FCS (lately stuck) so not yet implemented;
- Working on the integration of OpeNer and Panacea services

## Latvia

- No report

## Lithuania

- No report

## The Netherlands

- No report, NL was excused for this meeting

## Norway

- No report

## Poland

- Started the DSA/CTS (almost finished). We want to finish it next week. (I've just got an e-mail that they are not be able to process our application as they are in the middle of a transition period to our new tool)
- K-centre certificate

## Portugal

- No report

## Slovenia

- Migration to new hardware (finished; also many thanks to our Czech colleagues for excellent support with DSpace-based engine migration and upgrading)
- Production installation of <http://www.clarin.si/noske> and <http://www.clarin.si/kontext>
- Work on localising noSkE and KonText
- Ongoing adaptation of conversion scripts from and to TEI WebAnno 3 (<http://www.clarin.si/webanno/>)
- Ongoing work on exposing CLARIN.SI tools in WebLicht
- Support with preparing new entries for repository

## Sweden

- Negotiating/planning the extension period 2019-2024, 6 modules needed to go, final budget next week. Språkbanken will also become national infrastructure.
- Repository work and outreach in many constellations
- Awaiting B-centre assessment response
- A lot of new tools, preparing metadata
- Workshops,
- FCS 2.0 endpoint support to other centres

## United Kingdom

- No report

## 9 Any other business

*Dieter (CLARIN ERIC):* Pre-announcement: for the Parthenos project an effort is being made to make inventory of the CLARIN tools and services that are available. Daan Broeder (NL) will send some further communication to the SCCTC so to help in making this inventory. You will receive detailed mail in the next week.

**Next meeting:** A doodle will be sent out (we will take into account the NCF meetings).