

**Title** Minutes SCCTC 11 December 2019  
**Version** 1  
**Author(s)** LS  
**Date** 2019-12-11  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2019-1573



## Participants

Lene Offersgaard (DK), Krista Liin (EE), Vanja Štefanec (HR), Martin Matthiesen (FI), Samúel Thórisson (IS), Riccardo Del Gratta (IT), Roberts Dargís (LV), Luís Gomes (PT), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

Excused: Samúel Thórisson (IS)

## 0 Action points

#	Action	By whom	By when
1	Follow up Google Data Search: <ul style="list-style-type: none"> <li>- setup a short poll to be sent to all centres, to check their preferences</li> <li>- Provide a short report on the current crawling issues</li> <li>- discuss the issue with Google contacts if the current efforts to get the indexing working do not succeed.</li> </ul>	Dieter  Twan & Dieter  Pavel	asap

## 1 Agenda

1. Agenda
2. Approval minutes last meeting & action point status
  - a) ([CE-2019-1545](#))
  - b) Minutes and background on Google Data Search ([CE-2019-1549](#))
3. Google Data Search (oral Dieter)
4. Chair of SCCTC
5. Centre assessments:
6. Update on taskforce-related issues:
  - a. AAI: status blacklisting proposal (Martin M.)
7. Request for input workplan 2020 (all, oral)
8. Request for endorsement and feedback (please directly in the Google Doc) from the NCF Interoperability Committee: [Draft text for a CLARIN webpage on interoperability.](#)
9. Status update per country/member (please provide a short bullet-wise summary in the [google doc](#))
10. Any other business
  - a. Foreseen date and agenda Centre Meeting: 31 March + 1 April 2020, with the SCCTC on the first day, location: Utrecht

## 2 Approval minutes and action points

- a) The minutes of 7 November 2019 ([CE-2019-1545](#)) have been approved.
- b) The minutes and background on Google data Search ([CE-2019-1549](#)) have been approved.

## 3 Google Data Search

The Google Data Search was discussed during the NCF f2f meeting that was held alongside the General Assembly meeting in November 2019. A constructive discussion took place about the structured data on the landing pages of the VLO. There were no objections (both on the level of the NCF and SCCTC), but not all NCF members attended the meeting. For this reason, the Google Data Search needs to be formally approved in the NCF meeting of 12 December 2019.

In the new upcoming release of January 2020 the restructured scheme for URL's to be more Google friendly will already be implemented. Around May 2020 the fully structured data is planned to be implemented into the landing pages. We will need to see if Google will pick this up, we will have more clarity whether the inclusion of landing pages and structured metadata works in May 2020.

## 4 Chair of SCCTC

There is a broad agreement on appointing a co-chair for the centre committee on both the level of the NCF and Bod. We do not need the General Assembly's approval on this, but the new construction does require a formalization. Dieter and Franciska will have an in-depth discussion on how to best implement this formalization next week.

Martin comments that he had a very constructive meeting with Dieter about co-chairing the committee and he feels this new construction can definitely work.

## 5 Centre assessments:

Lene informs the committee that there is no report ready for approval yet. Two centres are being assessed. The deadline for the assessment team to finish the report is 16 December 2019.

**For the next centre meeting** Lene would like to discuss how much data a centre has to provide to become a centre. Should there for instance be a minimum amount of resources? Centres can have a small amount of resources but they could be very valuable.

## 6 Update on taskforce-related issues

### a) AAI: status blacklisting proposal (Martin)

Martin has created a [Google document](#) with a proposal and is waiting for feedback. The discussion has taken a slight turn. At first, we wanted to blacklist IDP's so users could not use them at all. A second scenario is to put them on a list and mention they

are not fully compatible. This scenario is trickier because we basically shame them. In the first option we completely hide them. Both scenarios' will be discussed in a separate meeting. Martin send out an [invitation to people](#) to have a meeting about this next year.

## **7 Request for input workplan 2020 (all, oral)**

Currently there is no full workplan 2020 yet. With the new chair system just in place input from the co-chair is needed, also the previous workplans contained tasks that the committee is not responsible for. The plan should be more on what the centre committee members are really focusing on.

Highlights of things that need to be discussed with the NCF for this plan are:

1. Following up with existing centre (activate them a bit more (if necessary))
2. Following up with National Consortia who have no centres (what is keeping the NC's from having a B-centre, how can we help)
3. Look at the services (not just focus on data but also knowledge exchange on services)
4. Evaluation of connection to the NCF

## **8 Request for endorsement and feedback (please directly in the Google Doc) from the NCF Interoperability Committee:**

[Draft text for a CLARIN webpage on interoperability.](#)

The SCCTC will have a look to see if this is a reasonable document and if it can be endorsed by the Centre Committee to be put on the website.

If no objections from the centre committee members have been received by email by Thursday 12 December 2019 before 10:00AM, the document will be endorsed at the NCF meeting (12 December) to be putt on the website.

## **9 Status update per country/member**

### **Austria**

- Second meeting of new CLARIAH-AT consortium
- Submission of GAMS as second Austrian B-Centre
- Decision on a new round of CLARIAH-AT projects has been made. The projects will start in early 2020.

### **Bulgaria**

- An agreement for financial support for the second year of CLaDA-BG was prepared and it is waiting to be signed at the ministry

## Croatia

- Representative of HR-CLARIN consortium participated at the CLARIN@University workshop, Utrecht, 4-5 November 2019.
- Received the generic approval from the Ministry for 2 FTEs for every national consortium involved in ERICs and this includes HR-CLARIN. Since NC is from the Faculty of Humanities and Social Sciences, after the agreement with University of Zagreb for fixed quota of 2 HR-CLARIN FTEs in the University lump-sum, we'll open the job vacancy call, TBC by 2019-12-31.

## Czech Republic

- Received equipment grant for 2020-2022, for the LINDAT/CLARIAH-CZ project. Working on contracts, consortium agreement etc. 1.2M EUR for 5 partners.
- Third fall tutorial by the joint LINDAT/CLARIAH-CZ project, on TEI (given by the Austrian colleagues from Graz).
- New data and online service - lexical semantic resource SynSemClass 1.0, inter-lingual synonym dictionary of verbs (Czech, English) with links to many external resources.
- Universal Dependencies release 2.5 (Nov. 15) - LINDAT/CLARIAH-CZ continues to be the official repository for UD releases.

## Denmark

- Workshop about digitizing teaching at the humanities faculties in DK, a new chance to reach out to more potential users in DK.
- New developer Simon hired - will start in January
- Upgraded DSpace repo to ubuntu 18.04

## Estonia

- Workshop "[Language Resources for Content Search and Metadata Search](#)" at [7th Estonian Digital Humanities Conference](#), 4-6 December in National Library was successful. We introduced Estonian KORP, Raba (with CLARIN FCS) and CLARIN VLO and made exercises according to research interests of participants. [Handouts](#) in English are available.
- CELR launched Estonian Federated Content Search called [Raba](#) in that workshop.
- Preparing strategy plan and work plan for 2020 to be presented to Council of CELR.
- Raba (<https://raba.keeleressursid.ee>) is a content search aggregator based on CLARIN FCS and KORP. Raba can search different text corpora using SRU/CQL, similar to FCS 2.0 protocol. In addition to [Estonian KORP](#) text corpora Raba is also connected to search interfaces for Estonian lexical resources and speech corpora. Lexicons are seen as FCS 1.0, where we have constrained word form search to lexical entries (but not e.g. the definition). Since our speech corpora are annotated with morphological information, they can be queried in more detail and return both text and audio clips. Another difference to CLARIN FCS is that there is no conversion to Universal Dependencies since we wish to include the intricacies of language-specific tagsets. According to resource, there can be different annotation levels or 'layers' (with different tagsets). In the backend it is possible to add mapping for each resource, but it is not required. If a layer is used in search, then the search is automatically constrained to resources that have that layer available. An example of searching for 'language' ('keel') in Raba: <https://raba.keeleressursid.ee/?&query=keel>

## Finland

- Participation in the workshop by CLARIN ERIC on Universities @ CLARIN

- New tool for Finnish NER launched at “From Text to Knowledge” Heldig Summit
- Launching the SAFMORIL K Center and acquiring new members
- Waiting for funding decision for 2020 onwards.
- Data/Software transfer to new HPC system still ongoing.

### France

- CLARIN-FR Tutorial was accepted for 2020 LREC conference. Preparing the outline.
- Discussion about FCS, still in progress in 2 C-Centers

### Germany

- Central Monitoring:
  - Bugfixing, Configuration
  - First discussions on integration with DARIAH-DE monitoring
- FCS and AAI
  - Discussions at CLARIN-D developer meeting on next steps for “small” AAI-solution for FCS Aggregator
  - Dirk will organize VC with FCS-TF, Dieter, Kai (BBAW) on this topic
- “Technology Watch” report on repository solutions is currently being worked on

### Greece

- No report

### Hungary

- No report

### Iceland

- CLARIN-IS takes an active part in ongoing discussions and decisions on standards and formats that will be used for tools and resources developed within the Icelandic National Language Technology Program.

### Italy

- ILC4CLARIN
  - Discussion on GDS: how to organize legacy landing pages
  - Repository server migration. Docker?

### Latvia

- Participation in the workshop by CLARIN ERIC on Universities @ CLARIN, Utrecht, 4-5 November 2019
- Meeting with representatives of the Ministry of Education and Sciences to discuss current situation and possibilities for additional funding.

### Lithuania

- We’ve received an official answer from the Research Council of Lithuania that CLARIN-LT will be included in the updated “Lithuanian Research Infrastructure (MTI) Roadmap”.
- CLARIN-LT has joined the CLARIN Knowledge Centre for Systems and Frameworks for Morphologically Rich Languages (SAFMORIL).

- Meeting of representatives from prospective partner institutions (VMU, KTU, VU, BPTI, MRU) to discuss the [consortium agreement](#), Kaunas, 5 December 2019

## The Netherlands

- Radboud University launched CLARIN K-centre on atypical communication <https://ace.ruhosting.nl>
- On November 13 Jan Odijk attended the ‘TOK-dag’ (Language Development with Children Day, University Library, Amsterdam) and presented a poster about the potential of GrETEL for semi-automatic assessment of language development of children based on (transcribed) spontaneous language
- On November 18, 2019, the CLARIAH Audiovisual Data and Media Studies work package (WP5) organized a one-day meeting on computer vision and audiovisual data at Utrecht University. The main organizers, [Jasmijn Van Gorp](#), [Christian Olesen](#) and [Tom Slootweg](#), invited several Dutch media scholars, as well as computer & data scientists, to explore the potential of computer vision for the analysis of audiovisual data held by cultural heritage intuitions (e.g. film, television, photography). The organizers furthermore invited two special guests, [Taylor Arnold](#) and [Lauren Tilton](#) (University of Richmond). Arnold and Tilton kindly agreed to participate in the expert meeting, before bringing the day to a close with an interesting public lecture on their [Distant Viewing Lab](#). See <https://www.clariah.nl/en/new/blogs>
- CLARIAH co-organised the Humanities sessions at this year’s eScience Symposium, which was held on 21 November 2019 in the Johan Cruyff Arena in Amsterdam. See <https://www.esciencesymposium2019.nl/> . See also <https://www.esciencecenter.nl/news/nlesc-holds-national-esceience-symposium-focussing-on-the-promises-challenge> .  
The theme of the symposium was ‘digital challenges in open science’. Our keynote speaker, **Eero Hyvönen** (University of Helsinki / Aalto University, Finland), other speakers **Marcel Broersma** (University of Groningen) , **Roser Morante**, **Piek Vossen & Antske Fokkens** (VU University Amsterdam) and **Daan Broeder** (Humanities Cluster / CLARIN ERIC).

The session was organized by **Jan Odijk** (Utrecht University) from CLARIAH and **Carlos Martinez Ortiz** from the eScience Centre, and was chaired by **Jan Odijk**.

- In addition to the Humanities session, there were posters and demo’s of projects co-organised and co-financed by CLARIAH and the eScience Centre through the Accelerating Scientific Discovery in the Arts and Humanities (ADAH) Call. [TICCLAT](#) (Text induced Corpus Correction and Lexical Assessment Tool) by **Martin Reynaert**; [NewsGac](#) (News Genres. Advancing Media History by Transparant Automatic Genre Classification): by **Marcel Broersma**, **Eric Tjong Kim Sang & Tom Klaver**; [EviDENce](#) (Ego Documents Events Modelling: How individuals recall mass violence): **Susan Hoogervorst & Meiert Grootes**.
- 29 November 2019: Symposium on the Public Infrastructure, Radio Kootwijk. Coorganized by CLARIAH and the NL Digital Heritage Community. More information: <https://www.clariah.nl/nieuw/nieuws/symposium-over-de-publieke-infrastructuur>
- NWO, our main funder, is inventorying the research infrastructure landscape in the Netherlands as a preparation for new rounds of funding. All existing and newly planned infrastructures have to describe the current state of affairs of their infrastructures and and initial indications for plans for the future

- 21 November, Publication of the first National Agenda for Artificial Intelligence <https://www.nwo.nl/en/news-and-events/news/2019/11/first-national-research-agenda-for-artificial-intelligence.html> . NLP is explicitly mentioned and many NLP researchers contributed to the agenda.

## Norway

- No report

## Poland

- Partial migration of web services to kubernetes cluster ( via Rancher)
- Participation in the conference: "Digital meetings with monuments" at the University of Wrocław (25.11.). Demonstration of an example of cooperation between CLARIN and art historians (Polish Institute of World Art Studies).
- CLARIN-PL Quality Commission meeting and review and summary of project progress for 2019
- Participation in the “2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing” (EMNLP, 3-7 November, Hong Kong). The title of the paper:Multi-Level Sentiment Analysis of PolEmo 2.0: Extended Corpus of Multi-Domain Consumer Reviews.

## Portugal

- Launching of PORTULAN CLARIN K-centre CLARIN Knowledge Center for the Science and Technology of the Portuguese Language
- Participation in the workshop by CLARIN ERIC on Universities @ CLARIN, Utrecht, 4-5 November 2019
- Meeting with representatives of the Ministry of Science and Technology to update about the current situation
- Fixing newly-discovered bugs
- Continued integration of tools into our workbench

## Slovenia

- As part of its work on RDA Node Slovenia, CLARIN.SI participated in the Conference on Open Research Data that took place in Nov. 14-15 in Maribor. A working group for coordinating the work of research data repositories in Slovenia was established;
- Annual meeting of the CLARIN.SI Management Board to take place in December;
- 8 new entries in the repository: frequency lists from various Slovene corpora and two corpus analysis tools

## South Africa

- Holiday season: no formal activities - universities closed
- Celebration of South African English (Febr 2020)

## Sweden

- No report

### United Kingdom

- Successfully added full eduGain support to the authentication system for OTA CLARIN Dspace repository
- The following requirements have been identified for potential centres in the UK - they need technical assistance and support for:
  - making web services CLARIN-compliant
  - adding Shibboleth authentication to online resources

### 3rd Parties

### USA

- No report

## 10 AOB

- a) Foreseen date and agenda Centre Meeting: 31 March + 1 April 2020, with the SCCTC on the first day, location: Utrecht.  
Please save the dates!

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