

**Title** Minutes SCCTC 7 November 2019  
**Version** 1  
**Author(s)** LS  
**Date** 2019-11-07  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2019-1545



## Participants

Pavel Straňák (CZ), Jan Hajič (CZ), Felix Helfer (DE), Krista Liin (ES), Vanja Štefanec (HR), Martin Matthiesen (FI), Roberts Dargis (LV), Tomasz Naskręć (PL), Luis Gomez (PT), Leif-Jöran Olsson (SE), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

Excused: Matej Ďurčo (AT), Lene Offersgaard (DK)

## 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 ([CE-2019-1518](#))
3. Centre assessments:
  - a. round 14: ongoing
4. Update on taskforce-related issues:
  - a. AAI: eduGAIN discovery service
  - b. AAI: IdP blacklisting
5. Document on Google Data Search ([CE-2019-1544](#))
6. Chair of the SCCTC: request to provide informal feedback on the idea of a duo-chair
7. Status update per country/member (please provide a short bullet-wise summary in the [google doc](#))
8. Any other business
  - a. Organize this year's centre meeting in April? (to avoid overlap with LREC and DARIAH annual event)

## 2 Approval minutes and action points

The minutes of 30 September 2019 ([CE-2019-1518](#)) have been approved.

### 3 Centre assessments:

a. **round 14: ongoing**

No report. Report will follow next centre meeting.

### 4 Update on taskforce-related issues

a. **AAI: eduGAIN discovery service**

Foreseen date to launch the update is in the week of 11 November. This will be separately communicated in the Centre news of 14 November.

Feeds on eduGAIN IDPs is already available via:

[https://infra.clarin.eu/aai/prod\\_md\\_about\\_edugain\\_idps.xml](https://infra.clarin.eu/aai/prod_md_about_edugain_idps.xml)

If you have any questions, feel free to ask.

b. **AAI: IdP blacklisting**

Dieter proposes to discuss the implementation with the AAI taskforce. Martin Matthiessen already emailed the taskforce and is waiting for a reply. The implementation timeline needs to be in place in this year. There needs to be a final agreement by the meeting of December, so implementing process can start right afterwards.

### 5 Document on Google Data Search (CE-2019-1544)

See: [https://office.clarin.eu/v/CE-2019-1549-report\\_google\\_dataset\\_search.pdf](https://office.clarin.eu/v/CE-2019-1549-report_google_dataset_search.pdf)

- The SCCTC's point of view on this is not in line with the original proposal put forward by Dieter in document CE-2019-1544.
- There was a clear majority in favour of exposing structured metadata to search engines like Google. The major motivation for this is the significant amount of traffic that a good listing on Google can generate.
- Many centres would prefer this metadata to be also included in the VLO landing pages, as to avoid additional implementation work on the side of their repository. It was pointed out that:
  - Doing this would mainly mean that the record pages at the Google Data Set search would display 2 links: one to the original repository and one to the VLO.
  - Currently, only a small subset of the record landing pages in the VLO (26.000 out of 1 million) are crawled by Google, despite several efforts to increase this crawling rate. This will be further investigated, but in the end there is no guarantee that Google will be crawling the whole VLO.
  - If the aforementioned issue can be solved, and if the VLO exposes the metadata for all its records as structured metadata that the Google Data Set search can process, this will be done for all the metadata providers.
    - Providing an opt-out function per repository is technically feasible but complex and costly: changes would need to be made to the centre registry, the harvester and the VLO importer.

- Everyone agrees that the use of embedded metadata does not affect the requirements for B- and C-centres.

Follow-up action points are:

- Dieter:
  - setup a short poll to be sent to all centres, to check their preferences
- Twan & Dieter:
  - Provide a short report on the current crawling issues [done, see document above]
- Pavel:
  - discuss the issue with Google contacts if the current efforts to get the indexing working do not succeed.

## 6 Chair of the SCCTC: request to provide informal feedback on the idea of a duo-chair

Martin Mathiessen is interested in co-chairing and proposes a system could be set-up for this. As co-chair the amount of work will be reasonable, and it ensures the connection with Dieter is sustained. There is no model for this in the bylaws. First an informal agreement is needed from the SCCTC committee. Both the SCCTC committee and the current co-chair Jan Haijc have no objections against the co-chairing model. DVU will explore the formalities for this and will discuss the idea with the BoD.

## 7 Status update per country/member

### Austria

- PARTHENOS ending (presentation on Wednesday)  
Resource Discovery Application: <https://parthenos.acdh.oeaw.ac.at/>

### Bulgaria

- [PARTHENOS Workshop for CEE countries](#)

### Croatia

- Co-organised the [NooJ for Digital Humanities Workshop, 23-25 October 2019](#).

### Czech Republic

- Organized first of three tutorials to all of Digital Humanities audience in the Czech Republic, as the first joint event under the new LINDAT/CLARIAH-CZ project
- New website ("front matter") for the LINDAT/CLARIAH-CZ project
- Finishing organization of the Nov. 19 National Day of Research Infrastructures in the Czech Republic, organized at LINDAT/CLARIAH-CZ premises in Prague
- Finishing arrangements for the CLARIN GA in Prague

### Denmark

- Looking for a new (web) developer, will announce it to CLARIN when official out
- Working on FAIRifying data - licensing of data is an issue - this is a collaboration with DARIAH in Denmark.

## Estonia

- Promoting CLARIN services and resources at the meeting of council of [Centre for digital humanities and information society](#) at University of Tartu
- Conference of [National Programme of Estonian Language Technology](#) in Tallinn on Nov 5th, presenting also CLARIN web-services and resources.
- Testing the local FCS aggregator [RABA](#) - Estonian written, spoken corpora and dictionaries (should be fully stable by end of the year)
- Transferring to KORP v7
- Setting up FCS endpoints for Clarin has been delayed, will be connected to the KORP7
- Early preparations to move to Kubernetes

## Finland

- Preparing transition to new Puhti HPC system
- Installing new version of Turku UD Parser
- Language Bank Rights update coming up
- Backlog review coming up.

## France

- Submission of a proposition for a “CLARIN-FR tutorial” for LREC 2020
- Organization of an “Information Day about European Research Infrastructures” for the French community - Paris 13 November
- Integration of members of “GDR TAL” (i.e. Research Group on NLP technologies) in our CLARIN-FR committee

## Germany

- good progress on updating central monitoring to Icinga2
  - probes have been migrated
  - map functionality is back
  - new probes (Switchboard services, SSL-certificate checks)
  - already live but still few open issues being worked on
  - next steps
    - national level: integration of DARIAH-DE/CLARIN monitoring
    - probably: future hosting in Göttingen at GWDG (close project partner)
- AAI integration in CLARIAH-DE: no actions needed but maybe DARIAH-DE will join SPF
- Metadata Integration of DARIAH-DE-resources into VLO and CLARIN-resources into Generic Search:
  - DARIAH-DE to be featured in centre registry to allow for VLO-integration
  - Mappings from most prominent CMDI-schemas created for GS-integration
  - goal: at least 75% of metadata covered

## Greece

- No report
- Started development of CLARIN:EL v2, which will include new functionalities, new front-end for improved user experience.
- Planning the extension of the Greek network through new members.
- Production of new dissemination materials (two new video tutorials on the use of the CLARIN:EL infrastructure  
<https://www.youtube.com/channel/UCExRiPpck8kwqnTtZxg8dQQ>)

## Hungary

- No report

## Iceland

- CLARIN-IS has been registered as a C-Centre
- Workshop on October 16 to celebrate the launching of a National LT Program. CLARIN-IS among the organizers. Steven Krauwer, Bente Maegaard and Kadri Vider belong to Expert Panel.

## Italy

- ILC4CLARIN
  - Preparation for CLARIN Conference

## Latvia

- Started to step up local clarin-dspace repository.
- Participated in “Workshop: CLARIN@universities”.
- Meeting with representatives of the Ministry of Education and Science regarding current status and possible additional funding

## Lithuania

- We have received official answers from the prospective partner institutions (VU, KTU, MRU, BPTI) to renew CLARIN-LT consortium agreement and are now working on the documentation to be signed by all parties.
- Participated in “Workshop: CLARIN@universities”.

## The Netherlands

- Nothing to report beyond what was reported wrt CLARIAH NL AAI solution.

## Norway

- Positive decision from the Research Council of Norway to fund CLARINO+ which will upgrade the CLARINO centres. The three-year project will start in the spring of 2020.
- Conducted awareness presentation about CLARINO+ and CLARINO services to the University of Bergen Library
- There has been some progress in Norway on making CLARIN services in bulk accessible to researchers (initially in Bergen)
- The CLARINO Bergen Repository received the CoreTrustSeal on Nov. 13 and has thereby completed its re-certification as a Type B centre.

## Poland

- Deployed new dedicated gpu server for cmc-tagger
- 10th edition of Workshops: CLARIN-PL in Research Practice (Kielce, Poland, 16-17.10.2019)
- Active participation at the Leipzig CAC: workshop paper + poster phd student (Wiktoria Kowszewicz Mieleszczenko);
- Extension of "WebSty" application with new elements: doc2vec - new version of the service with word2vec binary models from the fastText package; websim - service used to analyze semantic similarities between documents in the corpus; clusterer - service used for agglomeration grouping.

- Extension of “LEM” application: fog - a service that indicates the index of text readability

### Portugal

- Continued work on integration of new processing services into the PORTULAN workbench
- Improving usability of already integrated tools based on early users feedback
- Participation in “Workshop: CLARIN@universities”

### Slovenia

- Working on our CTS application and B Centre reassessment
- Parla-CLARIN recommendations for encoding Parliamentary corpora [presented](#) at the [TEI conference in Graz](#)
- First hardware received by JSI from the Ministry/EU grant: new server for CLARIN.SI services (e.g. repository)
- New entries to repository: [CMC training corpus Janes-Tag 2.1](#), [ŠUSS archive of questions and answers about the Slovenian language \(1998-2010\)](#), [Collocations Dictionary of Modern Slovene KSSS 1.0](#)

### South Africa

- Two more language celebrations (Siswati, isiNdebele) as part of SADiLaR’s contribution to UNESCO Year of Indigenous Languages programme
- Workshop on computational tools for language researchers and language practitioners (Durban)
- Took part in Parthenos Impact event (Rome)

### Sweden

- Openshift deployment finally coming closer, new Swe-Clarín web site close to first version, new resources prepared, Tour stop at Högskolan i Dalarna
- Planning further for user stored corpora
- Held the annual autumn workshop of the National Language bank with DH focus in Uppsala

### United Kingdom

- OTA DSpace repository launched 7th October (<https://ota.bodleian.ox.ac.uk/>). Preparations are now under way to move towards and application for CLARIN B-Centre accreditation.
- UK Research and Innovation Roadmap Programme was published at <https://www.ukri.org/research/infrastructure/>, including CLARIN in the ‘Opportunities to grow our capacity’ document; CLARIN-UK is also listed in the ‘InfraPortal’ at <https://www.infraportal.org.uk/>

### 3rd Parties

#### USA

- No report

## 8 AOB

Dieter suggest having the face-to-face centre committee meeting in April 2020 instead of May 2020 to avoid overlap with the LREC and DARIAH annual events and EOSC-hub week in May.

Estonia is unavailable on 23 and 24 April 2020. But there are no further restrictions. A doodle is sent around to see what days are most convenient. Estonia 23 24 April not available, avoid that date.

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