

**Title** Minutes SCCTC 5 September 2019  
**Version** 1  
**Author(s)** LS  
**Date** 2019-09-12  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2019-1497



## Participants

Matej Ďurčo (AT), Pavel Stranak (CZ), Lene Offersgaard (DK), Christophe Parisse (FR), Vanja Štefanec (HR), Martin Matthiesen (FI), Riccardo Del Gratta (IT), Daan Broeder (NL), Hemed Al Ruwehy (NO), Luis Gomez (PT), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

## 0 Action points

#	Action	By whom	By when
1	Actions for Shibboleth: Technically look how we can bridge OpenID Connect Service Providers to SAML Identity Providers. Daan will report in Leipzig ( <i>will be added to agenda</i> ).	Daan	Leipzig F@F meeting
2	Check which centres want to submit for assessment in the upcoming round (22 October)	Dieter	Before Leipzig meeting

## 1 Agenda

1. Agenda
2. Approval minutes last meeting & action point status ([CE-2019-1469](#))
3. Centre assessments:
  - a. round 13 ongoing (only CLARINO)
  - b. round 14 (Autumn 2019): Deadline 22 October 2019
4. Update on taskforce-related issues
5. Preparation of the face-to-face meeting in Leipzig: collecting agenda points
6. Status update per country/member (please provide a short bullet-wise summary in the [google doc](#))
7. Any other business

## 2 Approval minutes and action points

The minutes of 20 June 2019 are approved.

## 3 Centre assessments:

- a. round 13: ongoing (only CLARINO)

Update by Lene: We only have CLARINO for this round and we are waiting for them to answer some questions.

Jozef and Lene will go through the assessment and will send it out to the to the SCCTC for review. The checklist and questions asked will be evaluated. The plan is to have this summarized before the meeting in Leipzig.

#### **b. round 14 (Autumn 2019): Deadline 22 October 2019**

Centre planning to submit for this round:

- CLARIN Slovenia Centre

Other centres that might join this round:

- Matej: Graz colleagues are currently working on a submission. Whether they will be submitting for the upcoming round will be communicated soon.
- Christophe: CTS is not finished yet for the ORTOLANG centre, should this be done first before submitting for round 14?
  - Dieter: the assessment submission can be done, as long as the CTS procedure has been initiated. (Initiating the CTS procedure is a requirement for submitting for an assessment round)

Dieter will ask around if there are more centres planning to submit by asking all-centres.

#### **Updated version B-centre checklist**

Dieter and Lene are working on an updated version of the B-centre checklist, based on version 7.2 used by CLARINO (the version we agreed on during our last face-to-face meeting in Utrecht.)

Main changes:

- DOIs now part of the persistent identifier section
- an example table added to section 2.e
- the "how" blocks have been merged in the "details" sections
- the attribute aggregator is now mentioned in section 5
- URL of the Curation Module is updated
- curl examples added for the content negotiation

The plan will be sent to SCCTC for review and approval so this updated version can be presented in Leipzig.

## **4 Update on taskforce-related issues**

### **1. PID usage overview**

This document is related to the PID questionnaire and it is based partially on information we got from EPIC and it was an interesting exercise to have an overview of the centres, how they are using PIDs and from who they are getting the handles and who is having which prefix.

A few questions are raised:

1. Dieter: Language bank of Finland uses a PID from SURFsara, is that correct?  
Martin: Was surprised too. EPIC is sharing the prefix with an EUDAT node in Finland and they use PIDs hosted by EPIC.
2. Related to this procedure of issuing a non-intentional shared prefix. It turned out that by making this overview we found out that prefix 21.11115 is both used by Vienna (ACDH-ARCHE) and Lisbon (PORTULAN). Both centres were not aware and neither was EPIC. EPIC made a human mistake and this indicates that there are no waterproof procedures. Dieter marked this to the EPIC management and technical boards. The solution for the impacted centres can either be accepting the shared prefix or for PORTULAN (who has fewer PIDs) to ask for new prefix and set a forward from the already provided PIDs to the newly assigned prefix.

Daan: This happened between two CLARIN centres, but can it also be that this happens between CLARIN centre and outsider.

Dieter: not very probable. CLARIN centres are registered as one customer named CLARIN. Theoretically it could be possible but then it would have been noted by EPIC, this mistake happened because we are seen as one customer.

## 2. **Proposal to blacklist Identity Providers that are not releasing attributes** ([CE-2019-1487](#))

Frequently, CLARIN ERIC receives support request from researchers trying to use federated login to reach resources in CLARIN centres. In many cases, the login problems are caused by Identity Providers not releasing attributes that can uniquely identify a user (ePPN or ePTID). In case of attribute release problems, the administrators of the respective Identity Provider are contacted by the Service Provider operator and/or CLARIN ERIC, with some background information and the polite request to release the necessary attributes. In case an Identity Provider is contacted about lacking attribute release, and it turns out to be consciously unwilling to release attributes that enable to uniquely identify a user, the next suggested step in the proposal is that CLARIN should blacklist the Identity Provider.

### **Comments:**

Daan: we need to be explicit why they are blacklisted

Dieter: The list is not huge, and we only blacklist them when they give no response at all. When blacklisted the provider receive an email about it and are also given the opportunity to be white listed again.

## 5 **Preparation of the face-to-face meeting in Leipzig: collecting agenda points**

Add to agenda Leipzig F2F meeting:

- Action 1. *Technically look how we can bridge OpenID Connect Service Providers to SAML Identity Providers.* Daan will report in Leipzig
- Suggestion by Lene to brainstorm about things recommended by the repository, nice services that are offering that people use, are popular and work well.

The agenda for the F2F meeting will be distributed one week before the meeting. If you have anything to add, please inform Dieter.

## 6 Status update per country/member

### Austria

- Testing integration of Switchboard with ARCHE
- Resolved issues with harvesting of ARCHE to VLO
- Test implementation of LinkChecker based StormCrawler
- GAMS, Graz preparing for 14th round of Centre Assessments

### Bulgaria

- [International Conference "Recent Advances in NLP"](#)
- [3rd Conference on Biographical Data in a Digital World](#)

### Croatia

- Nothing to report

### Czech Republic

- CTS certification received, B-centre re-certified
- Preparing autumn tutorials (September–November) on data and services for humanities
- Adding new machine translation models in progress (en-de, de-en, ru-en, a en-ru)

### Denmark

- CTS certification received in July for the CLARIN-DK repository
- We are planning server OS upgrades and IP changes because UCPH IT demands for updates as a part of the change in our information security policy
- Informing researchers on the faculty about the option to FAIRify language data in CLARIN.
- Writing research applications.

### Dutch Language Union

- Unfortunately, the Dutch Language Union is no longer a CLARIN member. We are trying to resolve this, but it will not be possible to be a member in 2019 and 2020.

### Estonia

- No report

### Finland

- [Workshops](#) held at RDHum 2019 in Oulu on the Language Bank and Preserving AV Data
- Investigating [Oxford Common File Layout](#)
- CSC launched new super computer "Puhti"

## France

- We have been working on improving the metadata in our various centers so that we will be better referenced in the VLO. This has been made possible because we have now a better interaction between people using the VLO and people working in the centers.
- We also had a big work on juridical document and information related to contact with international projects and changes induced by the RGPD. Making juridical documents is really a slow process because this is outsourced, so it really takes time.
- We are working on the CoreTrustSeal validation for Ortolang and we would like to go through the CLARIN certification process as soon as possible.

## Germany

- No report

## Greece

- CLARIN:EL participates with demos, videos and presentations in the 84th International fair of Thessaloniki, in the stand of the General Secretariat for Research and Technology (7-15 Sept. 2019).

## Hungary

- No report

## Iceland

- We got a server for the dspace repository this month and are setting up the system. Hope we will be ready at the end of the month.
- A contract on the implementation of the Icelandic National Language Technology Program was signed on September 4. The role of CLARIN-IS in hosting and distributing the resources being built within the program is especially mentioned in the contract.
- CLARIN-IS is co-organizing a workshop on October 16 to celebrate the launch of the National LT Program.

## Italy

- ILC4CLARIN
  - Preparation for Leipzig: new services and dockers, currently under test

## Latvia

- No report

## Lithuania

- No report

## The Netherlands

- Dutch CLARIN centres IvdNT and RU as well as Tilburg University will use one exploitation environment for the OpenSONAR+ corpus search application, they will investigate having one fore-front and two backends setup. Current setup will be three separate instantiations of OpenSoNaR+ to spread the processing. OpenSoNaR+ gives access to the SoNaR corpus and the Spoken Dutch Corpus, incl. audiofiles, through CQP and a range of different user interfaces of increasing complexity
- Jan Odijk teaches a course in Utrecht to research masters in linguistics on Digital Methods in Linguistics in which CLARIN will play a role, inter alia WebLicht (DE), AutoSearch (NL), and PMLTQ (CS), and perhaps also Federated Search as one of the many instances using CQP as a query language

### Norway

- Positive decision from the Research Council of Norway to fund CLARINO+ which will upgrade the CLARINO centres. Not a big grant and only for three years (2020-2022), but at least Norway will remain a member and will sign an agreement with CLARIN ERIC.
- CLARINO Bergen Centre has revisited preservation policy, maintenance guarantee etc., resubmitted application for CTS certification and applied for recertification as a CLARIN type B centre.

### Poland

- Preparations for the workshops at UMCS Lublin in September
- Workshops at the Summer School of Sociology in Krakow
- Application for CTS certification has been submitted (for the 3rd time)
- Organization of Global Wordnet Conference (Wrocław July 2019)

### Portugal

- Dealing with hardware problems
- Faculty closed for summer holidays

### Slovenia

- We are in the process of buying new hardware from the Ministry/EU grant: supercomputer for neural computing, new server for CLARIN.SI services (e.g. repository), new server for language-related services at the University of Ljubljana
- New entries to repository: [Corpus of Informatics DSI 5.0](#), [DigiLing e-Learning Hub: e-Courses for Digital Linguistics](#), [Developmental corpus ccSolar 1.0](#)
- Participation in RDA Node Slovenia project (led by University of Ljubljana)
- Supporting the TSD conference taking place in September in Ljubljana, cf. <https://www.kiv.zcu.cz/tsd2019/index.php?page=partners>

### South Africa

- Currently engaged with B-Centre Certification Documentation
- Currently engaged processes required for CoreTrustSeal validation

### Sweden

- No report

## United Kingdom

- OTA CLARIN-DSpace repository launched at <https://ota-ga.bodleian.ox.ac.uk/repository/xmlui/>
- CO-organizing with BBAW CLARIN Workshop 'NLP for Historical Documents' 9-11 September, with participants from Oxford, Glasgow, Lancaster

## 3rd Parties

## USA

- No report

## 7 AOB

1. Still searching for SCCTC Chair, if you are interested or know about a colleague who might be interested let Dieter and Franciska know. **Deadline for candidacy is 16 September 2019.** It would be nice if someone else can take over from Dieter who has been doing it for several years. Please note that Dieter will remain Technical Director, the role of SCCTC chair and the directorship has been decoupled in the new CLARIN ERIC statutes.
2. Delegation from SADiLAR will visit Europe from 4-6 December 2019 (probably Paris) and would like to visit CLARIN centres. Are you interested in collaboration? Or would you like to meet SADiLAR high level people from SADiLAR? Please inform Dieter. More details will follow.

**Next meeting:** F2F meeting at 11:00 -13:00 Leipzig during CLARIN2019. A doodle for the next virtual meeting will be sent out shortly.