

Title Minutes SCCTC 21 May 2019
Version 1
Author(s) LS
Date 2019-05-21
Status Approved
Distribution Centre Committee
ID CE-2019-1451



Participants

Felix Helfer (DE), Jan Haijc (CZ), Lene Offersgaard (DK), Vincent Vandeghinste (DLU), Krista Liin (EE), Vanja Štefanec (HR), Martin Matthiesen (FI), Samúel Thórisson (IS), Riccardo Del Gratta (IT), Roberts Dargis (LV), Daan Broeder (NL), Koenraad Desmedt (NO), Tomasz Naskret (PL), Leif-Jöran Olsson (SE), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

Excused: Nicolas Larrousse and Christophe Parris (FR)

The SCCTC has no objections to having non-members (observers) at this f2f meeting.

0 Action points

#	Action	By whom	By when
1	Actions for Shibboleth <ol style="list-style-type: none"> 1. Make it clear what centres need to Shibbolize 2. Intensify support for Shibboleth, e.g. with a tutorial 3. Technically look how we can bridge OpenID Connect Service Providers to SAML Identity Providers. 	DVU	asap
2	Double check with ACDH-OEAW what their position on the EUDAT offer is	DVU	asap

1 Agenda

1. Agenda
2. Approval minutes last meeting & action point status ([CE-2019-1429](#))
3. Centre assessments:
 - a. round 13 (Spring 2019): procedure
 - b. round 14 (Autumn 2019): 22 October as suggested deadline
4. New version of B-centre checklist, background documents:
 - a. Current version ([CE-2013-0095 v6](#))
 - b. Proposed changes: for discussion (see [dropbox folder](#))
 1. Suggestions by Lene without track changes (CE-2013-0095-v7_FAIR)
 2. Suggestions by Lene with track changes and some edits by Dieter (CE-2013-0095-v7_1)
5. [Questionnaire: reasons not to setup a SAML service provider](#) (please fill it in before our meeting!)
6. SCCTC Governance
 - a. Updates of the bylaws ([CE-2012-0021 v3](#))
7. The EUDAT B2SAFE offer ([CE-2019-1397](#)) and performance indicators ([CE-2019-1397](#))
8. Status update per country/member (please provide a short bullet-wise summary in the [google doc](#))
9. Any other business
 - a. [CLARIN is hiring a System and Software Engineer](#)
 - b. [EOSC early adopter programme](#) [for information]

2 Approval minutes and action points

The minutes of 17 April 2019 are approved.

3 Centre assessments:

- a. round 13 (Spring 2019): procedure

The 13th assessment round will start soon, only CLARINO will participate in this round. Early beta checklist ready in July for CLARINO to fill in

- b. round 14 (Autumn 2019): 22 October as suggested deadline

The Slovenian centre also needs to re-apply but they were not informed on time to participate in the 13th round, so they will participate in the 14th round.

4 New version of B-centre checklist, background documents:

- a. Current version ([CE-2013-0095 v6](#))
- b. Proposed changes: for discussion (see [dropbox folder](#)):
 1. Suggestions by Lene without track changes (CE-2013-0095-v7_FAIR)
 2. Suggestions by Lene with track changes and some edits by Dieter (CE-2013-0095-v7_1)

Document CE-2013-0095-v7_1 is being discussed. The yellow parts in the documents are comments and will not necessary stay in.

Lene will create a proxy document where the terms like FAIR principles, EOSC and CTS and others (if any) are further described. Each term will start with a foreword explaining what CLARIN is doing with it. By creating a proxy document the checklist stays clean and does not get too messy with extra information on terms.

Underline that all the things happening in EOSC and SSH related initiatives are monitored by everyone in EOSC and CLARIN is seen as one of the most advanced SSH initiatives that has their certification in position. We should keep this up.

Comments concerning CLARIN B Centre Checklist (CE-2013-0095-v7_1)

Koenraad suggests referring to the statutes in a footnote about the linked third party.

2.c visibility of connection to CLARIN (updated)

Under the header 'details' the last sentence "... if this requirement is not met..." can be taken out. In any case if a centre wants to be recognized by CLARIN a centre needs to show they are connected to CLARIN, if not they should not apply for B-centre.

2.e Details about resources and services provided (updated)

Under the header 'Check procedure' it is suggested to rephrase the centre statement. We need to clearly describe what is required of a centre, why it is important to be active and how one can be active (DVU: this is a good to describe in the why and how block too). We expect more than just putting up a centre, CLARIN is a networking organization.

It is important to have a minimal requirement for centres, for instance a minimal should be that countries appoint a formal member for the centre committee, if countries are not nominating someone, we look at this case to case. Other requirements can be attending the meetings, conferences etc. Jan comments that if they have funding/financial issues we could accept some exceptions. Koenraad suggest changing the term funding into financing, this allows the option of self-financing not just from funding. Dieter agrees that this can be included in the funding situation part. An option is to rephrase the sentence under header 'Check procedure' to 'check the financing (instead of check the funding).

2.e Details about resources and services provided (updated)

Under the header 'Details' Koenraad asks if we should say metadata here. Dieter is not sure if the sentence needs to be kept in the document.

Daan suggests asking for a statement on why they keep data inside and why they keep data outside the VLO. He also mentioned that the FAIR part can also go into the proxy document.

2.f Registration in centre registry

Re3data.org harvests the centre registry and it is important to mention this, so people know they do not also need to fill Re3data.org.

Daan proposes to add a statement that information is going to be exposed in other sources and that it is openly available. Koenraad suggest adding this part in the introduction.

3.a Data offering & IPR (unchanged)

Dieter will refer to the CTS requirement on the availability of the services. Lene mentions we are more specific about the requirement then CTS.

3c. Licenses on data and metadata (new item)

Jan: Most metadata are not creative work. If metadata is assigned a license you make a statement that it is a creative work. Koenraad: you can attach a license to metadata. If there is a license it should allow copying.

Martin: Metadata should be available, and nobody should hide it because of a license. A license is expresses what can you do with the data.

Jan: Give it more time to discuss it and ask CLIC, it is an important topic.

Dieter: We can discuss a license for metadata later on if we find the additional time.

6.d User access to the repository (updated) (should this section still be part of the "6 Metadata" section?)

The question is if we require a nice front end for centres.

Lene: we should agree on something here to have a clear view about it.

Dieter: we can include that a centre ideally provides a nice front end, in case where it is not appropriate for the centre they should at least give a link to the VLO and indicate that there data is searchable via the VLO.

Planning of final B- Centre Checklist

- Approval of checklist by the centre committee in early June 2019, so CLARINO colleagues can submit it.
- Period afterwards we make sure we have a complete b-checklist by September 2019

5 Questionnaire: reasons not to setup a SAML service provider

See: [Questionnaire: reasons not to setup a SAML service provider](#)

There were 15 responses to the questionnaire.

First question: I have no expertise in-house to shibbolize a specific web application

The general conviction is that a lack of inhouse expertise is not an argument to not set up a SAML service provider. If a centre has not enough funding they can decide to make use of the CLARIN Mobility Grant and invite an external expert to help. Setting up a SAML SP is part of being a CLARIN centre.

Second question: We have not enough human resources to setup a service provider

Same argument concerning the mobility grants goes for this point, an external expert can be asked to help via the Mobility Grant.

Martin: concerning GDPR, a lot of data can be made available. Shibboleth can ensure, that the research data sharing is better controlled. Centres that say they only have open data, could have personal data in there that cannot be put out there

There are three action points for Shibboleth (**AP1**):

- a. Make it clear what centres need to Shibbolize
- b. Intensify support for Shibboleth, e.g. with a tutorial
- c. Technically look how we can bridge OpenID Connect Service Providers to SAML Identity Providers.

6 SCCTC Governance

- a. Updates of the bylaws ([CE-2012-0021 v3](#)).

The SCCTC has formally approved the updates of the SCCTC bylaws (version 3) without any objections.

New chair

Dieter informs the SCCTC that the committee is looking for a new chair. It is good to change chair, to make sure everything is not too much dependent on the same person, it is good to have an overall rotation in the committee. If you are interested or know of a good candidate in chairing this committee, please inform Franciska or Dieter. The election of a new chair could start at the CLARIN2019 conference in Leipzig.

Franciska: one addition to this subject. As long as Dieter is BoD member, he will also be attending the SCCTC.

7 The EUDAT B2SAFE offer (CE-2019-1397) and performance indicators

EUDAT B2SAFE offer [CE-2019-1397](#)) and performance indicators (https://office.clarin.eu/v/CE-2019-1397-2019_01_B2SAFE_CLARIN_performance_indicators.pdf)

Daan: quota prices are on the condition that CLARIN ERIC is brokering this. Channeled through CLARIN ERIC centres can receive this discount, but there should be enough interest by the centres.

For now, only Sweden is interested in the offer. In case of interest from a single centre, CLARIN ERIC in the middle as broker is not necessary, so Språkbanken will get in contact with EUDAT itself.

Most centres present at this meeting are currently using their national services or trough other kind of cooperation without any or much costs, see their comments below:

Norway: Offer not interesting for us.

CLARIN_NL/Meertens: SurfSARA is more or less equal with EUDAT. The operations manager is not interested in the EUDAT offer.

Estonia: Already cooperating with the national research network, inside the country. For now, not interested.

Finland: Uses national service for free.

Italy: Uses national service for free.

Jan informed us that Amazon is cheaper but has different services. With Amazon you have to pay to download the offline data. It is cheap to store in the cloud but expensive to get it back again. The Amazon Glacier pricing for storage is \$0.0045 per GB / Month = 54\$/TB/Year. Download (the fastest available): \$40/TB.

Dieter will double check with Austria/ ACDH what their position on the EUDAT offer is. **AP2**

Alexander Koenig asks if it is possible to share the list trough the all-centres mailing list. Sending the EUDAT offer to the public all-centers mailing list is sensitive. We can ask who is interested via Centre news and send the list to the interested person individually.

8 Status update per country/member

Austria

- Nothing to report

Bulgaria

- No report

Croatia

- Nothing to report

Czech Republic

- Work in progress on statistics of API calls for all services
- Website now accessible also at <http://lindat.cz>
- Universal Dependencies 2.4. released for [download](#) and [search](#)

Denmark

- Working on text for Tour de CLARIN
- Planning workshops and activities for rest of the year
- Looking for Danish research groups from Humanities that like to FAIRify their data

Dutch Language Union

- Together with Dariah-BE and the Impact Centre of Competence, we (CLARIAH-VL) held the [DATECH 2019](#) conference in Brussels on Digital Access to Textual Cultural Heritage.

Estonia

- 25th of April at [18th Annual Conference of Applied Linguistics 2019](#), presentation and demo about CELR and CLARIN web-services useful to Estonian researchers.
- Ongoing work debugging and updating KORP and RABA.

Finland

- Submission of new FIN-CLARIN funding proposal in mid-May
- Kielipankki Roadshow in Eastern Finland
- DHH2019 at the end of May
- Language Bank to receive Sensitive Data from the Institute for the Languages of Finland (Kotus)

France

- No report

Germany

- Kick-Off Meeting of the new CLARIAH-DE project: <https://www.clarin-d.net/en/current-issues/490-successful-start-to-clariah-de>
- New version of GermaNet was released: <http://www.sfs.uni-tuebingen.de/GermaNet/>
- On-going preparations of ESU courses
- Presentation and Participation at the DFG Conference on National Research Infrastructures

Greece

- Finalization of GDPR compliance of the CLARIN:EL infrastructure

Hungary

- HunCLARIN 3rd Roadshow in Pécs 2 MAY 2019: New Approaches and Advanced Tools in Computational Linguistics and Data Driven Methods (<http://clarin.hu/en/content/events>)

Iceland

- No report

Italy

- ILC4CLARIN
 - Internal wrap up for the meeting. (FCS and resources to offer)
 - Plan to move toward next CLARIN-DSPACE release (use docker?)
 - Software to wrap and make linguistic services available through WebLicht and LRS (Opener Tokenizer and Freeling it from Panacea). Code is here: <https://github.com/cnr-ilc/linguistic-tools-for-weblicht>

Latvia

- Practical seminar about Latvian treebank for CLARIN user community on May, 9
- Preparation to CLARIN NCF meeting in Riga
- Preparation of CLARIN 2018 post-proceedings
- Participation CLARIN-LT evaluation meeting organized by MOSTA on May, 10

Lithuania

- The Research Council of Lithuania is updating the Lithuanian Research Infrastructure (MTI) Roadmap 2019, which for the next five years will provide strategic directions for the development of RIs, long-term R&I policy and development priorities. CLARIN-LT is preparing an application to be further included in the Lithuanian RI Roadmap 2019.
- A new version of the Lithuania's participation in CLARIN ERIC plan for 1919-2021 has been prepared and will be shortly sent to the Ministry of Education, Science and Sport with a funding request.
- Cooperation with Research and Higher Education Monitoring and Analysis Centre (MOSTA) of Lithuania in preparing evaluation guidelines of long-term membership in international research infrastructures.

The Netherlands

- Data Science for Generating Linguistic Knowledge, Lecture by Jan Odijk at the Data Science Meets the Humanities Seminar, Leiden, 12 April, 2019
- Lex Heerma van Voss participated in an EOSC Sounding Board meeting

Norway

- Corpora searchable in CLARINO Corpuscle (<http://clarino.uib.no/corpuscle>) now cover 18 languages: Abkhazian (1) · Bulgarian (1) · English (17) · Faroese (1) · French (1) · Georgian (7) · German (1) · Lithuanian (1) · Mingrelian (1) · Norwegian (3) · Norwegian Bokmål (14) · Norwegian Nynorsk (7) · Old Georgian (1) · Old Norse (2) · Scots (1) · Slovenian (1) · Spanish (4) · Svan (1). Some corpora have restricted use.
- Treebanks searchable in CLARINO INESS (<http://clarino.uib.no/iness>) now cover 84 languages: Afrikaans (2) · Akkadian (1) · Amharic (1) · Ancient Greek (to 1453) (15) · Arabic (10) · Armenian (1) · Bambara (1) · Basque (7) · Belarusian (2) · Breton (1) · Bulgarian (8) · Buriat (2) · Catalan (5) · Chinese (11) · Church Slavic (9) · Classical Armenian (1) · Coptic (3) · Croatian (8) · Czech (21) · Danish (9) · Dutch (12) · English (31) · Erzya (1) · Estonian (8) · Faroese (2) · Finnish (18) · French (19) · Galician (9) · Georgian (7) · German (21) · Gothic (7) · Hebrew (7) · Hindi (8) · Hungarian (10) · Icelandic (3) · Indonesian (10) · Irish (7) · Italian (15) · Japanese (8) · Kazakh (5) · Komi (2) · Korean (4) · Latin (22) · Latvian (6) · Lithuanian (2) · Maltese (1) · Marathi (2) · Modern Greek (1453-) (8) · Nigerian Pidgin (1) · Northern Kurdish (2) · Northern Sami (27) · Norwegian (5) · Norwegian Bokmål (41) · Norwegian Nynorsk (15) · Old English (ca. 450-1100) (5) · Old French (842-ca. 1400) (2) · Old Norse (4) · Old Russian (20) · Persian (7) · Polish (31) · Portuguese (19) · Romanian (8) · Russian (14) · Sanskrit (3) · Serbian (2) · Slovak (4) · Slovenian (12) · Spanish (13) · Swedish (16) · Swedish Sign Language (3) · Tagalog (1) · Tamil (6) · Telugu (2) · Thai (1) · Turkish (8) · Uighur (4) · Ukrainian (4) · Upper Sorbian (2) · Urdu (5) · Vietnamese (4) · Warlpiri (1) · Wolof (3) · Yoruba (1) · Yue Chinese (2)
- The Research Council decision on a possible upgrade of CLARINO has been postponed to September 2019.
- Koenraad De Smedt participated in the RDA Language Data Interest Group session at RDA Plenary 13.

Poland

- Preparations for the 9th Language & Technology Conference in Poznań (May, 17-19, 2019): <http://ltc.amu.edu.pl/>
- Agnieszka Dziob will represent CLARIN-PL at the Conference Culture of Language Communication (Będlewo, May, 12-15, 2019): <http://kjp.amu.edu.pl/>
- Preparations for the workshop "CLARIN-PL Tools in Psychology"; cooperation with The University of Social Sciences and Humanities (Poznań, May, 23-24, 2019): <http://badanianarracyjne.pl/karuzela/warsztaty/>

Portugal

- B-Center certification for PORTULAN CLARIN

- Implementation of PORTULAN CLARIN (repository and workbench) continued

Slovenia

- Received reports on results of the 7 projects that CLARIN.SI financed in 2018
- Received and responded to another round of requested amendments to the application for research equipment for CLARIN.SI
- Added some new entries to the CLARIN.SI repository and concordancers, the most significant being siParl, the 200 million word corpus of Slovenian parliamentary debates 1990-2018, <http://hdl.handle.net/11356/1236>

South Africa

- Introductory ELAN workshop, Grahamstown
- Workshop: Text processing tools for SA languages, Grahamstown & Bloemfontein
- [Afrikaans](#) & Setswana language celebrations
- Sponsorship and support of [Digital Humanities Assoc. Of SA \(DHASA\) conference](#)

Sweden

- Roadshow for historians (discipline focused)
- Helpdesk branding almost done
- Implemented Matomo analytics for services to make comparable statistics
- Openshift migration unfortunately delayed once again

United Kingdom

- No report

3rd Parties

USA

- No report

9 AOB

- a. CLARIN is hiring [a System and Software Engineer](#) (deadline is 29 May 2019)
- b. For information: [EOSC early adopter programme](#) (deadline for application 15 June 2019)

Next meeting: Next online meeting will take place on Thursday 20 June 10:00u (CEST).