

Title Minutes SCCTC 4 June 2018
Version 2
Author(s) LS, DVU
Date 2018-06-18
Status Approved
Distribution Centre Committee
ID CE-2018-1223



Participants

Martina Trognitz (AT), Jan Hajic (CZ), Krista Liin (EE), Lene Offersgaard (DK), Vincent Vandeghinste (DLU), Martin Matthiesen (FI), Nicolas Larrousse (FR), Riccardo Del Gratta (IT), Daan Broeder (NL), Marcin Pol (PL), Leif-Jöran Olsson (SE), Twan Goosen (CLARIN ERIC, non-member SCCTC), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes), Alexander Koenig (EURAC, non-member SCCTC)

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

0 Action points

#	Action	By whom	By when
1	Follow up with EUDAT on the price and policy of B2Safe services	Dieter (CE)	Asap

1 Agenda

Agenda point 8 “Service Inventory Parthenos Project” was added at the request of Daan Broeder (NL) and the agenda was approved as follows:

1. Agenda
2. Approval minutes last meeting & action point status ([CE-2018-1207](#))
3. Update from the assessment committee (Lene)
4. Update Core trust seal
5. Detailed report by a taskforce: AAI
6. Proposal for new centre labels and types, for discussion ([CE-2018-1204](#))
7. Status update per country/member (all) -> please add a snippet for your country [to the reporting Google Doc](#)
8. Service Inventory Parthenos Project
9. Any other business

2 Approval minutes and action points

The minutes of the SCCTC meeting of 9 April 2018 are approved.

3 Update from the assessment committee

Lene (DK): Largest assessment around is still in progress, 10 centres re-assessed and there are two new centres. Lene thanks the rest of the assessment for providing the reviews. There is no complete report to send yet, but 5 centres can be re-accepted as B-centres:

1. Berlin-Brandenburg Academy of Sciences and Humanities
2. ASV Leipzig
3. MPI for Psycholinguistics
4. Bayerisches Archiv für Sprachsignale
5. Universität des Saarlandes

For these centres there are no special comments.

Because the assessment is so large Lene suggest handling these assessments in two steps. First, send the approvals for the 5 centres that have no special comments to the SCCTC now. Second, for the other centres that are more complicated and specific, the approvals will be sent out within two weeks.

Daan (NL) requests to have a discussion about the functionality of the curation model and what it can bring to the assessments. Twan (CE) agrees to discuss it during the curation session at the centre meeting.

Dieter (CE): Make the curation module a fixed part of the assessment. Independent checking mechanism. We need to rephrase the requirements. We should automate this as much as possible and make it clear what kind of tools we use for the next assessment round.

Ricardo (IT): suggestion to also add basic server security checks (e.g. a portscan).

This leads to a wider discussion about whether the assessment procedure should include such system-administration best-practice checks, e.g. with regards to security. It is concluded that the system administrators are the experts at the centres and that they should take up this responsibility, not the assessment committee.

4 Progress analysis based on the Centre workplan 2018 ([CE-2017-1099 version 4](#))

Dieter (CE) and the SCCTC went through the workplan, to check where we stand and have a look at the points that need follow-up.

Comments on Project activities -> EUDAT

Daan (NL): EUDAT mentioned to our director that we need to pay for their services in the future. This is something for CLARIN to take up with EUDAT. (see AP 2)

Martin (FI): also have to pay EPIC for our handles.

Dieter (CE): CLARIN has an agreement with EPIC for our centres, we are offering services and you do not need to have an individual agreement with them.

Comments on Central infrastructure activities

Daan (NL) Concerning the Centre registry, maybe we should have a meeting about services that are essential for CLARIN and services where you have to trust on bigger infrastructures like e-infra-central.

Dieter (CE) Meaning looking to connect Centre registry services with for instance e-infra-central registry?

Daan (NL) Yes, so create a vision what to keep within our own scope and what support development from outside we might need.

Dieter (CE) For technical parts we need our own things.

5 Proposal for new centre labels and types, more information on the further decision process ([CE-2018-1204 version 2](#))

Dieter (CE): This document has been introduced at the NCF. It will be split in two parts:

Part 1: Discuss the relabeling of the centres (section 2. Existing Centres) this is the least controversial part. This section should be discussed more within the SCCTC. We have to make sure we have an agreement about section 2 and send it to the NCF before their next meeting that takes place on 20-22 June in Wroclaw, Poland.

Part 2: The section about the new centre types, is more controversial and needs to be discussed at a later time. This should be done before the NCF meeting in September.

It is not an automatic given that these centre types will be renamed. It is a sensitive discussion.

Archive centre (instead of B-centre)

For some countries (FI, FR, CZ, DK and SE) the term *Archive Centre* should be avoided at any cost. From a political point of view, it will create a problem as for “archives” different rules and regulations apply. Also, the term *Archive* runs the risk being interpreted as less valuable (a “dead end” for data sets).

The label *Research Data Centre* is considered to be better. For CZ and PL, there is the wish not to have a *Data Centre*, since it could lead to confusion with the existing Data Centres (usually at large computing facilities) but adding the prefix *Research* would be an acceptable solution.

Metadata centre (instead of C-centre)

Daan (NL): Since metadata is one of our pillars, if a centre declares itself as metadata centre we need to check the validity of the metadata provided. I would like to see a formalization for that.

General thoughts on the centre labeling

Many of the members think that B and C are mostly seen as grades and the question that arises is, why is a centre B or C and not A? (with B and C you expect to go to A). It is confusing for people that are not familiar to CLARIN. It might be better to have names. If letters are to be used it should not be B and C but random letters (not as grades), maybe letter that define the expertise. It should be more a label than a hierarchy.

Riccardo (IT): Collect different tags that you can add to your centre. If you are a B-centre you are also a C-centre, and other tags can be freely composed.

Jan (CZ): If the labels need to be semantically empty we might want to consider colours.

Lene (DK): The term K-centre works well, as the K stands for Knowledge – this makes sense to outsiders. The terms B and C (and D) are really not working, and it is very difficult to get people to understand what they mean.

Martin (FI): Agrees with Lene. Keeping the letters and changing them is a better idea than naming it, by naming it you are narrowing the scope. For instance “D” as Data centre is wider and entails a lot of things. Best to rename the labels and keep the centres as they are.

Jan (CZ): D might be a good name but maybe we should try to avoid the first five letters and not use B or C because of the history that is known and not D as it is seen as degrading (for instance going from B to D centre).

Lene (DK): T-centre is still on the list Dieter (CE), this one will be removed.

Lene (DK) suggest making a clear suggestion to the NCF such as a fully work-out table.

Daan (NL): But there is not enough time for this during this meeting. We can convey our hopes and inform that we are unhappy with state of affairs.

Vincent (DLU): It is branding/marketing, we do not know how to convey this to the public. Maybe this is something for a marketing or communication office.

Dieter (CE): The initial proposal for new and renamed centre types are discussed and compiled together with the CLARIN communication officer. It needs input from the marketing side but also from the content side and we also need to keep in mind the politics and history of centres that already exists.

Vincent (DLU) suggest to first list the requirement instead of trying to come up with new names.

Martin (FI): B and C are the only ones that cause confusion, we can make a suggestion to change these two letters into other letters (B becomes R and C becomes M).

Daan (NL): suggest using two letters for example MO-centre for Metadata-only centre.

Dieter (CE): Will update the document with suggestions that have been made around the table. We need to get consensus on **if** we want to change things and if that is the case, on **how** we want to change things.

A small poll shows that a minority (3 people) is happy with the current B- and C-centre labels.

6 Status update per country/member (all)

Note: due to time constraints these written reports were not discussed at the meeting.

Austria

- Report here

Bulgaria

- No report

Czech Republic

- Report here

Denmark

- Started to deposit in the new DSpace repository, still solving minor issues with integration, metadata transformations, licences, branding.
- Collaboration/planning together with the Danish national infrastructure DIGHUMLAB.com
- Updated Privacy Policy

Dutch Language Union

- No report

Estonia

- Some problems with DSA/CTS that was sent back for a second time. 2 main problems:
 - SLA with our HPC – we will get on, it took time.
 - format conversion (under requirement 14 Data Reuse)
- Plans for a „FCS“ for Estonian resources (written, spoken corpora and lexicons) – adding endpoints to existing search interfaces and a user-friendly query interface.
- KORP – continue to add corpora to KORP and promoting it to SSH users.

Finland

- **In Progress**
 - Korp-cooperation / FCS still not implemented
 - Moving towards DevOps (new developer started)
 - Investigating possible co-operation with UzK on repository solutions
 - Investigating FREME as backend for Mylly.
 - HTML Version of [Mylly](#).
 - Coming up: Metashare update using Ansible/DevOps
 - Co-operation with National Library to distribute recent newspaper corpus
- **Done**
 - WWW-portal automatic installation now stable
 - New Language Bank Rights released.
 - CTS seal received.

France

- Report here

Germany

- Most centres are still waiting for first round of CTS-feedback: end of April the CTS board discussed the applications (and the feedback of the reviewers, which was not yet to be published) of all CLARIN-D centres to get a better overview of the infrastructure. Afterwards it was planned to revise the feedback documents and to then publish them to the centres.
- International Advisory Board of CLARIN-D to assemble on 19.6.2018
- Next developer meeting on 20.6.2018 - one topic will be the General Data Protection Regulation

Greece

- No report

Hungary

- No report

Italy

- CTS (formerly DSA) obtained April 2018
- FCS Still in progress

Latvia

- No report

Lithuania

- No report

The Netherlands

- No report

Norway

- No report

Poland

- Workshops in Poznań and Gdańsk
- We are waiting for the result of Submitted CoreTrustSeal (it is now in “Under Review” state)
- We prepared docker’s for NextCloud, D-Space and RabbitMQ webservices chain with full integration between them (export/import, NLP processing), all code is deployed on <https://github.com/CLARIN-PL> and <https://hub.docker.com/u/clarinpl/>
Docker Compose file: <https://nextcloud.clarin-pl.eu/index.php/s/LqjUVWC3FK9EgwE>

Portugal

- No report

Slovenia

- Working on localising DSpace messages
- Fixed hardware failure due to power outage
- Addressing bugs in repository (Slack, GitHub issues)

Sweden

- Continued FCS outreach
- Moving to OpenShift, also moving repositories to Github

United Kingdom

- OTA DSpace launch has been delayed due to staffing problems, but production site is now being built and planned for launch in late June 2018.

7 Task force reports

CMDI

- No report
-

Metadata curation

- No report

Federated Content Search

- LJO continued work with centres and their endpoints.
- Preparing for svn -> github move of aggregator
- Preparing Aggregator and reference endpoints to be buildable in OpenShift.

Persistent Identifiers

- No report

AAI

- Martin and Lene attended AAI/Metadata Workshop in Kiel, promoted CLARIN IdP as interim solution for AAI issues.
- U Cologne is releasing attributes to all CLARIN centres in eduGAIN that support the Data Protection Code of Conduct. Likely the first in Germany

8 Service Inventory PARTHENOS Project

Daan (NL): within PARTHENOS we have devised an extra task about inventoring tools in the humanities, Dorte Haltrup Hansen (Copenhagen) will start with this task. When the list of tools is a certain size we would like to give it to the centres, so they can provide feedback to this list. The request to the SCCTC is to please support this.

9 Any other business

Martin (FI): One of the partners of FIN-CLARIN has asked how we use our back-up, how we store, data, if it is possible to delete their data and how we encrypt data. They are creating a separate agreement (Third Party Agreement). What is the status of GDPR in CLARIN?

Dieter (CE): The GDPR discussions (and the questions around it) are complex. There is no simple answer to the question on how CLARIN is compliant. Overall it is good to remember that not so much actually changed if you compare it to the earlier privacy legislation and that it is important to strike the right balance between doing our job (providing language data and tools for research) and being reasonable when dealing with privacy-sensitive information.

Martin (FI): So CLARIN in all areas and other services is not storing any personal data?

Dieter (CE): CLARIN is not storing personal data/information in Piwik/Matomo, since it anonymizes the IP-addresses. For other service – especially mailchimp for the newsletter, which is hosted in the US – we are obviously dealing with email addresses, so we have informed the subscribers.

Nicolas (FR): What is the current state of the law in French? Our portal is located in France.

Dieter (CE): Services that CLARIN ERIC is responsible for as a legal entity, fall under Dutch law. For the individual countries it depends on your own national law.

Leif-Jöran (SE): It is important to make an inventory, which you can already do. Make sure to make a list of what you have, where you store it and what your using it for.

Next meeting: The next meeting will take place in the week of 25-28 June 2018. It will be a short meeting, only the most urgent points will be discussed. A doodle will be sent out shortly.