Title	Minutes SCCTC 5 September 2018
Version	2
Author(s)	LS/ DVU
Date	2018-09-05
Status	Approved
Distribution	Centre Committee
ID	CE-2018-1275

Participants

Jan Hajič (CZ), Dirk Goldhahn (DE), Mitchell Seaton (DK) Krista Liin (EE), Martin Matthiesen (FI), Riccardo Del Gratta (IT, Marcin Pol (PL), Luís Gomes (PT), Leif-Jöran Olsson (SE) , Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes), Darja Fišer (CLARIN ERIC, Director of User Involvement as invited participant)

Excused: Nicolas Larrousse (FR), Lene Offersgaard (DK), Matej Ďurčo (AT)

0 Action points

#	Action	By whom	By when
1	Follow up with EUDAT on the price and policy of B2Safe services.	Dieter (CE)	Asap
2	Contact all centres that have been participating in B2Safe uptake plan and see how they see the total picture.	Dieter and Willem (CE)	Asap
3	User delegation: Have a look at the plugin of the pilot version of the test instance of unity-idm	Marcin (PL)	Asap

1 Agenda

Sub point a. “Purpose of reporting google doc” was added to Agenda point 6 “Status update per country/member” at the request of MM Agenda and the agenda was approved as follows:

1. Agenda
2. Presentation by Darja Fišer on metadata issues related to the Resource Families (CE-2018-1236)
3. Approval of minutes last meeting & action point status (CE-2018-1248)
4. Update from the assessment committee (if someone is available, otherwise short written report)
5. Proposal for new B-centre document (skeleton to be distributed) ->no substantial document available yet
6. Status update per country/member (please provide a short bullet-wise summary in the google doc -> please notice that we have merged the SCCTC and the NCF report section)
   a. Purpose of the google doc
   b. Status update per country/member
7. Any other business:

    1. Background information: VLO benchmarking and scalability (CE-2018-1254)
    2. User delegation: short status update - to be distributed (Dirk, Marcin, Willem)

2 Presentation by Darja Fišer on metadata issues related to the Resource Families

See: (CE-2018-1236)

Darja Fišer: Last year we started doing survey and overviews of particular types of resources within the CLARIN Infrastructure. It was done two-fold, first to get a better idea of how much we already have of a particular type of resource and secondly, we wanted to evaluate if they can be easily found trough the VLO and if metadata records quality is good enough for the end-users (researchers from digital humanities, social sciences and human language technologies)
The overviews are based on extensive surveys of resources available in the VLO and the CLARIN repositories. We also provide lists of corpora for each type of resource that cannot be found in the VLO or have problematic metadata.

Darja asks to try and see if the list of issues mentioned in this document can be improved by people in this committee. As can be seen in the report the problems are trivial and can easily be solved and will make the infrastructure cleaner and clearer as a whole. This is no criticism but simply a collection of issues that can be improved for the future.

Dieter (CLARIN ERIC): Potential ways to follow up on these issues is through a specific GitHub repository: https://github.com/clarin-eric/resource-families-issues/issues (it is publicly accessible).Here we can follow up and make someone responsible so specific metadata issues that are addressed.

Darja suggest the next steps to be as follows:

- Designated person per centre, it will be good to gather people in charge at Centre level who can take care of those issues.
- Examine issues for particular centre in three stages:
    1. Which issues can easily be solved
    2. Which issues can be solved midterm
    3. Which issues can be solved long term or cannot be done at all
- Have a milestone in about 6 to 12 months and then check where we stand and have a further discussion.

Dieter (CLARIN ERIC) suggests to:

- Have a look at the complete amount of issues and try to first get this list fully completed between now and two weeks (by Jakob Lenardič)
- Have a round of adoption of tickets, invite centres assign themselves and then take two weeks for adoption centrally.
- Go through issues that have not been adopted and divide these.
- Try to first get this list fully completed, between now and two weeks (by Jakob)
- Put out a call for adoption of metadata issues, where centres can assign themselves (two weeks)
- After two weeks centrally contact persons who can adopt the orphaned issues.

Darja thanks everyone for being so open to this, a lot has been invested and the infrastructure gains a lot of quality with taking this step.

Dieter (CLARIN ERIC) thanks Darja for joining us

3 Approval minutes and action points

The minutes of the SCCTC meeting of 28 June 2018 are approved.

4 Update from the assessment committee

Dieter (CLARIN ERIC): Lene is excused for this meeting so currently no details are available. We can follow up by email if anything needs to be approved or taken into preparation for the assessment meeting in Pisa.

5 Proposal for new B-centre document

Dieter (CLARIN ERIC): Document is not yet ready. We have not yet come up with something that is mature enough to distributed. There are some ideas, but nothing to discuss at this point. We should have a starting structure that we can put forward as a document for our Pisa meeting.

Luís Gomes (PT): We are ready to start filling the document for becoming a new b centre b, should we start filling the old document or is the new one available very soon and should we use that one?

Dieter (CLARIN ERIC): You can safely start with the old document. In terms of requirements these will remain the same and will not change in the new document. If, in the meanwhile, the new document comes available you can easily copy/paste the data from the old to the new document.

6 Status update per country/member

a. Clarify the purpose of the reporting google doc

Dieter (CLARIN ERIC) explains who the audience of the document is and what it is intended for: The document was created to get a common view of what happens in a country in general and on a technical level. The audience is the CLARIN Board and of course colleagues in other countries. People appreciate these overviews to see what is happening in the community. Because it was not always clear what should be National Consortia news and Centre Committee news, it was decided to group everything in one Google document.

Martin (FI): It is sometimes empty, we should get info from everyone and make sure everyone fills it out.

Dieter (CLARIN ERIC): We will ask it more explicitly to all members. CLARIN Office can pursue this and can send a reminder. We just need a short update in e.g. three bullets. Even if nothing happened, this can be reported as well. LS for SCCTC and QB for NCF will follow up on this.

b. Status reports per country/member

Austria

- CLARIAH-AT Meeting: Preparation of next three years

Bulgaria

- No report

Croatia

- No report

Czech Republic

- Updated neural machine translation models English<->Czech. Service still in beta. https://lindat.mff.cuni.cz/services/transformer/
- Progress with ministry negotiations towards LINDAT/CLARIAH-cz
- Corpus search tool Kontext with LINDAT corpora now ready for prime time and available from the main menu at https://lindat.mff.cuni.cz

Denmark

- Data and tools moved to DSpace repository.
- New resources have been uploaded in the repository (examples Grundtvig's Works Corpus, STO lexicon - morphology, syntax, semantics in LMF format)

Dutch Language Union

- Collection and parsing of Wablieft-corpus (easy-to-read newspapers in Dutch)

Estonia

- Received 3rd feedback from DSA reviewers. Still have to do some minor changes, but basically should have filled all the requirements by now.
- Contracted an IT-team to work on developing content search. Based on CLARIN FCS, but to include also lexicons and possibly (some) annotated speech corpora in Estonia. The project will run until end of 2018.
- Funding decision for Estonian National Programme in Language Technology (including CELR funding) came in at beginning of July.

Finland

- In Progress
    - Korp-cooperation / FCS still not implemented
    - Moving towards DevOps (new developer started)
    - Investigating possible co-operation with UzK on repository solutions (test system expected 10/2018)
    - HTML Version of Mylly.
    - Co-operation with National Library to distribute recent newspaper corpus
    - Improving tools documentation
    - Improving internal process descriptions
- Done
    - FREME-evaluation
    - Update of Language Bank Rights (lbr.csc.fi)

France

- Report here

Germany

- Merging of CLARIN-D and DARIAH-DE helpdesks: first discussions have started
- Next CLARIN-D developer meeting: 13.09.2018 in Saarbrücken
- DSA/CTS: Still working on common feedback for some general remarks of the board (making contractual information and other documents available; MoU on resource handover between centres)

Greece

- Tour de CLARIN Greece finalised

- Setting up a new repository for the new Greek network member (National Center for Social Research)
- Preparing training workshops for the new members' teams on the use of the infrastructure.

Hungary

- No report

Italy

- ILC4CLARIN at the latest release 2018.02 (set up in August)
- Close to the Clarin Annual Conference (wait for u all)

Latvia

- No report

Lithuania

- No report

The Netherlands

- CLARIAH Summerschool Media Studies, 2-6 July, Amsterdam. Web site
- Evening Lecture by Jan Odijk at ESSLLI, `Boosting Linguistics with CLARIN', 14 aug, slides (also filmed, will soon appear, probably via CLARIN website)
- Made detailed plans and budget for successor project CLARIAH-PLUS (to start 1 Jan 2019) (and this is still on-going)
- First version of faceted search for software fully based on CMDI records has become available (already in July). New updates will follow soon. http://portal.clarin.nl/clariah-tools-fs
- CLARIAH Workshop on Provenance, The Hague, 3 Sep 2018. Programme

Norway

- Preparing a proposal for a grant from the Research Countil towards an upgrade of the CLARINO infrastructure (deadline in October; massive competition from other proposals)

Poland

- We've got Core Trust Seal Application Feedback. We will check it after 15.09.2018.
- Changes in NextCloud and new plugins as Draw.io
- We added the WSD module to the Inforex - a Collaborative System for Text Corpora Annotation and Analysis.
- Reassessment procedure - We have check all suggestions from doc's and fixed / corrected / applied them.

Portugal

- Adaptation and integration of 11 tools into the workbench of clarinportulan. Work in progress for two more tools.
- Deployment of workbench is ongoing (already allocated hardware and started installation).
- Started work on welcome page for clarinportulan.

Slovenia

- First CLARIN.SI call for small projects (up to 8k EUR), good turnout (8/10 projects accepted for financing). Projects should be completed in 2018.

- Ministry sent call for proposals for research equipment grants for Slovenian ESFRI infrastructures 2018-2021, we are busy preparing our proposal (for cca 400k EUR).
- New resources in repository: 23-language JRC EU DGT Translation Memory Parsebank DGT-UD 1.0; Spoken corpus Gos VideoLectures 3.0 (37 talks, 16 speech); Serbian Training corpus SETimes.SR 1.0 (86 726 tokens manually annotated on the levels of tokenisation, sentence segmentation, morphosyntactic tagging, lemmatisation, syntactic dependencies, and named entities)

Sweden

- Some movement to decide on "nationally usable" Metadata profile. Planning workshop for promotion
- FCS work integrate into infrastructure, OCR and other classics infra into OpenShift (hopefully done by December)
- New Swe-Clarin roles in relation to new National Language Bank (national infrastructure), Swe-Clarin Consortium meeting next week
- (Review of ISO-standard)

United Kingdom

- UK period as an Observer has expired. We are working with the Consortium and the Arts and Humanities Research Council (now part of the new 'UK Research and Innovation') on an application for renewal.
- Lancaster University Corpus MOOC starts in September.
- Several UK participants will take part in the Oral History workshop in Munich in September.

3rd parties

USA

- Development of a new browser search engine for all of TalkBank, based on XML, JSON, and MongoDB. We will report on this work at the upcoming meeting in Pisa.
- Movement of all of TalkBank resources (14 terabytes) to the Carnegie Mellon Campus Cloud facility.
- Creation of new methods for web playback of media that have not yet been transcribed.
- Progress on the creation of methods for ASR diarization of daylong recording of spoken interaction using the Speech Kitchen (speechkitchen.org) toolbox and HomeBank training data.
- Movement of DOI creation from EZID to DataCite.
- Coordination of work on the TRJS transcriber with OrtoLang (Christophe Parisse).

7 Any other business

1. Background information: VLO benchmarking and scalability (CE-2018-1254)
   This is an informative, short document that Twan Goosen has been creating when testing the scalability of the VLO. It can be of interest to the committee, if you have questions on this topic please contact Twan directly (twan@clarin.eu).

2. User delegation: short status update - to be distributed (Dirk, Marcin, Willem)

   Dieter (CLARIN ERIC): In short, ideally, we would like to have a solution application writing into next cloud. Willem has set up a pilot version of the unity, see summary document . Still work needs to be done to glue it all together

Dieter (CLARIN ERIC): Is there any update regarding this pilot?

Marcin (PL): No news due to vacation time. Marcin will have a look at the plugin after his vacation. We will Keep this as an action point on the agenda also for Pisa. **See AP 3**

Dirk (DE): Currently nothing to report. The last thing that happened was an update from Willem at the end of August, where he specified to look into setting up the unity instance.

Dieter (CLARIN ERIC) requests Dirk: to bring this up in Saarbrücken and ask them to check whether WebLicht can still connect and make sure they have this check on the radar. Dirk agrees to do so at the developers meeting.

3. Communication about change in VLO harvested viewer

Mitchel (DK): the VLO harvested viewer has changed, is this communicated to the centres. How is this being communicated?

Dieter (CLARIN ERIC): Good point, the traditional way is in the issue of the centre news that is send by email, but this month it has not been distributed yet. Because of the summer holidays in August the last issue was sent in July. We also make sure to announce it when the VLO viewer works better (it is still somewhat slow). It will be announced soon but not sure if it is in the next issue of the centre news.

Next meeting: 8 October 2018 (11:00-13:00), f2f meeting at CLARIN Annual conference 2018 in Pisa.