

**Title** Minutes SCCTC 30 June 2020  
**Version** 1  
**Author(s)** LS, MMA, DVU  
**Date** 2020-06-30  
**Status** Approved  
**Distribution** Centre Committee  
**ID** CE-2020-1696



## Participants

Martin Matthiesen (Chair, FI), Marinos Ioannides (CY), Jan Hajič (CZ), Thomas Eckart (DE), Simon Gray (DK), Lene Offersgaard (DK), Krista Liin (EE), Christophe Parisse (FR), Vanja Štefanec (HR), Riccardo Del Gratta (IT), Tomasz Naskręt (PL), Luis Gomes (PT), Leif-Jöran Olsson (SE), Dieter Van Uytvanck (CLARIN ERIC), Linda Stokman (CLARIN ERIC, minutes)

Excused: Daan Broeder (NL), Martin Hennelly (SA)

## 0 Action points

#	Action	By whom	By when
1	Prepare a broader discussion on the FAIR Data Maturity Model within the competence of the PID Task force this week and send out an invitation for this. (see questions Martin M. sent to SCCTC by email on 27 April 2020)	Daan & Martin M.	May 2020
2	Create a mailing list for the Fedora Commons interest group	Dieter	10 July 2020

## 1 Agenda

1. Agenda:
  - a. The agenda is approved with the following addition to agenda item 7 (Any other business): 'Request to invite Lene as a liaison to the centre assessment committee'
2. Approval minutes last meeting & action point status
  - a. Minutes of 26 May meeting ([CE-2020-1673](#))
3. Centre assessments
4. Evaluation [Fedora Commons online meeting](#) and follow-up: creation of a [short use case document](#) for CLARIN centres
5. Discussion & brainstorm, topics:
  - a) Draft strategy for technical infrastructure ([powerpoint, slide nr 6](#)) [Dieter]
  - b) *Implementation and maintenance of the quality control for all CLARIN tools* (request from NCF) ([NCF work plan, section 2.2](#)) [Dieter]
6. Status update per country/member (please provide a short bullet-wise summary in [the Google Doc](#))
7. Any other business
  - a) A new afternoon option for meetings (to make CMU attendance possible)
  - b) Request to invite Lene as a liaison to the centre assessment committee

## 2 Approval minutes last meeting & action points

- a) The minutes of the 26 May meeting ([CE-2020-1673](#)) are approved.

## 3 Centre assessments

Lene reports that the 15<sup>th</sup> round is still ongoing on. Currently the South African Centre is being assessed, the report is expected to be ready for the next centre committee meeting.

## 4 Evaluation Fedora Commons online meeting and follow-up

Link to Fedora Commons online meeting:

<https://docs.google.com/document/d/1U8w-0DF4yoLTHscfW3d9U1-GGICz-ImjUuN4Fy9M1Dk/edit>

Follow-up: creation of a short use case document for CLARIN centres:

[https://docs.google.com/document/d/1-AS\\_ZpaRj-aNMpkHBU-UWEjuzOVhrVJlqjkP9yCNCgs/edit](https://docs.google.com/document/d/1-AS_ZpaRj-aNMpkHBU-UWEjuzOVhrVJlqjkP9yCNCgs/edit)

Thomas informs the committee that during the Fedora meeting with David Wilcox, it was decided to create the short use case document with CLARIN's requirements to a repository solution as a basis for further discussion with Duraspace/Lyrisis.

Anyone who is using Fedora Commons or is otherwise interested in the topic can add suggestions to the document.

A Fedora Commons interest group will also be set-up. This group will function as a contact point, can push the improvement or can be used for comments or requests in the context of the development. (a similar thing could be setup for DSpace)

Dieter will create a mailing list for the Fedora Commons interest group. **AP2**

After the mailing list is created, people from different repository systems can be invited, also centres not present at this meeting, such as the Austrian centres, ACDH-CH and ZIM centre and the Dutch centres, Max Planck and Meertens/ HuC, might have an interest in joining.

## 5 Discussion & brainstorm, topics

- a) Draft strategy for technical infrastructure ([powerpoint, slide nr 6](#)) [Dieter]

The BoD is working on a new strategy document and is requesting input from several committees on initial ideas.

### **Technical infrastructure bullet**

After some discussion in the NCF it was suggested to name some new technologies, such as deep learning. If you are working on new technologies on the side of your centre it is good to include it here. Related to this, it would be good to think about how we give this a place at discussion fora like the centre committee. For instance, initiate an interest group for upcoming technologies with a free format, where people can exchange ideas, experiences, get opinions or give short presentations. Would be good to have this, many people have been asking for this.

Concerning the final point on the slide about quality assurance and curation, this is in a way a continuation of what we already started with monitoring centre assessments. It is important to find a balance in what we can check automatically to make sure things work or to have manual intervention where needed.

The centre committee is asked to have a look at the slide and see if anything is missing or can be added as a part of this document.

### **Comments:**

Marinos asks if the current policy activities of the EC are directly linked to what has been explained especially with regards to Europeana.

- Dieter: We have some active collaboration with Europeana see: [clarin.eu/europeana](http://clarin.eu/europeana) where you can find for some pointers. This is a good thing to mention and Dieter will take a note, to make sure we also mention the Digital Culture Heritage as part of the strategy for the technical infrastructure.
- Marinos asks the committee to disseminate (and contribute to) the running EC Consultation: <https://ec.europa.eu/digital-single-market/en/news/public-consultation-opportunities-offered-digital-technologies-culture-heritage-sector>  
The EC running consultation is a public consultation in view of the evaluation and possible revision of its main policy instrument to support the digital transformation of the cultural heritage sector. Marinos mentioned it is extremely important to distribute within the centres networks. As it is very important for the sector of the humanities. Dimension of Europe is very important so far as policy and future investments are concerned. Marinos wanted to take the opportunity to highlight this running consultation so that CLARIN can take an active position in this.
- Martin Matthiesen is wondering what the second part of the first point ‘innovative / experimental features are integrated pro-actively’ means in practice, because CLARIN.eu cannot really demand this from the centres or if they do it might not happen. Should innovative/ experimental features always be integrated?

- Dieter: It is more intended as a kind of indication that we should not get an old and rusty vehicle. It is good to keep a clear and fresh mind of things and evaluate different features. It is about the mindset, try to see how the universe around us is changing and try to be more up to date to that instead of sticking to the concept that we had at the launch of the infrastructure.

**Conclusion:** The points in the strategy are well received in the committee and there are no immediate objections. The points are not specific yet, but they make sense to the committee.

- b) *Implementation and maintenance of the quality control for all CLARIN tools* (request from NCF) ([NCF work plan, section 2.2](#)) [Dieter]

This part of the NCF work plan concerning the quality control for all CLARIN tools that is focused on their proper description, robustness and reliability. This is mainly related to the fact that when we tested web applications that are connected to the Switchboard, we found out there was a high spread in terms of how robust applications are. There were applications that could not process files bigger than 100KB. In case an application cannot deal with a format or size it would be good to get a clear error message. These kinds of issues brought up the point that it might be good to do a quality check to see if applications work reasonably, if they are robust and to have more information about their limitations.

Dieter asks the committee how this point can best be addressed and if there are any ideas of bringing this robustness further, such as doing some basic checks, have recommendations for robustness etc.

**Points made by the committee that the NCF can consider:**

- We should stay out of quality measuring because it really depends on what you use your tools for. Some tools can work very well in one domain and not work well in others. If we want to use quality measurements, we would need to exactly define what that means.
- Many of the tools CLARIN provides come out of Scientific Research, we do not want to get into a scientific discussion, but it is good to have some indications. We could have a higher “quality/robustness” standard for commonly used tools and some lower “quality/robustness” standards for bleeding edge research.
- The term quality can create confusion. When using a term like robustness it is clear that it is about the software domain. When using the term “quality” it could be interpreted as checking the scientific content of the resource or tool.
- Proper documentation with description and robustness is important, but we have to be careful with quality measures and benchmarking, because it might hinder the implementation process if people are trying to submit to

the benchmarking. One suggestion would be to add 'where appropriate' to section 2.2 of the work plan.

Dieter will report these points to the NCF.

## 6 Status update per country/member

### Austria

- No report

### Bulgaria

- No report

### Croatia

- No report

### Czech Republic

- Lexical and corpus search projects continuing, new versions of tools and services to be included soon (NER 2.0, UDPipe)
- Participation at AI policy meetings (national, V4, EE)
- Establishing the International Advisory Board of LINDAT/CLARIAH-CZ (letters sent)
- Submitted project to Call 49 ("LT4EU"), with CUNI/LINDAT participation
- Opened new building in the Troja campus of the University, with several members of the team located there, contract for new servers signed for there

### Cyprus

- No report

### Denmark

- Further security consolidation of clarin.dk servers
- Broader set of corpora available in FCS korp endpoint + bugfixes.

### Estonia

- Performance problems with KORP
- Continued work on Zammadi helpdesk integration
- Moving servers to the new building

### Finland

- FIN-CLARIAH Road Map application sent to Finnish Academy
- Korp update still ongoing.
- CentOS6 EOL 30.11. requires moving/updating of 4 servers (ongoing)
- Missing funding forces CSC to minimize service levels for the Language Bank

### France

- One French centre (Cocoon) is now covered by the federated content search.
- There is work in progress for another centre (Ortolang) to achieve the same.
- The Ortolang centre is still at work for applying for CoreTrustSeal, but we have made progress in this direction.
- A French K Centre (CORLI) is nearly ready and in the process of obtaining agreement from CLARIN.

## Germany

- Rejected follow-up proposal for German CLARIN project
- Ongoing work on AAI integration in FCS
- Follow-up action about future use of Fedora Commons in CLARIN (video conference, requirements document)
- Final work on migration of WebLicht infrastructure and web services to TCF5 started

## Greece

- No report

## Hungary

- No report

## Iceland

- No report

## Italy

- ILC4CLARIN
  - Time to finalize ALIM into CLARIN: expected before August.

## Latvia

- No report

## Lithuania

- Nothing to report

## The Netherlands

- No report

## Norway

- The Research Council of Norway has appointed Koenraad De Smedt as a member of a national committee for assessing the practice and impact of data licensing.

## Poland

- Two training online seminars (mini-workshops): 1. use of Inforex platform in linguistic research (23 June for UŚ). 2. use of the Inforex platform in media research (June 5th for researchers from Jagiellonian University).

## Portugal

- Matomo integration, started collecting access statistics
- Adding file processing ability to several tools of our workbench
- JSON-RPC API for several tools of our workbench
- Continued work on integrating new tools

## Slovenia

- Added new entries to the repository
- Start of work on the new ParlaMint project
- Preparation of information about research data repositories in Slovenia in the scope of the RDA Node Slovenia project

## South Africa

- No report

### Sweden

- New Swe-Clarín website ready for deployment, helpdesk rollout next
- More resources in preparation for releases before vacation period
- Finalizing ongoing work with partners on AAI integration in FCS

### United Kingdom

- A funding proposal is in preparation to be submitted to the UKRI to fund arts and humanities infrastructure for ten years, including funding for full membership of CLARIN ERIC and funding for the OTA as a trusted data repository.
- Lancaster University summer schools are under way as virtual events

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

- Because of the Summer Holidays it is agreed that the next meeting will take place in September.