# Meeting of the CLARIN Standards Committee (CSC)

## *Vienna (AT), April 10th, 2015, 9:00 – 12:30*

Participants: Andreas Witt (chair), Jan Odijk, Pavel Rychly, Dieter van Uytvanck (Centre Committee chair), Dan Broeder, Inke Schuurman, Axel Herold, Karlheinz Moerth (co-chair), Neeme Kahusk

- A. Witt provides a brief overview of CSC related events since the last meeting
  - CMDI-1.2 ISO standardization did not make any progress
  - A. Witt and K. Moerth participated in a Dariah ERIC meeting in Rome
  - K. Moerth is the coordinator of the DARIAH Standards working group
  - in Horizon 2020 project applications, standards will be a key issue
- Main discussion: Should CSC »enforce« the use of standards in CLARIN and what kind of enforcement could that be? The following lists the points of general agreements:
  - Standards in CLARIN often emerge on the national level (e.g., TCF, Folia). CSC will have to focus beyond that level.
  - The ultimate goal of CLARIN lies in the interoperability of its resources and tools. Standards are not applied for their own sake but primarily to achieve interoperability or because of the availability of certain tools.
  - The Centre Committee will provide a list of mappings from MIME types to existing tools as an additional input for the list of recommended standards.
  - There should be no restriction to just a single standard (e.g. for a given type of annotation) for different resource types.
  - A set of standards for which CLARIN aims at full support (CLARIN recommended standards), must be identified. Under this premise, »enforcement« will be achieved by a voluntary adoption rather than by strict rules. CMDI and FCS are two examples where de-facto standards emerged from a large scale voluntary adoption.
  - The Centre Committee will act as a link to the individual centres for feedback on the CSC recommendations, because the centres are the first target group for the CSC recommendations.
  - The second target group for the CSC recommendations is data providers and/or projects outside CLARIN that want to use the infrastructure. There could be a FAQ list or an interactive chooser (based on a decision tree) guiding this group from a given type of resource to the recommended standard.
  - Still, CLARIN should be liberal in accepting legacy/non-standard data (even plain text, PDF, or word processor formats) from external providers and provide converters to the recommended standards.
  - Work on a set of CLARIN recommended standards should first focus on written text (corpus) data as this comprises most of the data available to CLARIN at the moment. The multimodal data may also be a good starting point as this data is much more dependent on a set of only a few existing tools. It doesn't show much variability.
- CSC collaboratively drafts a position paper on its role and some possible means of standard enforcement within CLARIN. Upon finalization, the paper will be forwarded to the Centre Committee for feedback and after another revision to the Board of Directors for approval.
- A. Witt provides a brief outlook on CSC related events coming up

- A. Witt will attend ISO meeting in Japan, June 2015 (finalization of transcription of speech and of corpus query lingua franca)
- A. Witt will attend DH2015 in Sydney
- A. Witt will attend the next TEI meeting Lyon, possibly others as well