



*Friday October 26*

## **Towards Integration of the CLARIN Technical Infrastructure, Content, and Services: on Metadata**

Why are there access policies for metadata – shouldn't they be freely available? Sometimes it is good to confirm the obvious: metadata should be freely available. Sometimes also because of ethical constraints. Could metadata be written in a non-personal way so that they can be published?

It is important to raise awareness of means and scope of anonymisation, e.g. "baker of the village" is not anonymous. Metadata must be open and searchable by everyone. Sensitive data should be eliminated and put into a second set of internal metadata.

CLARIN should be more ambitious, shouldn't only publish own data and metadata, but become the first place of access for any researcher looking for data. We should integrate at least links to other humanities data available from other partners. It should be easier for other researchers to integrate metadata of their resources.

The user perspective is lacking at the moment. A platform for users is built, but the users are still absent. Users think about all data that exist, not only about CLARIN data.

Rather than focusing on individual tools and their interoperability, work processes should be more in focus. User studies are important, but we have to do more and have to imagine what users might need in 10 years. That might be very different from now.

Should complete resources be redone in new standards that are already in some de facto standards? Users wouldn't appreciate to relearn resources and tools that they already learnt. And funding agencies wouldn't appreciate that work either.

We need to find different ways, and the situation isn't so bad – we already have quite a lot of conversion scripts. It's not only about tool integration but also about user integration.