

DELAD-CLARIN Workshop 27-28 January 2021: Introduction



Henk van den Heuvel, CLST Radboud University



DELAD initiative



- Initiative to collect and share corpora of speech with disorders (CSD):
- <http://delad.net/>
- Partners are a mix of researchers, infrastructure specialists, legal experts
- DELAD organises annual workshops since 2015 where these groups convene
- Since 2017 under [CLARIN](#) header and support

Topics addressed:

- Examples of CSD
- Guidelines for collecting and sharing CSD
- Ethics and legal aspects
- Levels of anonymisation
- Layered access of data
- Integration of CSD in the CLARIN infrastructure
- Formats
- Relevant metadata

DELAD Background

- Second CLARIN-DELAD Workshop, Utrecht NL, SURF premises, 28-29 Jan. 2019
 - Type II workshop
 - Involve further research partners to DELAD initiative
 - Ethical & Legal aspects of CSD sharing (CLIC members)
 - Specs & priorities of a CSD centered webportal in CLARIN context
 - Possible role of Talkbank
 - How to best spend 3 PM ICT developer
 - Outcomes:
 - Reaffirmation that CLARIN is the Data Trust, to provide the data fence around CSD.
 - DELAD should apply to become a Task Force within CLARIN focusing on practical issues on sharing CSD.
 - As a result, the DELAD website should be updated with a CLARIN flag and contain relevant guidelines for collecting, sharing and storing CSD.
 - Talkbank is seen as a good CLARIN site to host CSD, especially if a European storage cloud and stricter access policy can be realised.
 - Collaboration DELAD, ACE, TLA, Talkbank in data curation
 - Work together on contributions for the CLARIN AC 2019 in Leipzig.
- This workshop will show the results of these actions



DELAD CLARIN Proposal



Goals of this workshop:

- Extend DELAD network with new participants
- Explore with the participants the potential of the new CLARIN K-Centre on Atypical Communication Expertise (henceforth ACE) for hosting CSD of DELAD members
- Exchange deeper insights on Data Protection Impact Assessments (DPIAs),
- Discuss voice conversion as a means to pseudonymize speech

Practical matters



- Recordings will be made of plenary sessions
 - Edited versions will be made available for CLARIN dissemination
- Breakout rooms: at breaks and roleplay to tomorrow
- Participants list: please check, see mail
- Presentations will be collected in Googlefolder (Maria?)

Programme for 27 January

- 10:00 – 10:15 Introduction (Henk van den Heuvel)
- 10:15 – 11:00 Research presentations:
 - Croatian written and spoken corpora of speech with communication disorders (Gordana Hržica)
 - Oral and written documents of mental health patients (Silvia Calamai & Rosalba Nodari)
- 11:00 – 11:15 Break
- 11:15 – 12:00 Research presentations:
 - Using electromagnetic articulography for the purpose of studying speaking styles and speech disorders (Katarzyna Klessa, Anita Lorenc, & Łukasz Mik)
 - Parkinson's disease: A French corpus collected using MonPaGe protocol (Veronique Delvaux)

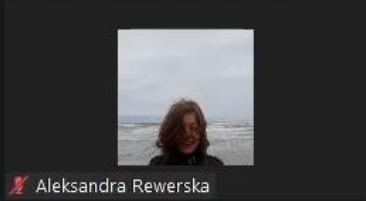
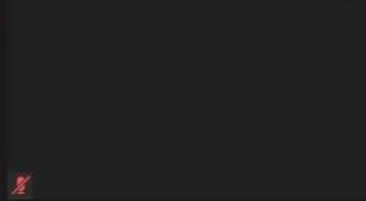
Programme for 27 January

- 12:00 – 13:45 Lunch break
- 13:45 – 14:00 Introduction to roleplay of 28 Jan. (Esther Hoorn)
- 14:00 Group picture
- 14:00-14:50 Help from DELAD and CLARIN Centre for Atypical Communication Expertise (ACE) in sharing CSD (Henk van den Heuvel)
- 14:50-15:00 Break
- 15:00 – 16:00 How to access & deposit existing data at CLARIN centres, profiles of metadata, licenses (Paul Trilsbeek)

Practical matters



- Recordings will be made of plenary sessions
 - Edited versions will be made available for CLARIN dissemination
- Participants list: please check, see mail first day for the link
- A second group photo at 12:00 (just a second)
- Presentations will be collected in central folder: please mail to H.vandenheuvel@let.ru.nl



Programme for 28 January

- 10.00-10.15 Welcome and presentation on the DPIA in research and education (Esther Hoorn)
- 10.15-10.25 Discussion on expectations (Esther Hoorn)
- Break to get into your role
- 10.30-11.20 The roleplay Data Protection Impact Assessment
 - Part 1 (research general) 10.30-10.50
 - Part 2 (building an app with insurance company OR archiving at ACE) 10.55-11.20
- 11:20-11.30 Break
- 11.30-12.00 Evaluation (step 3)
- 12:00 Group photo
- 12:00 – 14:00 Lunch
- 14:00 – 15:00 Keynote: Use voice conversion for pseudonymisation? (Rob van Son)
- 15:00 – 15:15 Break
- 15:15 – 16:00 Wrap up session
 - **What have you learnt in the workshop?**
 - **What do you see as next steps for DELAD?**
- 16:00 Chat over drinks