Participants

Martin Matthiesen (MM) Chair Finland
Matej Đurčo Member Austria
Jesse de Does Member Belgium
Ivan Georgiev Member Bulgaria
Vanja Štefanec Member Croatia
Jan Hajíč (JH) Member Czechia
Simon Gray Member Denmark
Krista Liin Member Estonia
Etienne Petitjean Observer France
Thomas Eckart Member Germany
Dimitris Galanis Member Greece
Samuel Thórisson Observer Iceland
Riccardo Del Gratta (RDG) Member Italy
Roberts Darģis Member Latvia
Andrius Utka Member Lithuania
Menzo Windhouwer Member Netherlands
Hemed Al Ruwehy Member Norway
Tomasz Naskręt Member Poland
Luis Gomes Member Portugal
Cyprian Laskowski Member Slovenia
Leif-Jöran Olsson Member Sweden
Langa Khumalo Observer South Africa
Martin Wynne Observer United Kingdom
Brian MacWhinney Third-Party USA (TalkBank-CMU)

Dieter van Uytvanck CLARIN ERIC Netherlands
Julia Misersky (JM) CLARIN ERIC Netherlands
Lene Offersgaard Assessment Committee
Jennifer Frey (JF) EURAC Italy
João Silva User Involvement Committee Portugal
### Action points

<table>
<thead>
<tr>
<th>#</th>
<th>Action</th>
<th>By whom</th>
<th>By when</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995.1</td>
<td>Follow-up: Recommendations on PIDs outcomes: <a href="https://docs.google.com/document/d/1hLhtKWVe5ffFWvU28mukAGC0yl4gaOv6XNT4MVkioQ/edit#">https://docs.google.com/document/d/1hLhtKWVe5ffFWvU28mukAGC0yl4gaOv6XNT4MVkioQ/edit#</a></td>
<td>Daan &amp; Dieter</td>
<td>Feb 2022</td>
</tr>
<tr>
<td>1995.2</td>
<td>Check documentation status of attribute aggregator and prepare a small amendment to the B-centre requirements with a strong recommendation to configure the attribute aggregator</td>
<td>Martin &amp; Dieter</td>
<td>February 2022</td>
</tr>
<tr>
<td>1995.3</td>
<td>Discuss end of terms of the chairs and potential renewal with SCCTC and CAC chairs</td>
<td>Dieter</td>
<td>February 2022</td>
</tr>
<tr>
<td>1995.4</td>
<td>Create inventory of who is awaiting a response and relay this to the CTS</td>
<td>Martin</td>
<td>To be revisited</td>
</tr>
<tr>
<td>1995.5</td>
<td>Add link to CMDI Taskforce report in the NCF report</td>
<td>Dieter</td>
<td>February 2022</td>
</tr>
</tbody>
</table>

### Agenda

1. Agenda: Request for changes? (2 min)
2. Approval minutes last meeting (16.11.2021) & action point status (5 min)
3. Centre assessments (10 min)
4. Series on Sensitive Data: Sensitive data @CLARIN-IT: Two case studies (ILC4CLARIN; RDG + Eurac; Egon Stemle) (20 min)
5. Status update per country/member (please provide a short bullet-wise summary in the Google Doc): (15 min)
6. AOB (5 min)

The agenda is approved as is.

We welcomed Jennifer Frey (JF) who presented a case study on sensitive data (see below) standing in for Egon Stemle from EURAC.
Approval of minutes last meeting & action points

Approved without comments, all APs (with the exception of 1995.5) have been carried over from 16-11-2021 with new prospective finishing dates.

Centre assessments

- Skipped due to absence of LO
- MM suggested to discuss the CTS procedure in an upcoming meeting due to the waiting time (three months) being fairly long

Series on Sensitive Data

Presentation of the DiDi corpus - ERCC/EURAC (JF)
- EURAC is part of CLARIN-IT since 2017
  - ongoing: B-centre certification
  - hosts non-standard language corpora (dialectal, learner, social media etc.)

- DiDi corpus of South Tyrolean CMC: a linguistic corpus study (2013-2015, pre-GDPR) to document and analyse choice of language use and variety
  - made use of collecting and sharing user-generated sensitive data from Facebook (main platform for digital communication in the area)
  - data collected was (in part) personal, private, sensitive, and harmful/compromising → legal and ethical considerations to be made
  - Data collection: development of a Facebook app, integration of FB API to retrieve socio-demographic metadata and access tokens (voluntary access by FB users); backend script to access the data with access tokens from users
  - Legal considerations: Consent, terms of licence and privacy policy, no pre-filled forms, explicit and transparent about how data was used + scope of future use, possibility for users to withdraw, full manual anonymisation (as part of the normalisation, see below)

Questions

MM: Manual anonymisation, was it a lot of work? JF: Normalising of language data needed to be done anyway due to the dialect, and as part of this anonymisation was carried out

2) Presentation of two case studies: Archivio Vivo and ReadLet (RDG)

A) Archivio Vivo Case Study (RDG not directly involved, question can be directed to him via email and he will pass those on)
- a regional project (CLARIN main technical partner)
- motivation: audiovisual data (and connected digital material) of the twentieth century is at the risk of being lost
  - project produces a digitalised archive of Caterina Bueno (named after Tuscan folklore singer)
  - can be used as a model for other audiovisual archives
- pseudonymisation, informed consent, data protection (GARR as a main partner of CLARIN-IT for data protection in the cloud)
- Users can get access via the Vue App: From user registration to login to accessing the resource, there are back-end checks to ensure authentication (uniqueness of identity of user)

B) ReadLet Case Study
- Study on developmental difficulties in language development and language disorders; children (8-11 yo) were recorded during reading text from a tablet, movement of their finger on the text also recorded
  → collects personal and technical information
- uses a web app (children recorded via a tablet, authentication with password)
- Data recording and retrieval uses the following flow: encrypted data/HTTPS protocol > server (API) > pseudonymisation > server (AES encryption)

Questions
MM: How are users informed of how their (sensitive) data is being handled?
Archivio Vivo uses a local login and plans to use CLARIN federated login
ReadLet: Testing sessions are managed by teachers in schools

Status update per country/member
Short bullet-wise summaries can be found in:
https://docs.google.com/document/d/1CClFth0DHZlXCANu4TxZtX79_vJql5JeKcZt9NsKm/ edit#

AOB

JH suggested to share previous and current NCF reports on our agenda so we have a reminder of the developments from the last meeting to the current one

Next meeting

JM has created a Doodle for the last week of January