

# Aspects & requirements of a webportal

Henk van den Heuvel (CLST, Radboud University Nijmegen)

Data

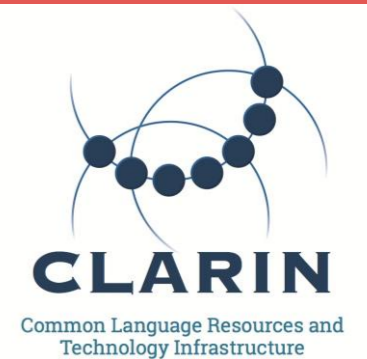
IPR/Ethics

Anonymisation

Metadata

Access

DELAD portal



What limits on type of data and disorder?

- “speech data judged as atypical by the researcher offering the data to the database”
- typical data only if it has been recorded as a control for an atypical dataset
- Speech and language (audio / video /text / film)

Data formats:

- digital audio;
- digital video;
- transcriptions in the absence of audio/video recordings (may have value for historiographical research)
- instrumental data : EPG; EMA; acoustic; nasometric, Ultrasound, fMRI, etc.
- case data and assessment results (how much? What should be obligatory and what optional?)
- results of analyses

Location:

- Local/on site
- CLARIN federated using the national CLARIN data centres
- CLARIN central / one site

Versioning

FAIR delivery

## European and national regulations

- Differences per country
- Effect of GDPR
- Issues: whether data of deceased research participants could be used/shared; how long the data could be kept after the lifetime of the research project; procedures for data anonymization etc.).
- What if a country/institute does not have an ethics board
- Declaration on webportal by contributors that they adhered to these standards?

Distinction between “for education only” and “for research only” (or both). A relevant distinction?

## Contents of consent form (of research projects)

- different levels of consent (e.g. the participants give consent to the use of the data at conferences/meetings, for education purposes only, in publications, to be shared with other researchers through a database etc.).
- Correct legal formulation to share data in other research or inclusion in archive

Opt-out options for participants included in a dataset

Further copyrights and IP rights to consider

Which data should be anonymised:

- Raw data (audio, video)
- Transcriptions

What should be anonymised:

- Names
- ...

How should this be anonymised:

- Coding schemes
- Automatic/ manual

Data

IPR/Ethics

Anonymisation

Metadata

Access

DELAD portal

Which metadata is minimally required

Which metadata is nice to have

Metadata formats

Language of metadata (labels and values): English only?

### Levels of Accessibility

- “secret locker” : for data without any current consent;
- Metadata accessible, but raw data inaccessible
- Parts of metadata/raw data accessible
- All raw data accessible (not realistic for all types of data)
- Register and request combinations

### Location:of the data

- Local/on site
- CLARIN federated using the national CLARIN data centres
- CLARIN central / one site

### Versioning of datasets

### Persistent Identifiers

Data

IPR/Ethics

Anonymisation

Metadata

Access

DELAD portal

Integration in VLO / in Talkbank?

Target users: only researchers?

Login required? Federated login?

Levels of search interface: Simple (use of keywords); Advanced (use of metadata categories); Expert (use of 'regular expressions' in metadata queries)?

Connection to other databases (e.g. CHILDES; Talkbank, Voicebank, Phonbank etc.)?

Links to publications arising from use of the database (should this be obligatory in order to gain access to the data?; how to police this?) (NB: links to publications/citations form good evidence in making the case for further funding for the database)

Who acts as gatekeeper to the data? Do we allow unrestricted uploads or is there an approval protocol? Which criteria should be formulated for that?

Who puts the data onto the database?

Maintenance: who and on which funding

Links to other initiatives such as Talkbank