DARIAH-EU

CLARIN
Common Language Resources and Technology Infrastructure

FIN-CLARIAH

Authors:
Mietta Lennes
FIN-CLARIN, University of Helsinki
Sirpa Kovanen
Research Services, University of Helsinki
Martin Matthiesen
CSC – IT Center for Science
Krista Lagus
Maria Valaste
Faculty of Social Sciences, University of Helsinki
Veronika Laippala
University of Turku

# WP 2.1 Social Data Science

## GOALS

- Help collect shared language resources, e.g., surveys with open-ended questions, interviews, media data
- Develop a common framework for licensing and storing unstructured text and audio with personal data
- Protective measures for storing and sharing data in the special categories GDPR
- Use text and audio processing tools developed in WP1.1 and WP1.2 for enriching and processing the data in open-ended questionnaires or interviews

## DELIVERABLES

- Q3 (Sep 2022): D2.1.1 Licensing agreements for personal data
- Q6 (Jun 2023): D2.1.2 Licensing agreements for special categories

## SUPPORT FOR COLLECTING AND MANAGING RESEARCH DATA

- Guidelines: www.kielipankki.fi
- Advice: fin-clarin@helsinki.fi
- Brochure about the Language Bank of Finland for research participants: www.kielipankki.fi/support/info-for-research-participants/
- The researcher can submit the basic metadata about a new resource via an e-form: urn:nbn:fi:lb-2021121421
- The Language Bank publishes the preliminary metadata and assigns a persistent identifier (PID) to the record, in case the resource is to be deposited with the Language Bank.
- The resource becomes discoverable and citable.
- The metadata record may be modified and completed at a later stage.

### CLARIN Licence Categories

- PUB — Publicly available
- ACA — Academic use
- RES — Individual access

**The CLARIN licensing framework allows researchers to specify the licence conditions for shared research data in a uniform manner. In each licence category, additional conditions can be selected and further details may be added to the licence as free text.**

## DEPOSITION AGREEMENT

- Deposition agreement template: www.kielipankki.fi/support/dela
- Meet with the depositing researcher, to discuss technical issues, safeguards and requirements regarding personal data and/or copyrighted content.
- Help the depositor select a licence category and additional conditions. www.kielipankki.fi/support/clarin-eula/
- Help in writing the personal data processing terms and conditions, if required (licence condition +*PRIV*).
- Select the role of the Language Bank (University of Helsinki):
  1. The data controller allows the Language Bank to independently grant access for secondary purposes of use.
  2. The data controller chooses to remain in charge of granting access for secondary use. The Language Bank acts as a data processor.
- Prepare a draft of the agreement, ask for comments from the rightholders and the data controller, and finalize.
- Have the deposition agreement signed.

## PUBLISH LICENCE

- Translate the data processing terms and conditions, if required (fin+eng).
- Publish the licence page with a PID.
- Link to the licence from the public metadata of the resource.

## After the resource has been received and made available by the Language Bank:

## MANAGE ACCESS

- Federated login (HAKA, eduGAIN, etc.) is required for resources under ACA or RES licences.
- Language Bank Rights (LBR, lbr.csc.fi) is used for managing the application process and access permissions for restricted resources (RES).
- When applying for access to a RES resource in LBR or before downloading an ACA resource, the applicants must agree to the resource-specific licence, including the data protection terms and conditions.

## COLLECT PRIVACY NOTICES

- The end-users of resources containing personal data are required to submit the general title of their project and the link to the privacy notice concerning their purpose of processing: urn:nbn:fi:lb-2022052522
- The submitted details are published on the web portal of the Language Bank.

## FUTURE WORK

- Test and develop the deposition procedure for social science datasets in collaboration with the Faculty of Social Sciences at the University of Helsinki and with the University of Turku.
- Test and develop technical solutions for sharing sensitive data, e.g., Secure Desktop by CSC: research.csc.fi/-/sd-desktop
- Share information about tools and methods that could be helpful for protecting data during research, e.g., semi-automatic pseudonymization or data encryption.

UNIVERSITY OF HELSINKI

DARIAH-FI

KIELIPANKKI The Language Bank of Finland

FIN-CLARIN