



## WP 2.1

# Personal and Copyrighted Research Data

Many researchers need datasets that cannot be processed or shared openly due to personal data or copyright restrictions. The Language Bank of Finland already offers licenses, guidelines and services for restricted datasets. However, we need to develop and to document workflows and good practices in order to process, to store and to share more sensitive data requiring additional safeguards.



### EXAMPLE DATASETS

- matriculation examination essays (text)
- open-ended questionnaires (text)
- interviews (audio and video) or their transcripts
- sign language data (video)
- social media data (text, images, video, ...)

### CURRENT SERVICES

**Language Bank Rights (LBR/REMS by CSC)** is available for access management (federated authentication & authorization).  
<https://lbr.csc.fi/>

**Sensitive Data (SD) services (CSC)** are available for using confidential data within research teams. In this WP, the SD Desktop will be tested with audiovisual data.  
<https://docs.csc.fi/data/sensitive-data/>

**Deposition license agreement templates (fin+eng)** can be used for making data available via the Language Bank of Finland.  
<https://www.kielipankki.fi/support/dela/>

**CLARIN End-User Licenses (PUB/ACA/RES)** and public end-user licenses. For personal data, resource-specific data protection terms and conditions are included.  
<https://www.kielipankki.fi/support/clarin-eula/>

### DELIVERABLE 2.1.1 – Sep 2024

**Main goal:** Add more support for storing and processing unstructured resources containing personal data.

- Create and document a potential workflow for using audio and video data in SD Desktop.
- Investigate the legal possibilities for depositing social media data in the Language Bank.

### DELIVERABLE 2.1.2 – Sep 2025

**Main goal:** Enable academic researchers to share copyrighted datasets for the purpose of ‘verification of research’.

- Negotiate with UHEL and UEF to deposit the *Finnish presidential elections 2024 in social media* ([urn:nbn:fi:lb-2024030501](https://nbn-resolving.org/urn:nbn:fi:lb-2024030501)) and *Nordic Tweet Stream 2013-2023* ([urn:nbn:fi:lb-2024032221](https://nbn-resolving.org/urn:nbn:fi:lb-2024032221)) in the Language Bank.
- Investigate possibilities for combining SD Desktop with data from other sources (e.g., KAVI; see WPs 1.1 & 2.3).