

Privacy Notice for the Use of the *Giellagas Corpus of Spoken Saami Languages* corpus at Kielipankki (The Language Bank of Finland)

27 October 2025

This Privacy Notice concerns the study “*Evaluating Self-Supervised Automatic Speech Recognition on Northern Sámi Dialect Data*” (Master’s-level research project).

Data Controller:

Uppsala University

Thunbergsvägen 3H

752 38 Uppsala

Sweden

Researcher for this project: Gina Welsh (gina.welsh.7005@student.uu.se)

Data Protection Officer:

Jacob Håkansson

Segerstedthuset, Dag Hammarskjölds väg 7

752 37 Uppsala

Sweden

dataskyddsbud@uu.se

1. What is the purpose of the processing and what personal data will be processed?

The purpose of processing the data is to use it as a source of dialectal audio data for a Master’s-Level Research project about automatic speech recognition in Northern Saami dialects. The personal data processed would be audio recordings and corresponding transcripts which are linked to areas where Northern Saami is spoken. Personal data is processed solely for purposes intended to further research, knowledge, and insight for the benefit of the public. The data are obtained from an existing archival corpus and no new personal data are collected directly from individuals.

2. How will the personal data be stored and processed?

The corpus data will be stored on Uppsala University’s secure research servers, protected by institutional firewalls, multifactor authentication, and access control. Data are processed locally using encrypted drives and are not transferred to cloud services. Access is limited to the researcher named above and the supervisory team.

3. Other disclosure of personal data during research

Your personal data will not be disclosed to third parties. Only aggregated and anonymised research results will be published or shared. No identifiable audio or transcript data will appear in any publication.

4. Personal data to be processed in “Evaluating Self-Supervised Automatic Speech Recognition on Northern Sámi Dialect Data” (Research and Development Project)

This research involves the processing of audio data which may indirectly reveal speakers’ ethnic origin and linguistic background through dialectal features. However, no direct identifiers such as names or contact details will be processed.

5. Description of the personal data processed

This privacy notice is made to gain access to the *Giellagas Corpus of Spoken Saami Languages* to use as a primary source of audio data for a Master’s level project about automatic speech recognition of North Saami dialects. The corpus is obtained through the Language Bank of Finland; thus, personal data are not collected directly from the data subjects.

The metadata information for this data set can be found at: <https://clarino.uib.no/comedi/editor/lb-201407302>

The corpus download link is: <http://urn.fi/urn:nbn:fi:lb-2020112928>

More information about the corpus can be accessed here: <https://www.kielipankki.fi/corpora/giellagas/>

6. The lawful basis for processing personal data in scientific research

The processing of personal data in this project is based on Article 6(1)(e) GDPR – processing is necessary for the performance of a task carried out in the public interest, namely, scientific research conducted under the auspices of Uppsala University.

For special categories of personal data (e.g. data that may reveal ethnic or linguistic origin through speech recordings), the processing is further based on Article 9(2)(j) GDPR – processing necessary for scientific or historical research purposes subject to appropriate safeguards in accordance with Article 89(1).

In exceptional cases where consent is required by the data provider or repository, processing may also rely on Article 6(1)(a) GDPR (the data subject’s consent).

7. Transfer of personal data outside the EU/EEA area

In this research, your personal data will not be transferred outside the EU/EEA area.

8. Protection for personal data

No direct identification data is collected. The personal data processed in this research will be protected using user ID, password, and multi-factor authentication processes implemented by Uppsala University's IT department. The data is handled only by the data controller with a password secured computer and encryption software. All processing complies with Uppsala University's Information Security Policy and the General Data Protection Regulation (EU 2016/679)

9. Retention period of personal data

The data will be retained until February 2025 to allow for the completion and examination of the Master's project. After this, the local copy will be deleted in accordance with Uppsala University's research data policy. The original corpus remains under Kielipankki's control.

10. Rights of the data subject

Cancellation of consent (GDPR, Article 7)

You have the right to cancel your consent if the processing of personal data is based on consent. Such a cancellation has no impact on the lawfulness of consent-based processing conducted before the cancellation of consent.

Right to access your personal data (GDPR, Article 15)

You have the right to get to know whether and which personal data of yours are processed. If you wish, you can also request a copy of your personal data to be processed.

Right to rectification (GDPR, Article 16)

If there are any inaccuracies or errors in your personal data to be processed, you are entitled to request that these be rectified or supplemented.

Right to erasure (GDPR, Article 17)

You have the right to demand in some cases that your personal data be erased. However, the right of erasure is not applicable if the erasure would prevent or greatly hinder reaching the goal of processing in scientific research.

Right to restriction of processing (GDPR, Article 18)

You have the right to restrict the processing of your personal data in some cases, like when you challenge the correctness of your personal data.

Right to lodge a complaint with a supervisory authority (Article 13(2)(d))

If you believe your personal data have been processed unlawfully, you have the right to lodge a complaint with the Swedish Authority for Privacy Protection (Integritetsskyddsmyndigheten, IMY).

Website: <https://www.imy.se/>

Deviating from the rights

In some individual cases, it is possible to deviate from the described rights on the grounds stipulated in the GDPR and the Data Protection Act insofar as the rights would prevent or greatly hinder reaching the goals of scientific or historical research or statistical purposes. The need for deviating from the rights is always assessed case-specifically. It is also possible to deviate from the rights if the data subject cannot, or cannot any longer, be identified.

Profiling and automatised decision-making

In this research, your personal data will not be used for any automatic decision-making. In this research, the purpose of the processing of personal data is not to assess your personal qualities, i.e. profiling, but personal data and qualities are considered from the perspective of broader scientific research.