

Data Protection Impact Assessment

This is a brief assessment of whether the open publication of the dataset “Replication Data for: Listening to Accents: Comprehensibility, accentedness and intelligibility of native and non-native English speech” (henceforth *Dataset*) in TROLLing is aligned with applicable legal and research-ethical regulations and guidelines.

As TROLLing is located in Norway and the study was carried out in an EU country, the assessment is mainly based on the following legal and research-ethical frameworks:

- Common European guidelines on research integrity as described in the document “The European Code of Conduct for Research Integrity” (ALLEA - All European Academies, 2017).
- Common European legal regulations as described in the document “Ethics and data protection” (European Commission, 2021a) and in the Personal Data Act (Personopplysningsloven, 2018), which is the Norwegian implementation of the General Data Protection Regulation (GDPR).
- Common European research-ethical guidelines for the Social Sciences and Humanities (SSH) as described in the document “Ethics in Social Science and Humanities” (European Commission, 2021b), and as implemented in Norway in the document “Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora” (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021).

In addition, more specific (e.g., regional, institutional and/or domain-specific) regulations and guidelines are applied in case they are more restrictive than the common European and national Norwegian ones, or in other ways may be relevant for the assessment, to the extent such documents exist and could be provided by the data depositor.

The assessment was carried out in November 2022 by Philipp Konzett (TROLLing data curator) and Gil Verbeke (author, data depositor). The author/data depositor is responsible for the truthfulness of the assessment. The research organization(s) with the administrative responsibility for the Dataset (see the Producer field in the metadata record of the Dataset) is/are responsible for any legal and research-ethical consequences that may arise due to actions taken based on the assessment.

To assess whether the open publication of the Dataset in TROLLing is aligned with applicable legal and research-ethical rules and guidelines, the following four issues were addressed:

- 1) Does the Dataset contain personal data?
- 2) Did the participants provide valid consent to open publication of the Dataset? If the answer is no:
- 3) Are there other considerations that suggest that it is legally and research-ethically justifiable to publish the Dataset with open access? If the answer is yes:
- 4) Is it possible to inform the participants in a research-ethically responsible way about the (open) publication of the Dataset?

Question 1: Personal data?

Does the Dataset contain personal data? (Here we disregard publication/bibliographic and contact information about researchers who have been involved in the research project, as it is necessary for scholarly publications to contain such information.)

We may address this question by answering the following questions from the guide “Control of anonymity” issued by the Norwegian Centre for Research Data (Norwegian Centre for Research Data, n.d.), rendered in *italics* below:

1a) Do the data contain directly identifiable personal data such as personal identification numbers, names, addresses or phone numbers?

None of the data files in the Dataset contain directly identifiable personal data about the participants.

1b) Do the data contain indirectly identifiable personal data? This is determined by:

- *combining different information, e.g. geographical information/variables (municipality number, ward or district) with other background information such as income, education, occupation or position.*

The scrambling key connecting participants with the data has been deleted, following the information provided to the participants in the content form; see file "Listening_to_Accents_Informed_Consent.pdf".

The only information that potentially could identify the identity of individual participants, is the information provided in the documentation of the Dataset that the data were collected from 33 Flemish English as a Foreign Language (EFL) learners. However, as the total population of Flemish EFL learners is by far much larger than the selected sample, it is by all practical means not possible to identify any members of the sample.

The conclusion is therefore that it is not possible to indirectly identify persons by combining different information contained in the Dataset.

- *checking the sample size. In a small sample, information such as gender or age can make the personal data indirectly identifiable.*

Despite of the small sample size (33 Flemish EFL learners), the Dataset does not contain information about the participants that can make the participants indirectly identifiable; see the previous point.

- *checking the composition of the sample. If, for example, a sample's gender or age distribution is biased, it may be enough to make the data indirectly identifiable.*

The only potentially person-identifying information in the Dataset is the information that the participants are Flemish English as a Foreign Language (EFL) learners. As demonstrated above, this information cannot be said to be person-identifying.

- *checking the data's degree of detail. If, for example, the occupation and education variables are very detailed (four or five-digit level), the data set is more likely to be indirectly identifiable.*

The only potentially person-identifying information in the Dataset is the information that the participants are Flemish English as a Foreign Language (EFL) learners. As demonstrated above, this information cannot be said to be person-identifying.

Conclusion: The Dataset does not contain personal data. Thus, the processing of the data contained in this Dataset is no longer subject to applicable data protection regulations as described in the Personal Data Act (Personopplysningsloven, 2018).

Question 2: Valid Consent?

Did the participants provide valid consent to open publication of the Dataset?

The Dataset does not contain personal data and processing the data contained in the Dataset is thus not subject to data protection regulations as described in the Personal Data Act. Therefore, from a legal point of view, there is also no requirement for consent to further processing of the data in the Dataset. On the other hand, research-ethical principles expect that any processing of information from research projects in which natural persons are or have been involved should in principle be based on valid consent from the participants. This also holds for processing the data after the project has finished:

From a research ethics perspective, and in accordance with the principles of fair and transparent data processing, if you intend to use or make their data available for future research projects, it is best practice to obtain their additional, explicit consent to the secondary use of the data.
(European Commission, 2021a, p. 12)

The participants involved in the project resulting in the data contained in the Dataset consented among other thing to the following:

- “I give consent to the experimenter to use my data and share my data in a repository”
- “all data will be anonymised”

See the consent form in the file “Listening_to_Accents_Informed_Consent.pdf”.

It is questionable whether the wording “share my data in a repository” explicitly implies that the data will be shared with **open** access.

Conclusion: It is questionable whether the participants provided valid consent to open publication of the Dataset.

Consequently, other considerations regarding the open publication of the Dataset are to be evaluated (see next question).

Question 3: Other considerations in favour of open publication?

Are there other considerations that suggest that it is legally and research-ethically justifiable to publish the Dataset with open access?

Article 18 of “Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora” (‘Research Ethics Guidelines for the Social Sciences and Humanities’) applicable to research carried out in Norway allows for exceptions to the requirement for consent: “There are situations where it is justifiable to conduct research on persons, even if a research ethics consent is not obtained.” (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021, p. 20).

In the case in question, the researchers have obtained valid consent to do research about the participating persons, and that the data will be shared in a repository, but it is questionable whether the provided consent can be said to qualify as explicit consent to publish the anonymized research data from the project with **open** access. As the scrambling key connecting participants with the data has been deleted, it is no longer possible to contact the participants to obtain new consent.

According to NESH (email of April 5, 2022), one should therefore examine whether there are other considerations that suggest that it is legally and research-ethically justifiable to publish the Dataset with open access.

There are several circumstances that support the consideration of the open publication of the Dataset being legally and research-ethically justifiable:

- 1) Although it is questionable whether the participants' consent to share the data in a repository explicitly implies **open** access publishing of the data, the obtained consent as described in the consent form does not explicitly exclude that the Dataset can be openly shared at the end of the project.
- 2) The open publication of the data does not entail any risk of inconvenience for the participants, because
 - a. the participants cannot be identified (see Question 1 above);
 - b. none of the original potentially personally identifiable information was of the type "special categories of personal data".
- 3) On the other hand, publishing the data with open access will allow others to gain more insight into the topic and to build on their research or other kinds of knowledge production (e.g., in teaching activities or student assignments). Thus, the expected benefits from publishing the data openly outweigh by far the potential – or rather non-existing – risks of doing so.
- 4) As the Dataset has been anonymised, the processing of the Dataset is no longer subject to data protection regulations (see Question 1 above).
- 5) NESH article 6 urges researchers to share their research data as openly as possible: "Research material and results should be made available to others as openly as possible, to facilitate learning, testing and criticism." (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021, p. 12)
- 6) NESH article 25 calls for research results to be reported back to the participants: "Researchers must report the research results back to the participants in a comprehensible and responsible way" (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021, p. 23).

Conclusion: Despite that it is questionable whether the participants provided informed consent to open publication of the Dataset, there are several other, more weighty considerations that support the conclusion that it is legally and research-ethically justifiable to publish the Dataset with open access.

Question 4: Information to participants?

Is it possible to inform the participants in a research-ethically responsible way about the (open) publication of the Dataset?

Even if NESH under certain conditions allows the processing of personal data without consent (see Question 3 above), an information requirement still applies; cf. NESH article 18: "Researchers must in any case take care of the responsibility to inform [the participants]." (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021, p. 20)

As the Dataset is anonymous and the scrambling key connecting participants with the data has been deleted, it is not possible anymore to contact the participants.

However, according to NESH (email of April 8, 2022), the requirement to inform participants can be seen in the context of NESH article 25. Return of results: "If direct contact with each individual participant is disproportionately difficult or not possible, public dissemination of the research results can help fulfil this requirement." (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021, p. 23)

Conclusion: The actual publication of the Dataset with open access can be said to fulfil the requirement that the informants must be informed in a responsible way about the publication of the Dataset.

Overall Conclusion

Based on the assessment above, our overall conclusion is that the open publication of the data in the Dataset is in line with applicable legal and research-ethical rules and guidelines.

References

- ALLEA - All European Academies. (2017). *The European Code of Conduct for Research Integrity. Revised Edition*. ALLEA - All European Academies. <https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>
- Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH). (2021). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora* (5. utgave). De nasjonale forskningsetiske komiteene. <https://www.forskningsetikk.no/globalassets/dokumenter/4-publikasjoner-som-pdf/forskningsetiske-retningslinjer-for-samfunnsvitenskap-og-humaniora.pdf>
- European Commission. (2021a). *Ethics and data protection*. European Commission. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf
- European Commission. (2021b). *Ethics in Social Science and Humanities*. European Commission. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-in-social-science-and-humanities_he_en.pdf
- Norwegian Centre for Research Data. (n.d.). *Control of anonymity*. Retrieved 28 April 2022, from https://o.nsd.no/arkivering/en/control_anonymity.html
- Personopplysningsloven. (2018). *Lov om behandling av personopplysninger* (LOV-2018-06-15-38). Lovdata. <https://lovdata.no/dokument/NL/lov/2018-06-15-38/>