

## DPIA report for the role play with DELAD experts

### Format for documenting the risks and mitigation measures

This document for reporting addresses the main issues that need to be assessed in an official Data Protection Impact Assessment (DPIA). This format is based on [The data protection impact assessment according to article 35 GDPR, a practitioner's manual](#).

In an official DPIA report you should also address.

- a. Description of the processing operation (now limited to the case description).
- b. Further information and documentation (such as the information sheet, consent form and data management plan).
- c. Documentation of the legal basis.

#### Case description:

Alice, currently a post-doctoral researcher in AI & Data Science, has developed an algorithm for voice conversion as a major deliverable of her PhD project. This technique can alter the speech signal in a way so that the speakers could not be recognised by the third party by their voice while preserving the clarity of their speech. This kind of pseudonymised speech data is potentially useful for research in many other disciplines or research design, such as qualitative investigation using interviews where participants' verbal responses are usually audio-recorded for further analysis, or quantitative studies where the speech samples might be played to other listeners for auditory-perceptual judgements or other kinds of evaluation.

The algorithm was developed based on speech samples collected from typical adult speakers of English, German, and Dutch. The next step of Alice's research, as part of her present post-doctoral project, is to test this algorithm on other languages, speech samples of speakers of a wider age range (including children), and speech data from speakers with different types of speech disorders. Anita and her supervisor have recently heard of a couple of online resources where researchers can archive and share their speech data for the purpose of education and secondary analysis. They found a description of a dataset that comprises speech samples of 60+ Polish-speaking children with speech difficulties associated with hearing impairment<sup>1</sup>. Moreover, the format of the speech data files seems to be compatible with that required by the algorithm, which means that they can conveniently use the data files to run some tests right away. Therefore, Anita and her supervisor are thinking about using this dataset as a test of their algorithm on children with speech disorders, and want to find out how to gain access to it. The dataset is available under restricted access conditions. The repository sees it as its responsibility to contact the representatives of data providers for permission.

This is the report of a DPIA meeting of stakeholders of the Polish dataset discussing if and under which conditions the dataset can be made available for this research purpose in the light of the GDPR.

#### Roles:

- Alice, postdoc researcher, who submitted a request for the Polish dataset
- Anita, the researcher who collected the Polish dataset
- Erin, a lawyer who supports researchers in privacy by design for research
- Kate who represents the Polish data subjects as a member of a local association of hearing impaired speakers
- Archie is responsible for the data centre where the Polish dataset is stored
- Alan is the ICT specialist at this data centre
- Aine is a member of the ethics assessment committee that will review this proposal

---

<sup>1</sup> See also: <https://phonbank.talkbank.org/access/Clinical/PCSC.html>

Tip: Where possible, refer to the overarching principle of data minimisation, the six protection goals and the key concept cards.

- Confidentiality
- Availability
- Integrity
- Unlinkability
- Transparency
- Interveniability

Structure your response along the following three topics: risks, measures and final assessment.

### 1.Risks

Does the currently envisaged design of the data collection and processing ensure compliance with the respective data protection goals and the key concepts of the GDPR? If not, document under what circumstances a risk to the data protection goal is possible. Also document points of discussion.

Which risks are data subjects likely to face as a result of an infringement of their rights and the goals of GDPR?

*Anita: Audio is personal information, falls under GDPR, no personal questions in the record material, only prompted texts. There is metadata on family context and health.*

*Erin: GDPR allows sharing data for scientific research with appropriate safeguards. How about the content of the material: Children utter words. Safeguards should be clearly described.*

*Anita: Identity of speakers is limited in voice recordings because voices have changed since children are now grown up.*

*Kate: contact point for data subject is needed and must be explicitly mentioned so that data subjects know where they can go with their questions and concerns.*

*Erin: And mention how you will deal with their rights. Article 14 GDPR specifies what information needs to be available for the data subjects.*

*Archie: There is also a scientific risk if data subjects now request data erasure. If data subjects require their data to be erased at this stage, then this would be in conflict with the public interest of research institutes for reproducibility of research results. This is not possible anymore if part of the data disappears. Also this must be made clear on the website.*

*Aine: information needed to-reuse the data from ethical point of view:*

- documentation and clarity in terms of consent,
- there was no consent. So we need to resort to research need as public interest/legitimate interest of further processing for scientific research and that needs to be documented. DPIA is relevant for Ethics committee

*Only for similar research purpose, what is the risk here, since the request pertains to another type of research that was originally envisaged.*

*Only temporal use, delete it later on.*

*Erin: GDPR mentions measures such as pseudonimisation and the general principle of data minimisation. The DPIA helps to clarify who has to take which measure. Who is the controller? There can be more controllers. Then the essence of their roles must be clear to participants.*

*Repurpose, so ask participants again, or assess if the further re-use can also fall under the legitimate interest of scientific research and safeguards are in place to protect the rights of the data subjects. Important to involve representatives in the assessment and learn from their needs.*

*Alan: Technical protection measures need to be balanced with the goal of safeguarding the integrity of the data: Risk of encryption: Loss of encryption key -> data loss. Encryption also hampers playback of the data as it still sits in the archive.*

*Technical solutions*

*We retain the data and cannot erase. We have the necessary safeguards for further processing.*

## **2. Measures**

Which measures were discussed to address (high) risks to the data subjects?

Indicate which stakeholders should be involved to implement the measure.

*In the archive there is information about the point of contact for data subjects if they have questions.*

*Kate: Should also be added to website.*

*Clarify the safeguards that we have in place (action Alan).*

*Alice proposes another type of research than originally intended, but if the data is properly safeguarded then it is important that Alice's algorithm is also tested on other languages to avoid biases in the model.*

*Advice from ethics board about this innovative use would boil down to the question if they think a DPIA is needed. Ethics committee wants to know if back engineering can trace back to individuals in the dataset. The measure is that an expert is consulted about this.*

*License specified the terms of use in the conditions of the GDPR. Who can use the data (categories of recipients), for which purposes,*

*Check: what does it say about data retention?*

*Need to further specify terms of use for follow-up research, specifying the type of research allowed for follow-up research. Requests must be assessed by researchers.*

*This is in the interest of data subjects and researchers who want to use the data.*

## **3. Final assessment**

What is the conclusion of your group regarding necessity and proportionality of the collection and processing of personal data?

Did you conclude that the research and the development of the algorithm is feasible when taking identified risks and mitigation measures in account?

*Document all safeguards for sharing this data as discussed and put it on the webpage at the place where possible data subjects can find the information.*

*For innovative research we will state that we will consult the ethics committee about a DPIA. and we will call for external technical expert advice to ensure that the risks of back engineering the identities of data subjects are very low.  
We agree that sharing the data is justified by legitimate interest.  
Point of contact where data subjects can obtain information of what is done with the data and requests for re-use (according to Article 14 GDPR).*