The Interaction of Personal Data, Intellectual Property and Freedom of Expression in the Context of Language Research

Aleksei Kelli
University of Tartu, Estonia
aleksei.kelli@ut.ee

Krister Lindén
University of Helsinki, Finland
krister.linden@helsinki.fi

Pawel Kamocki
IDS Mannheim, Germany
kamocki@ids-mannheim.de

Kadri Vider
University of Tartu, Estonia
kadri.vider@ut.ee

Penny Labropoulou
ILSP/ARC, Greece
penny@ilsp.gr

Ramūnas Birštonas
Vilnius University, Lithuania
ramunas.birstonas@tf.vu.lt

Vadim Mantrov
University of Latvia, Latvia
vadims.mantrovs@lu.lv

Vanessa Hannesschläger
OeAW, Austria
vanessa.hannesschlaeger gmail.com

Riccardo Del Gratta
ILC, Italy
riccardo.delgratta ilc.cnr.it

Age Värv
University of Tartu, Estonia
age.varv@ut.ee

Gaabriel Tavits
University of Tartu, Estonia
gaabriel.tavits@ut.ee

Andres Vutt
University of Tartu, Estonia
andres.vutt@ut.ee

Esther Hoorn
University of Groningen, The Netherlands
e.hoorn@rug.nl

Jan Hajic
Charles University, Czechia
hajic@ufal.mff.cuni.cz

Arvi Tavast
Institute of the Estonian Language, Estonia
arvi@tavast.ee

Abstract

Language researchers are usually aware of intellectual property and personal data (PD) requirements. The problem, however, arises when these two legal regimes have conflicting requirements. For instance, when copyright law requires the acknowledgement of the author, but personal data law enshrines the data minimisation principle. It is a practical question for a language researcher whether he should name the author of the text used for, e.g., building a language model, or follow the data minimisation principle not to name the author.

The access right that a data subject has introduces similar conflicts. The question is what the scope of the access right is. Does it cover only processed personal data, or does it extend to data derived from PD?

Aleksei Kelli, Krister Lindén, Pawel Kamocki, Kadri Vider, Penny Labropoulou, Ramūnas Birštonas, Vadim Mantrov, Vanessa Hannesschläger, Riccardo Del Gratta, Age Värv, Gaabriel Tavits, Andres Vutt, Esther Hoorn, Jan Hajic Charles and Arvi Tavast 2022. The Interaction of Personal Data, Intellectual Property and Freedom of Expression in the Context of Language Research. *Selected papers from the CLARIN Annual Conference 2021. Ed. by Monica Monachini and Maria Eskevich.* Linköping Electronic Conference Proceedings 189, pp. 76–87. DOI: https://doi.org/10.3384/9789179294441

The interaction of the freedom of expression with PD protection entails several problems. The question is whether researchers can publish their research results containing personal data. The General Data Protection Regulation establishes a general framework that needs to be implemented by EU member states. We analyse different implementations based on examples from several EU countries.

1 Introduction

There is an awareness that intellectual property[1] and personal data[2] (PD) protection are relevant in language research. These two regimes are often applicable simultaneously, and their requirements might seem contradictory. Therefore, we have chosen three specific cases[3] to outline the interaction of intellectual property and personal data protection and provide preliminary guidance.

Firstly, we explore the interplay between the data minimisation principle and the right to be acknowledged as the author (the attribution/paternity right). On the one hand, the data minimisation principle enshrined in the General Data Protection Regulation (GDPR) requires processing[4] as little personal data as possible (Art. 5 (1) c)). According to the European Data Protection Board (EDPB) "Data minimisation substantiates and operationalises the principle of necessity" (2019: 21). On the other hand, the Berne Convention Art. 6bis, which sets the international standard and binds all current EU member states (and almost all of the remaining world), gives authors the attribution (paternity) right. The relevant question here is whether a researcher who has collected language data containing copyrighted content (for further discussion on the process of development of language technologies from the legal perspective, see Kelli et al 2020) should attribute the author of the content or follow the data minimisation principle and remove all personal data (e.g., the author's name) that is not necessary for processing.

The second case concerns intellectual property protection and the data subject's access right. A researcher might need to decide what data the access right covers in practical terms. Is it only raw personal data[5] or personal data derived from raw personal data?

Thirdly, we discuss the impact of personal data protection on freedom of expression since publications constitute research outcomes. The two previous questions do not require comparative analysis, but the situation is different in this case. Therefore, we rely on the General Data Protection Regulation implementation model of the following European Union countries: Austria, Czechia, Estonia, Finland, France, Germany, Greece, Italy, Latvia, Lithuania, and the Netherlands. Even though not all EU countries are studied, we can draw preliminary conclusions about the implementations. Adding more countries would not change the general picture.

[1] Intellectual property can be defined as "rights resulting from intellectual activity in the industrial, scientific, literary or artistic fields". Art. 2 of the Convention Establishing WIPO. IP is traditionally divided into three main categories: 1) copyright; 2) related rights to copyright; 3) industrial property.

[2] The General Data Protection Regulation (GDPR) defines personal data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Art. 4 (1)).

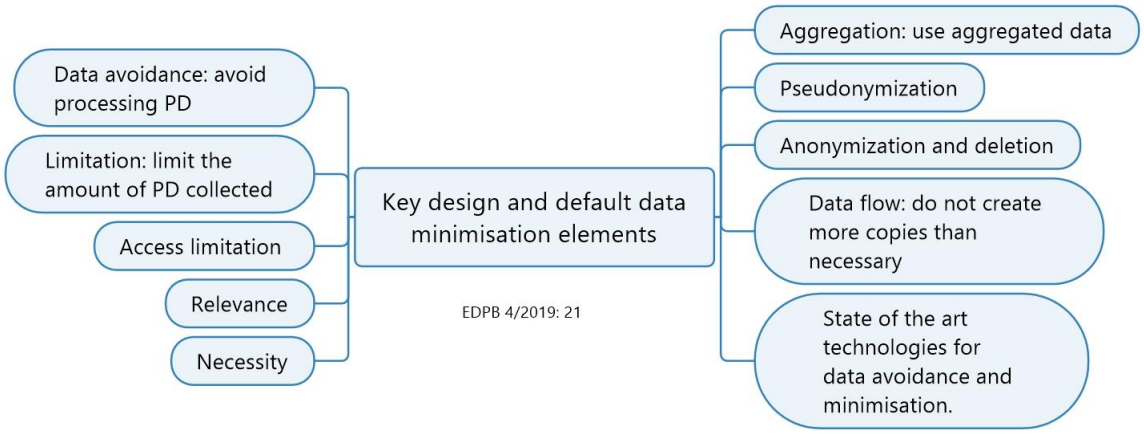
[3] It should be mentioned that there are several IP and PD protection interaction points whose systematic mapping is outside the scope of this article. Therefore, we chose cases that could potentially be relevant for language researchers.

[4] The GDPR defines processing of personal data as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" (Art. 4 (2)).

[5] In the context of this article, the concept of raw personal data refers to information such as age, height, weight, nationality, income, physical characteristics and so forth. Derived personal data is based on raw personal data (e.g., subject's profile as a consumer, the estimation of person's life expectancy and so forth).

2 The data minimisation principle and the right of attribution

According to the data minimisation principle, PD must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed" (GDPR Art. 5 (1) clause c). The European Data Protection Board (EDPB) outlines the data minimisation obligation for default processing with reference to GDPR Art. 25 (2) as containing the following elements: 1) amount of personal data collected (unnecessary data is not collected); 2) the extent of their processing (processing is limited to what is necessary); 3) the period of their storage (the retention period is no longer than necessary); 4) their accessibility (the access is limited to what is necessary) (2019: 12-14). The following graph visualises the data minimisation principle as conceptualised by EDPB (2019: 21):



The focus of the article is not on the data minimisation principle as such but on the identification of the data subject. The European Data Protection Board is of the following opinion: "Minimising can also refer to the degree of identification. If the purpose of the processing does not require the final set of data to refer to an identified or identifiable individual (such as in statistics), but the initial processing does (e.g. before data aggregation), then the controller shall delete or anonymise personal data as soon as identification is no longer needed. Or, if continued identification is needed for other processing activities, personal data should be pseudonymised to mitigate risks for the data subjects' rights" (2019: 21).

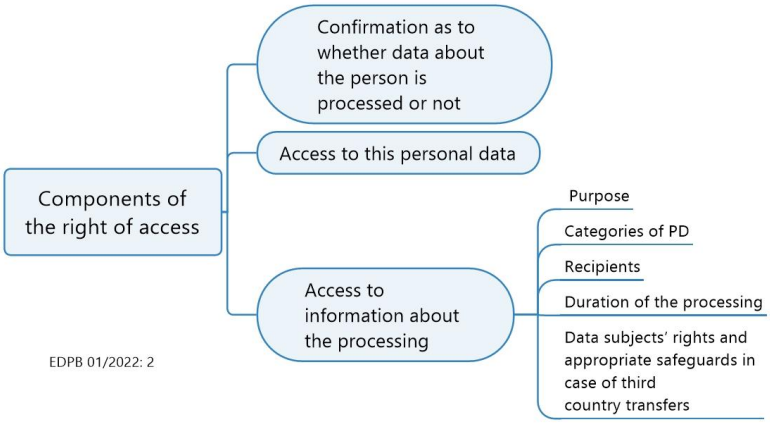
Pursuant to Art. 6bis (1) of the Berne Convention, "the author shall have the right to claim authorship of the work". The InfoSoc Directive also contains the obligation to identify the source (incl. the author's name), e.g., in the context of quotation or research exceptions (Art. 5 (3), esp. (a) and (d)). The EU case law reiterates the obligation (e.g., C-145/10). Copyright laws of different European Union member states contain the same principle. For instance, according to the Estonian Copyright Act "The author of a work has the right to appear in public as the creator of the work and claim recognition of the fact of creation of the work by way of relating the authorship of the work to the author's person and name upon any use of the work (right of authorship)" (§ 12 (1) clause 1). In other words, there is a legal obligation to acknowledge the author of a work. Therefore, it is compatible with the GDPR since it names compliance with a legal obligation as a legal basis for PD processing (Art. 6 (1) c)).

An overarching theme for this and the following section concerns legal obligations relating to derived data, e.g., data derived through text and data mining (TDM). For further discussion, see Kelli et al. (2020). Interestingly, the TDM exception contained in the DSM Directive does not require attribution. However, it should be borne in mind that the TDM exception only limits the reproduction right and does not allow any communication to the public. If the results of TDM are disseminated (e.g., based on quotation or research exception), the attribution right has to be honoured.[6]

[6] The attribution right exists only in case of the existence of copyrighted content. In the EU case law, it is pointed out that 11 consecutive words could be copyright protected (C-5/08). However, it does not say that less than 11 words are not copyrighted. For further discussion, see Kamocki 2020.

3 The right of access and intellectual property protection

In combination with the right to be informed and the principle of transparency, the access right forms a foundation for exercising the data subjects' rights. The access right requires the controller to provide information on the processing of PD, as well as access to the data (GDPR Art. 15). The European Data Protection Board conceptualises the right of access as follows (2022: 2):



The first question for research organisations and researchers (data controllers) is the scope of the access right. The question is whether the access right applies to raw personal data or personal data derived from raw personal data. In other words, this question asks what PD covers.[7] The Court of Justice of the European Union (CJEU) has not been remarkably consistent. For instance, it has explained that "There is no doubt that the data relating to the applicant for a residence permit and contained in a minute, such as the applicant's name, date of birth, nationality, gender, ethnicity, religion and language, are […] 'personal data' […] As regards, on the other hand, the legal analysis in a minute, it must be stated that, although it may contain personal data, it does not in itself constitute such data" (C-141/12 paragraphs 38, 39). In another case, the CJEU held that "the written answers submitted by a candidate at a professional examination and any comments made by an examiner with respect to those answers constitute personal data" (C-434/16).

Understandably, the concept of personal data should be interpreted consistently and extensively. However, there is no legal clarity on whether data derived from PD should be made available. WP29 (2016: 9) suggests in the context of the right of portability (see GDPR Art. 20) that "user categorisation or profiling are data which are derived or inferred from the personal data provided by the data subject, and are not covered by the right to data portability".

However, in its very recent draft guidelines, the EDPB (2022, paragraph 96) clearly states that not only the raw data provided by the data subject but also personal data derived and inferred from such data should be provided to the data subject who requests access to his or her personal data. It should be noted that non-personal data, even derived or inferred from the data subject's personal data, are not concerned with such requests. The data surrounding PD does not have to be made available as well.

Within the context of language research, the question is whether the data subject could require access to a language model trained using his PD. First, a model that does not contain any personal data is not concerned with the right of access. Moreover, the language model containing PD can be protected by intellectual property rights (database copyright, database *sui generis* right and trade secret[8]). The rightholder should have an exclusive right to decide who can access it. The GDPR accommodates this line of argument in its Recital 63, explaining the nature of the access right: "That right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in

[7] For the concept of PD, see WP29 2007.

[8] Article 21 of the trade secrets directive defines a trade secret as information not generally known, having commercial value and its holder has taken steps to keep it secret.

particular the copyright protecting the software. However, the result of those considerations should not be a refusal to provide all information to the data subject".

To sum up, the access right does not cover access to a language model containing PD as a whole, especially when this model can be considered a trade secret. In this context, IP rights prevail over data protection.

4 Data subject's rights and freedom of expression

4.1 General background

The data subject has the right to object to the processing and obtain the erasure, restriction or rectification of PD concerning himself (GDPR Art. 16, 17, 18, 21). These rights may conflict with the author's right to make his work available. This question can be framed as an interaction of personal data protection and freedom of expression (FoE). Personal data protection is not an absolute right (GDPR Rec. 4). Furthermore, freedom of expression is guaranteed by all major international human rights treaties and European legal acts, such as the Universal Declaration of Human Rights (Article 19), the Convention for the Protection of Human Rights and Fundamental Freedoms (Article 10), the Charter of Fundamental Rights of the European Union (Article 11). Therefore, the GDPR allows Member States to limit the data subject's rights to reconcile PD protection with the freedom of expression. According to GDPR Rec. 153, "This should apply in particular to the processing of personal data in the audiovisual field and in news archives and press libraries".

There are two intriguing questions concerning the interaction of personal data protection and freedom of expression:

1) how to strike a fair balance between personal data protection and freedom of expression in research settings? Freedom of expression is usually framed in the context of newspapers publishing facts about public figures. Freedom of academic expression is somewhat unclear. Still, it has been interpreted to apply to, e.g., x-ray pictures of medical case studies as standard practice. Such accompanying material is publicly disclosed in a scientific journal to illustrate the published case. There is no need to obtain the consent of the x-rayed person for this purpose. Usually, a person cannot be directly identified from such an x-ray. However, if the medical condition is rare, the individual may still be identifiable with the help of additional information. As the GDPR defines data concerning health as special categories of PD (Art. 9), this is an especially delicate example.

2) how and where to draw a line between processing for academic expression and research purposes. Research publication requires prior research. The question is whether this research is covered with the freedom of academic expression. We admit that the processing could be covered by the freedom of expression except when data is present in the research publication. It is important to emphasise that the principles of data minimisation, purpose limitation, accuracy, fairness (GDPR Art. 5), and other requirements need to be followed. Research quality and funding conditions often require the publication of research data to ensure reproducibility and verifiability of research results. Therefore, there is tension between the requirements on providing open data and protecting personal data. For further discussion, see Kelli et al. (2018).

4.2 Implementation models of EU countries to strike a fair balance between freedom of expression and personal data protection

The main aim of the General Data Protection regulation is to create a uniform regulatory framework throughout the European Union. However, some aspects of personal data protection are delegated to the EU member states. Striking a fair balance between freedom of expression and processing personal data is one of them.

According to Art. 85 (1) of the GDPR "Member States shall by law reconcile the right to the protection of personal data pursuant to this Regulation with the right to freedom of expression and information, including processing for journalistic purposes and the purposes of academic, artistic or literary expression". The GDPR Art. 85 (1) further specifies: "For processing carried out for journalistic purposes or the purpose of academic artistic or literary expression, Member States shall provide for exemptions or derogations from Chapter II (principles), Chapter III (rights of the data subject), Chapter IV (controller and processor), Chapter V (transfer of personal data to third countries or international organisations), Chapter VI (independent supervisory authorities), Chapter VII (cooperation and consistency) and Chapter IX (specific data processing situations) if they are necessary to reconcile the right to the protection of personal data with the freedom of expression and information".

Analyzing the implementation routes of selected EU countries to guarantee academic freedom of speech exemplifies several differences. The approach of the EU countries varies from countries that do not have any specific provisions to counties with a very detailed regulatory framework. Some countries are placed between the two.

4.2.1 Countries without specific provisions

For instance, the **German** Federal Data Protection Act (BDSG) does not contain any rules specifically implementing Article 85 of the GDPR. The existing broad derogation for research and archiving purposes (Article 27 of the BDSG) is based on Article 89, not 85 of the GDPR. It seems to be deemed sufficient by the legislator (Deutscher Bundestag, 2018). Specific state acts regarding media and journalistic expression exist in many federal states (Länder), e.g., Hessisches Pressegesetz or Landesmediengesetz Baden-Württemberg.

4.2.2 Countries with a general provision

Several studied countries have a general provision (with minor additions) limiting the applicability of the General Data Protection Regulation to protect freedom of expression. These countries are **Austria** (the Austrian Data Protection Amendment Act), **Finland** (the Finnish Data Protection Act), **Latvia** (the Latvian PDPA), and **Lithuania** (the Republic of Lithuania Law on Legal Protection of Personal Data). These general provisions as such are not very informative and, due to their similar character, are not presented here.

Although these countries have a literal implementation model, some still have some additional norms. For instance, Art. 7(2) of the Republic of **Lithuania** Law on Legal Protection of Personal Data further provides that the Inspector of Journalist Ethics shall monitor the application of the GDPR and the Law on Legal Protection of Personal Data and ensure that this legislation applies to the processing of personal data for journalistic purposes and academic, artistic or literary purposes. Therefore, deviating from the general principle, the Inspector of Journalist Ethics, not the State Data Protection Inspectorate, is responsible for supervising the processing of personal data for journalistic purposes and academic, artistic or literary purposes. Currently, the Lithuanian case law and the public decisions made by the Inspector of Journalist Ethics is too fragmented to make any conclusive statements about the interplay of the FoE and PD protection, but it seems that the intention is to interpret the concepts of "journalistic purposes" and "academic, artistic or literary purposes" broadly, as foreseen in Recital 153 of the GDPR.

Article 32(3) of the **Latvian** PDPA states that when processing data for academic, artistic or literary expression, provisions of the GDPR (except for Article 5) shall not be applied if all of the following conditions are present: 1) Data processing is conducted by respecting the right of a person to private

life, and it does not affect interests of a data subject which require protection and override the public interest; 2) Compliance with the provisions of the GDPR is incompatible with or prevents the exercise of the rights to freedom of expression and information.

As one may observe from this quoted provision, it is essentially based on Article 85 (2) of the GDPR and contains two parts. The Latvian legislator reacted to a necessity to provide exemptions or derogations from certain chapters of the GDPR. As one may observe from the phrase 'except for Article 5' contained in the quoted provision, the Latvian legislator chose to apply Article 5 (i.e. principles relating to the processing of personal data) to be observed while processing data for academic, artistic or literary expression from all provisions included in the relevant chapters of the GDPR. At the same time, the Latvian legislator provided two cumulative preconditions referred to in the quoted provision for processing data for academic, artistic or literary expression in order to avoid the application of the rules included in relevant chapters of the GDPR. Therefore, if at least one of these two cumulative preconditions is not met, the GDPR in full should be applied for such processing.

4.2.3 Countries with an elaborate provision

There are also countries with a more elaborate approach to the interaction of personal data protection and freedom of expression, such as **Czechia, Estonia, France, Greece, Italy** and the **Netherlands**.

Czechia seems to be the EU country with the most detailed regulation. In **Czechia**, the GDPR is implemented by a completely new law (No. 110/2019 Coll.). Art. 17 gives the legal basis for personal data processing for journalistic, academic, artistic or literary expression: (1) Personal data may also be processed if it serves, in a reasonable manner, journalistic purposes or purposes of academic, artistic or literary expression. (2) The processing of personal data for the purposes referred above is not subject to authorisation or approval of the Office and enjoys the right to protection of the source and content of information, even in the case of the processing of personal data in a manner allowing remote access.

Articles 18 to 22 are devoted to the exceptions related to the right of the subject to be informed, exceptions related to the protection of the source and contents of the personal information, exceptions to the right for corrections, deletion and restriction of processing, and the right to appeal. Some of the exceptions are, however, constrained in specific cases. Art. 23 provides further limitations allowed in the GDPR.

The last paragraph of Art. 23 contains a catch-all phrase related to the topic: (3) Where the exclusion or limitation of certain rights or obligations would be likely to result in a high risk to the legitimate interests of the data subject, the controller or processor shall, without undue delay, adopt and document appropriate measures to mitigate such or similar risk. It might be interesting to note that the word *academic* has indeed been used in line with the usual legal text practice, even though outside of the legal domain, it has a meaning very similar to the French expression 'académique' (see the discussion below about France), with several tens of "Academic" research institutes formed across the country.

In practice, the formal and often informal guidelines of the Czech Office for personal data protection[9] are that reporting according to the law should be minimised to clear cases of processing personal data, and only in the case it is not covered by other provisions of the law. For example, if human subjects performing tasks (non-medical) in a research project are being paid by the same institution, their personal data are being collected for processing their salaries and covered by the reporting done once by that institution for the purpose of employment. In such a case, no other reporting is necessary provided the writings, speech recording, survey results or other data collected from the subjects are anonymised (or collected anonymously) before they are stored and processed, which is often the case in linguistic research focused on data collection for machine learning in the area of language technology, where in fact individual differences are to be suppressed anyway to get generalised behaviour of the models and resulting software tools and applications.

The **Estonian** Personal Data Protection Act (Estonian PDPA) has two sections to protect freedom of speech (Section 4 and 5). Section 4 of the Estonian PDPA regulates the processing of personal data for

[9] Available at https://www.uoou.cz/en.

journalistic purposes[10] and is not addressed here. Section 5 concerns the processing of personal data for academic, artistic and literary expression, which is the focus of the article. According to Section 5 of the Estonian PDPA "Personal data may be processed without the consent of the data subject for the purpose of academic, artistic and literary expression, in particular, disclosed if this does not cause excessive damage to the rights of the data subject". The Explanatory Memorandum to the Estonian PDPA emphasises that the regulation applies inter alia to books, motion pictures, visual art, biographies and other content that does not qualify as journalism. According to the memorandum, consent as a legal basis for processing PD for academic, artistic and literary expression is not required. The reason is that consent can be withdrawn, which could have an adverse impact on the freedom of expression. Although the law allows processing PD without consent, it is necessary to strike a fair balance between FoE and privacy (2018: 13-14).

Article 80 of the **French** Data Protection Act attempts to reconcile data protection and freedom of expression in France. It derogates from two general principles: the storage limitation and the prohibition of processing sensitive data (including data about criminal convictions and offences). It also limits information rights, access, rectification and restriction, and derogates from the rules on data transfers. This framework applies only when necessary to safeguard freedom of expression and information, and only when the data are processed: 1) for academic ('universitaire'), artistic or literary expression, or 2) for journalistic purposes by professional journalists, in a way that respects ethical rules (deontology) of the profession. The Article clearly states that other laws and codes regarding violations of privacy and reputational damage continue to apply.

One can be surprised by the adjective 'universitaire' in Article 80 of the French Copyright Act (expression universitaire, artistique ou littéraire) rather than 'académique' (as in 'academic, artistic or literary expression'). However, the same wording is used by the French version of Article 85 of the GDPR. This is because 'académique' has a very restricted meaning in French (related to the Académie Française) and should not be interpreted as limiting the derogatory framework to processing made by scholars with a university affiliation.

Article 28 of the **Greek** Personal Data Protection Act, corresponding directly to the GDPR (Art. 85), aims to reconcile the right to personal data protection with the right to freedom of expression and information, "including the processing for journalistic purposes and for purposes of academic, literary or artistic expression". More specifically, in the framework of these objectives, Paragraph 1 of this Article explicitly enumerates cases where the processing of PD is allowed: "(a) when the subject of the data has given his explicit consent, (b) for PD that have been publicised by the subject, (c) when the right to the freedom of expression and the right to information overweighs the right to PD protection, especially for topics of general interest or when the PD relates to public persons, and (d) when it is restricted to the necessary measure to ensure the right of expression and the right of information, especially with regard to sensitive categories of PD[11], and criminal cases, and security-related measures, taking into account the right of the subject to his private and family life." We can deduce that the Article looks more into the 'journalistic purposes' rather than 'academic purposes'. Paragraph 2 of the same Article provides the exceptions and derogations for processing for such purposes, which are mentioned in Article 85 of the GDPR.

Article 136 of the **Italian** Personal Data Protection Code (PDPC) implements Art. 85 of the GDPR. It regulates journalistic as well as academic works. Article 137 defines the categories of PD that can be processed without the data subject's consent. Namely, such categories are special categories of PD and

[10] The Estonian Personal Data Protection Act § 4: "Personal data may be processed and disclosed in the media for journalistic purposes without the consent of the data subject, in particular disclosed in the media, if there is public interest therefor and this is in accordance with the principles of journalism ethics. Disclosure of personal data must not cause excessive damage to the rights of any data subjects".

[11] This is really a different approach from the Netherlands. The Dutch variant links Art. 85 GDPR to academic expressions. That leads to legal uncertainty whereas the main principles of the GDPR are oriented towards purposes and legal grounds. The Greek approach seems a clarification of the legal ground of the public interest Art. 6 GDPR. The emphasis on public persons is in line with the case law on freedom of expression. This raises the question to what extent journalistic purposes really are comparable to academic purposes.

PD data related to criminal convictions and offences (GDPR Art. 9, 10). Other sections further restrict these categories to "Safeguards applying to the processing of genetic data, biometric data, and data relating to health" (section 2-f) and "Processing entailing a high risk for the performance of a task carried out in the public interest" (section 2-p). Article 137 (3) provides: "It shall be allowed to process the data concerning circumstances or events that have been made known (communicated/disseminated) either directly by the data subject or on account of the data subject's public conduct".

In the **Netherlands** Art. 85 GDPR is implemented in a broad way in article 43 of the Uitvoeringswet AVG. The Article speaks of the reconciliation of rights for journalistic and academic expressions. As the Dutch lawmaker explains in the parliamentary discussion (Memorie van Antwoord UAVG 2017-2018, 34 851, first chamber) Recital, 153 of the GDPR calls for broad implementation. Yet, still, an assessment of the proportionality has to take place. If needed, transparency requirements and access rights of Chapter 3 GDPR can be disregarded. This raises the question of how researchers should deal with transparency requirements inherent to academic research and verifiability. No exemption is possible from the obligation on data protection by design (Art. 25 GDPR). So, a balancing act and applying data protection principles, like data minimisation, is still mandated. Yet, for instance, when necessary, in urgent cases, based on this derogation, a researcher can refrain from filling in questions about the intended processing, where the institution generally would want this in their shared role of controller, based on the obligation to maintain records of processing activities (article 30 GDPR). This is in line with article 1.6 of the Netherlands Higher education and Research Act, which states that academic freedom is taken into account in universities of the Netherlands. A data protection impact assessment as a method for privacy by design could be used to document the balancing act and design a protocol for verifiability.

4.2.4 Different implementation models and a way forward

Different implementation models raise the question of their potential impact. As a general observation, we would emphasise that Article 85 of the General Data Protection Regulation itself is rather vague. There is a good reason for this. The right to freedom of speech does not have a clear scope. Freedom of speech (also academic freedom of speech) is probably differently defined in different EU countries. This means that what could be protected as academic freedom of speech in Estonian is not necessarily identical to France or Greece. The cultural differences are probably reflected and reinforced in divergent GDPR Art. 85 implementation models that are not limited to laws but also extend to legal practice. Since research is becoming increasingly international, this could be a problem.

Personal data protection and freedom of expression are both human rights. This means that one is not prioritised over another. The critical issue is to strike a fair balance between them, as personal data protection should not affect academic freedom of expression. While openness promotes transparency and accuracy of the research, protective measures promote confidentiality. Both aspects are needed to preserve trust in research among fellow researchers and the general public. The tension might be resolved using the principles on Findability, Accessibility, Interoperability and Reusability of data (FAIR 2016) advocated by the European Commission (COM(2020) 66). In practical terms, it means making the metadata open and providing a clear protocol for accessing the content even if it is not offered openly on the internet.

5 Conclusion

We reached the following preliminary conclusions. Firstly, the data minimisation and the attribution right are not contradictory concepts. The acknowledgement of the author is compatible with the GDPR as the compliance with a legal obligation. The attribution does not concern all personal data but only data that is copyrighted.

Secondly, the access right primarily applies to raw personal data. There is no legal clarity regarding the access to data derived from personal data. The information not containing personal data is not within the scope of the access right (even if the information is derived from personal data). The access right

could be exercised to get access to personal data derived from raw personal data. This is not the case when such access could conflict with intellectual property rights and trade secret protection. Trade secret protection considerations could be more relevant here.

Thirdly, personal data protection usually does not take precedence over the freedom of expression and cannot hinder the academic freedom of speech and the author's right to disseminate his work. However, there could be restrictions on how it can be disseminated. Conducting research may also be covered by academic freedom of speech. Although the General Data Protection Regulation provides a framework to enhance freedom of speech in the field of academic research, its implementation by different EU countries diverges. It is not necessarily compatible with the GDPR's aim to establish a uniform framework for processing personal data for research and academic expression and could have a negative impact on the dissemination of research results. Therefore it is advisable to rely on the principles of Findability, Accessibility, Interoperability and Reusability of data.

References

Austrian Data Protection Amendment Act. Entry into force 2018. Available at https://www.ris.bka.gv.at/eli/bgbl/I/2018/31 (3.4.2021).

Berne Convention. *Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886*. Available at https://wipolex.wipo.int/en/text/283698 (9.2.2022).

Charter of Fundamental Rights of the European Union. - OJ C 326, 26.10.2012, p. 391-407. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT (4.2.2022).

Convention Establishing WIPO. Convention Establishing the World Intellectual Property Organization (as amended on September 28, 1979). Available at https://wipolex.wipo.int/en/text/283854 (8.2.2022).

Convention for the Protection of Human Rights and Fundamental Freedoms. Rome, 4.XI.1950. Available at https://www.echr.coe.int/documents/convention_eng.pdf (4.2.2022).

Czech Republic Law No. 110/2019 Coll. Available at https://www.zakonyprolidi.cz/cs/2019-110 (4.2.2022).

C-434/16. *Case C-434/16.* Peter Nowak v Data Protection Commissioner (20 December 2017). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62016CJ0434&qid=1617649690306 (5.4.2021).

C-141/12. *Joined Cases C-141/12 and C-372/12.* YS (C-141/12) vs Minister voor Immigratie, Integratie en Asiel, and Minister voor Immigratie, Integratie en Asiel (C-372/12) v M, S (17 July 2014). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0141&qid=1617633666683 (5.4.2021).

C-145/10. *Case C-145/10.* Eva-Maria Painer vs Standard VerlagsGmbH and Others (1 December 2011). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62010CJ0145&qid=1618044850444 (10.4.2021).

C-5/08. *Case C-5/08.* Infopaq International A/S vs Danske Dagblades Forening (16 July 2009). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1555243488182&uri=CELEX:62008CJ0005 (10.4.2021).

COM(2020) 66. *A European strategy for data.* Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions**.** Brussels, 19.2.2020. COM(2020) 66 final. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN (15.1.2022).

Deutscher Bundestag, 2018. *Deutscher Bundestag,* Ausarbeitung: Die Öffnungsklausel des Art. 85 der Datenschutz-Grundverordnung, WD 3 - 3000 - 123/18. Available at https://www.bundestag.de/resource/blob/560944/956f5930221c807984d40c1df2af5abf/WD-3-123-18-pdf-data.pdf (26.04.2021).

DSM Directive. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. *OJ L 130, 17.5.2019*, pp. 92-125. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1572352552633&uri=CELEX:32019L0790 (10.4.2021).

Dutch Implementation Act General Data Processing Regulation, In force from: 16th of May 2018, available at: Uitvoeringswet Algemene verordening gegevensbescherming, Permanent link to Artikel 43 Uitvoeringswet

Algemene verordening gegevensbescherming. Available at https://wetten.overheid.nl/BWBR0040940/2021-07-01 (20-1-2022)

EDPB 2022. *European Data Protection Board.* Guidelines 01/2022 on data subject rights - Right of access. Version 1.0 (for public consultation). Adopted on 18 January 2022. Available at https://edpb.europa.eu/system/files/2022-01/edpb_guidelines_012022_right-of-access_0.pdf (1.2.2022)

EDPB 2019. *European Data Protection Board.* Guidelines 4/2019 on Article 25 Data Protection by Design and by Default. Version 2.0. Adopted on 20 October 2020. Available at https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf (8.2.2022).

Estonian Copyright Act. *Copyright Act.* Entry into force 12.12.1992. Available at https://www.riigiteataja.ee/en/eli/504032021006/consolide (9.2.2022).

Estonian PDPA. *Personal Data Protection Act.* In force from: 15.01.2019. Available at https://www.riigiteataja.ee/en/eli/523012019001/consolide (6.2.2022).

EU Charter of Fundamental Rights. *Charter of Fundamental Rights of the European Union.* 2012/C 326/02. OJ C 326, 26.10.2012, p. 391-407. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (1.4.2021).FAIR 2016. *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0.* Available at https://www.force11.org/fairprinciples (15.1.2022).

Explanatory Memorandum to the Estonian PDPA (2018). Available at https://www.riigikogu.ee/tegevus/eelnoud/eelnou/5c9f8086-b465-4067-841e-41e7df3b95af (7.2.2022).

Finnish PDPA. *Data Protection Act (1050/2018)*. Available at https://www.finlex.fi/en/laki/kaannokset/2018/en20181050.pdf (8.4.2021).

French Data Protection Act. *Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*. Available at https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/ (27.4.2021).

French Intellectual Property Code. *Code de la propriété intellectuelle.* Available at https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006069414/ (27.4.2021).

German Federal Data Protection Act. *Bundesdatenschutzgesetz vom 30. Juni 2017 (BGBl. I S. 2097)*, das durch Artikel 12 des Gesetzes vom 20. November 2019 (BGBl. I S. 1626) geändert worden ist. Available at: https://www.gesetze-im-internet.de/bdsg_2018/BJNR209710017.html (27.4.2021).

GDPR. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *OJ L 119*, 4.5.2016, p. 1-88. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1555312258399&uri=CELEX:32016R0679 (1.4.2021).

Greek Data Protection Act. *Personal Data Protection Law (4624/2019)*. Available at: https://www.e-nomothesia.gr/kat-dedomena-prosopikou-kharaktera/nomos-4624-2019-phek-137a-29-8-2019.html (26.4.2021).

InfoSoc Directive. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. *Official Journal L 167*, 22/06/2001 P. 0010 – 0019. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1555254956114&uri=CELEX:32001L0029 (4.4.2021).

Italian PDPA. *Personal Data Protection Code containing provisions to adapt the national legislation to Regulation (EU) 2016/679* of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Available at https://www.garanteprivacy.it/documents/10160/0/Data+Protection+Code.pdf/7f4dc718-98e4-1af5-fb44-16a313f4e70f?version=1.3 (8.4.2021).

Kamocki, Pawel. 2020. When Size Matters. Legal Perspective(s) on N-grams. *Proceedings of CLARIN Annual Conference 2020. 05 – 07 October 2020*. Virtual Edition. Ed. Costanza Navarretta, Maria Eskevich. CLARIN, 166-169. Available at https://office.clarin.eu/v/CE-2020-1738-CLARIN2020_ConferenceProceedings.pdf (10.4.2021).

Kelli, Aleksei, Arvi Tavast, Krister Lindén, Kadri Vider, Ramunas Birštonas, Penny Labropoulou, Irene Kull, Gaabriel Tavits, Age Värv, Pavel Stranák, Jan Hajic. 2020. The Impact of Copyright and Personal Data Laws on the Creation and Use of Models for Language Technologies. In: Kiril Simov, Maria Eskevich (Ed.). *Selected*

Papers from the CLARIN Annual Conference 2019 (53−65). Linköping University Electronic Press. Available at https://ep.liu.se/ecp/172/008/ecp20172008.pdf (10.4.2021).

Kelli, Aleksei, Tõnis Mets, Lars Jonsson, Krister Lindén, Kadri Vider, Ramūnas Birštonas, Age Värv. 2018. Challenges of Transformation of Research Data into Open Data: the Perspective of Social Sciences and Humanities. *International Journal of Technology Management and Sustainable Development*, 17 (3), 227−251.

Latvian PDPA. *Latvian Personal Data Processing Law.* Available at https://vvc.gov.lv/image/catalog/dokumenti/Personal%20Data%20Processing%20Law.doc (26.4.2021).

Lithuania PDPA. *The Republic of Lithuania Law on Legal Protection of Personal Data.* Available at https://www.e-tar.lt/portal/legalAct.html?documentId=43cddd8084cc11e8ae2bfd1913d66d57 (26.4.2021).

Netherlands Higher education and Research Act, In force from: 8th of October 1992, Available at: https://wetten.overheid.nl/BWBR0005682 (7.2.2022).

Trade secrets directive = Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. – OJ L 157, 15.6.2016, p. 1-18. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0943 (4.2.2022).

Universal Declaration of Human Rights. Available at https://www.un.org/en/about-us/universal-declaration-of-human-rights (4.2.2022).

WP29 2016. *Article 29 Working Party (WP29)*. Guidelines on the right to data portability. Adopted on 13 December 2016. Available at https://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf (5.4.2021).

WP29 2007. *Article 29 Working Party (WP29)*. Opinion 4/2007 on the concept of personal data. Adopted on 20th June. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (5.4.2021).