

---

# LEGAL ASPECTS OF PROCESSING PERSONAL DATA IN DEVELOPMENT AND USE OF DIGITAL LANGUAGE RESOURCES: THE ESTONIAN PERSPECTIVE<sup>1</sup>

Liina Jents, Aleksei Kelli

Faculty of Law

University of Tartu

Kaarli 3, 10119 Tallinn, Estonia

Phone: (+372) 56 508 059 (Aleksei Kelli)

E-mail: [aleksei.kelli@ut.ee](mailto:aleksei.kelli@ut.ee); [liina.jents@borenius.ee](mailto:liina.jents@borenius.ee)

Received on 17 February 2014, accepted on 13 March 2014

doi:10.13165/JUR-14-21-1-08

## Introduction

Today, only around one million people speak Estonian as their mother tongue. The majority of the speakers (around 900,000 people) live in the territory of modern Estonia. Also, approximately 160,000 people speak the language in Russia, US, Sweden, Canada, Finland and many other countries<sup>2</sup>. Under these circumstances, it is crucial not to underestimate the importance of preservation and development of national languages, especially in small countries, such as Estonia, where the number of speakers is small and decreasing.

---

1 This publication has been supported by the European Social Foundation through the Research and Innovation Policy Monitoring Programme.

2 Estonian Institute. *Estonian in a World Context*. [interactive]. [accessed on 26-01-2014]. <[http://www.estonica.org/en/Society/The\\_Estonian\\_Language/Estonian\\_in\\_a\\_world\\_context/](http://www.estonica.org/en/Society/The_Estonian_Language/Estonian_in_a_world_context/)>.

In addition to cultural and national identity functions that a language has, it is also an important communication tool in politics, business and everyday life. In Estonia, the importance of preservation and development of the Estonian language has also been recognized on the constitutional level<sup>3</sup>. The preamble of the Constitution of the Republic of Estonia states that “[w]ith unwavering faith and a steadfast will to strengthen and develop the state [...] which shall guarantee the preservation of the Estonian nation, language and culture through the ages”<sup>4</sup>.

As of 1 May 2004, the Estonian language is also one of the 23 official languages of the European Union. The European Union considers preservation and development of national languages of its Member States very important, and promoting multilingualism is one of European Union’s priorities<sup>5</sup>. However, communication between European citizens is inevitably challenged by these multilingual barriers. Language technology and linguistic research can make a significant contribution to removing those barriers<sup>6</sup>. Language technologies also play a substantial role in mobile information services, computer-assisted language learning software, eLearning environments, self-assessment tools and plagiarism detection software<sup>7</sup>.

Today, language-technology applications that are available for the Estonian language<sup>8</sup> include spelling checkers, machine translation, speech synthesis (converting written text into speech), and speech recognition<sup>9</sup> (converting speech to written text)<sup>10</sup>. None of these systems at their present level of development really understand or speak natural languages in the human sense. Instead, they use sophisticated technology to imitate human language behaviour and extract unambiguous information from human contribution<sup>11</sup>.

3 For further discussion, see Kelli, A.; Tavast, A.; Pisuke, H. Copyright and Constitutional Aspects of Digital Language Resources: The Estonian Approach. Tallinn. *Juridica International*. 2012, 19.

4 *Eesti Vabariigi põhiseadus* [The Constitution of the Republic of Estonia] (as entering into force on 3.7.1992). RT 1992, 26, 349; RT I, 27-4-2011, 1 [interactive] (in Estonian). [accessed on 26-01-2014]. English translation available via <http://www.legaltext.ee/>.

5 *An Inventory of Community Actions in the Field of Multilingualism* – 2011 update. Commission Staff Working Paper. Brussels, July 7, 2011, SEC(2011) 926 final.

6 Liin, K. et al. *The Estonian Language in the Digital Age/Eesti Keel Digiajastul*. White Paper Series/Valge Raamatu Sari, 2012, p. 40.

7 *Ibid.*, p. 39.

8 For the Estonian language, a number of technologies and resources exist, but considerably less than for major languages, such as English. This is naturally understandable due the enormous difference in speakers of these two languages. It should also be noted that in Estonia digital language resource databases are generally not created for direct commercial purposes, but mainly for language research. Despite that, they can and do have commercial uses, as well.

9 During the last decade, the research on automatic speech recognition in Estonia has been carried out mainly at the Laboratory of Phonetics and Speech Technology, Institute of Cybernetics at Tallinn University of Technology.

10 See Kelli, A.; Tavast, A.; Pisuke, H., *supra* note 3, p. 42.

11 *Ibid.*, p. 41.

There are two main types of systems that “acquire” language skills in a similar manner to humans. The first system is called the statistical approach. According to language researchers, “[s]tatistical (or ‘data-driven’) approaches obtain linguistic knowledge from vast collections of concrete example texts. While it is sufficient to use text in a single language for training, e.g., a spell checker, parallel texts in two (or more) languages have to be available for training a machine translation system. The machine learning algorithm then “learns” patterns of how words, short phrases and complete sentences are translated”<sup>12</sup>. The second approach to language technology and to machine translation in particular is called the rule-based approach. In order to build such rule-base systems, “[e]xperts in the fields of linguistics, computational linguistics and computer science first have to encode grammatical analyses (translation rules) and compile vocabulary lists (lexicons). This is very time consuming and labour intensive”<sup>13</sup>.

Speech interaction is one of many application areas that depend on speech technologies, i.e., technologies for processing spoken language. Using machine learning techniques, language models can also be generated automatically from speech corpora, i.e., large collections of speech audio files and text transcriptions<sup>14</sup>. Those speech audio files are often accompanied by different information about the speaker, e.g., his age, mother tongue, place of residence (not the exact address, but the city or village), etc.

There are several legal issues relating to digital language resources and language technologies. These issues can be divided into two main categories. The first category concerns copyright and to some extent related rights<sup>15</sup>, and the second one is focused on personal data protection. There is some preliminary research conducted in the first category<sup>16</sup>. Since the previous research is also relevant for this article, then the results are concisely summarized in the following paragraph.

The main issues of digital language resources in copyright context relate to the definition of language resources and mechanisms allowing their creation. Drawing on the existing knowledge base, digital language resources can be defined as “databases whose content consists of many written and oral texts. Selection of the texts to be included in this database is a creative<sup>17</sup> process that requires collection and

12 Liin, K. *et al.*, *supra* note 6, p. 40.

13 *Ibid.*, p. 40.

14 *Ibid.*, p. 52.

15 Mainly rights of performers and *sui generis* database makers.

16 See Tavast, A.; Pisuke, H.; Kelli, A. Õiguslikud väljakutsed ja võimalikud lahendused keeleressurside arendamisel (Legal Challenges and Possible Solutions in Developing Language Resources). *Eesti Rakenduslingvistika Ühingu Aastaraamat*. 2013 (9): 317–332; Kelli, A.; Tavast, A.; Pisuke, H. *supra* note 3.

17 Since identification and collection of relevant texts require time and resources, then digital language resources could also be protected as *sui generis* database. Directive on the legal protection of databases regulates the protection of *sui generis* databases. Directive 96/9/EC

systematisation of the material for inclusion in this database”<sup>18</sup>. Development of such digital language resources requires an extensive use of copyrighted texts. This can be based on contractual model (consent is sought from right holders) or exception model (development and use is based on copyright exception, and therefore, not requiring right holders’ consent). From a societal perspective, it is best to move forward with the exception model. However, it would be recommendable to include a specific copyright exception regulating text analysis and data mining<sup>19</sup>.

Personal data protection in the context of digital language resources has not gained sufficient attention in the legal research literature. The aim of this article is to fill this gap.

The development of digital language resources involves use of material, which could include or even be considered itself personal data. Since it is crucial to use contemporary texts as “raw material”, then it is likely that actual personal data can be part of the processed written and oral texts. Personal data protection is not even confined to life time of data subjects, but, similarly to copyright protection, it is extended beyond. According to the Estonian Data Protection Act<sup>20</sup>, “[a]fter the death of a data subject, processing of personal data relating to the data subject is permitted only with the written consent of the successor, spouse, descendant or ascendant, brother or sister of the data subject, except if consent is not required for processing of the personal data or if thirty years have passed from the death of the data subject”<sup>21</sup>. Additional relevant question within the context of use of oral speech is whether a person’s voice can be considered her personal data. Based on the above mentioned, it can be concluded that development of digital language resources involves processing of personal data, in addition to use of copyrighted material. According to general principle, the processing requires the consent of a data subject or it is allowed by law. The authors’ main argument is that it is allowed to develop digital language resources without the consent of a data subject. The reason is that the development of language resources is done to enhance public welfare. This is vitally important for small languages, such as Estonian. In case law, it is interpreted and applied in a

---

of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. OJ L 77, 27.3.1996, p. 20–28.

18 See Kelli, A.; Tavast, A.; Pisuke, H., *supra* note 3, p. 41.

19 It should be added to the 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 (InfoSoc Directive). OJ L 167, 22.6.2001, p. 10–19. At the moment, there is the Public Consultation on the review of the EU copyright rules, which could lead to inclusion of this exception. See [http://ec.europa.eu/internal\\_market/consultations/2013/copyright-rules/index\\_en.htm](http://ec.europa.eu/internal_market/consultations/2013/copyright-rules/index_en.htm) [accessed on 15-02-2014].

20 *Isikuandmete kaitse seadus* [Data Protection Act] (as entering into force on 01.01.2008). RT I 2007, 24, 127; RT I, 30.12.2010, 2 (in Estonian). [interactive]. [accessed on 26-01-2014]. Unofficial translation available via <http://www.legaltext.ee/>.

21 Estonian Data Protection Act, § 13(1).

way, which prohibits development of digital language resources. Then, it would result sooner or later in extinction of small languages.

In the following sections, the authors examine processing of personal data for scientific research. Thereafter, focus is put on specific aspects of processing personal data in the context of digital language resources.

## 1. Processing of personal data for scientific research and other purposes

The Estonian Personal Data Protection Act (PDPA), the Data Protection Directive<sup>22</sup> and the Proposal for a Data Protection Regulation<sup>23</sup> define personal data in a similar way. According to all definitions, personal data constitute information, which relates to an identified or identifiable natural person<sup>24</sup>. More detailed characteristics of personal data are given through the data subject's definition<sup>25</sup> that also bounds the personal data concept with the terms "an identified natural person" and "a natural person who can be identified". The Estonian law also makes a difference between ordinary personal data and sensitive personal data. The processing of sensitive personal data requires following additional rules prescribed by the law and registration of processing with the competent authority. Sensitive personal data is among other data revealing ethnic or racial origin, data on genetic information, biometric data<sup>26</sup>.

---

22 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, 23/11/1995, p. 0031–0050.

23 Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. COM(2012) 11 final.

24 Subsection 4(1) of PDPA defines personal data as "any data concerning an identified or identifiable natural person, regardless of the form or format in which such data exist". According to Article 2 of the Data Protection Directive, personal data means "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity". According to Article 4(2) of the Proposal for a Data Protection Regulation, "personal data means any information relating to a data subject".

25 Under Article 4(1) of the Proposal for a Data Protection Regulation, 'data subject' means an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

26 PDPA, § 4(2).

Although the definitions in the Estonian law, Directive and Proposal for Regulation slightly differ, but what they have in common are two main elements: 1) the data is related to a natural person and 2) the natural person is identified or identifiable. There are no difficulties determining what personal data are for an identified natural person. It is data that allow direct identification and can easily be associated with a data subject and reveal his identity. To determine what is “identifiable” can be more complex. Identifiability exists if there is a mere possibility to establish a link between the data and a person. The mere possibility to identify is sufficient to consider this data personal data<sup>27</sup>. It can be stated that a person is identifiable when the data collected regarding a person are specific enough to allow a connection between the respective data and the person concerned. For instance, this is the case if the data include the age and school of a person and there is only one person at this age in this school. A crucial issue here is how far the “mere possibility” goes.

Is it enough when establishment of a link is possible or should we also consider the effort to do it? The posed question is answered by Recital 26 of the Data Protection Directive, which pays particular attention to the term “identifiable”. It reads as follows: “whereas to determine whether a person is identifiable account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person”. This means that a mere hypothetical possibility to single out the individual is not enough to consider the person as “identifiable”<sup>28</sup>.

The next issue concerns the question from which moment does “identifiable data” become “non-identifiable” or “anonymous data”? In practice, it is very difficult to draw the line between the two. There is no definition of “anonymous data” in the Estonian law, the Data Protection Directive or the new Proposal for Regulation. The Directive and the Proposal for Regulation only provide that the principles of data protection should not apply to data rendered anonymous in such a way, that the data subject is no longer identifiable<sup>29</sup>. Several countries have a definition for anonymous data or the law provides terms rendering the data anonymous. For example, under German law, “[r]endering anonymous means the modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labour be attributed to an identified or identifiable individual”<sup>30</sup>. In Data Protection Aspects Guidelines

27 Rosier, K.; Vereecken, I. *Data Protection Aspects within the Framework of Socio-Economic Research*. Institute for Employment Studies, Report 415, 2004, p. 3.

28 *Opinion 4/2007 on the Concept of Personal Data*. Article 29 of Data Protection Working Party. 01248/07/EN WP136. 2007 [interactive]. [accessed on 26-01-2014]. <<http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/>>.

29 Recital 23 of the Proposal for Regulation and Recital 26 of the Data Protection Directive.

30 Section 3(6) of the German Federal Data Protection Act. January 14, 2003 [interactive]. [accessed on 26-01-2014]. Translation available at: [http://www.gesetze-im-internet.de/englisch\\_bdsch/englisch\\_bdsch.html](http://www.gesetze-im-internet.de/englisch_bdsch/englisch_bdsch.html).

drawn up within the European Commission's RESPECT project, anonymous data are defined as "[d]ata that cannot be qualified as personal data, since they do not (anymore) allow direct or indirect identification of the data subject"<sup>31</sup>. The reason why it is important to make this distinction is that unlike from processing of identifiable data, processing of anonymous data is not subject to legal data protection requirements. Therefore, if data processed in the course of scientific research can be considered anonymous data, then there is no need for the consent of the data subject or to rely on the scientific research exemption.

This brings us to the next question. Namely, what constitutes the processing of personal data? According to the Estonian law, the processing of personal data is any act performed with personal data, including collection, recording, organisation, storage, alteration, disclosure, granting access to personal data, consultation and retrieval, use of personal data, etc.<sup>32</sup> The law does not make a difference if personal data are processed manually or automatically – it is still processing of personal data in the meaning of the Personal Data Protection Act. It was already established in the end of the nineties that the concept of processing also includes operations performed by software and hardware without the knowledge of the data subject, and hence, they are invisible to them<sup>33</sup>. It should be noted that the list of activities in the law is illustrative and under law any act performed with personal data can be considered processing of personal data. In case of scientific research, the processing of personal data happens when researchers collect the data from data subject, record it to its system, organize it and store it. It is important to remember that processing carried out to render data anonymous is also considered processing of personal data. Therefore, in order to escape the legal requirements arising out of processing of personal data, the processed data must already be anonymised<sup>34</sup>. Even if researchers collect the personal data themselves and make it anonymous, it is still considered processing of personal data, for which the consent of the data subject is needed. Until the moment data are made anonymous, the researcher must comply with all the legal requirements for the processing of personal data. The person, who operates with this anonymous data, does not require the consent of the data subject anymore, and the consent is also not required to transfer the data or make it available to this person as the data are anonymised (it is not personal data anymore).

---

31 Rosier, K.; Vereecken, I., *supra* note 27, p. 4.

32 PDPA, § 5.

33 Recommendation 1/99 adopted on February 23<sup>rd</sup> 1999, on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware by the Working Party on the Protection of Individuals with regard to the Processing of Personal Data set up by Directive 95/46/EC of the European Parliament and of the Council of 24 October, 1995, 5093/98/EN/final.

34 Rosier, K.; Vereecken, I., *supra* note 27, p. 4.

The key element in data protection regulations is the term “purpose”. The purpose of data processing is what defines the scope of processing and determines whether such processing is considered lawful. The Data Protection Directive and national laws, *inter alia* the Estonia Data Protection Act, refer to scientific and statistical purposes as a particular category of purpose. According to the explanatory memorandum of the Personal Data Protection Act<sup>35</sup>, scientific research should be interpreted widely, as in Subsection 2(7) of the Organisation of Research and Development Act<sup>36</sup>. The Organisation of Research and Development Act defines research as “independent creative work of a person undertaken by means of scientific research in order to increase the stock of knowledge of man, nature and society and their interaction”<sup>37</sup>. It is also stated in the explanatory memorandum that in general such scientific research should be conducted by research and development institutions registered in the respective register maintained by the Minister of Education and Research<sup>38</sup>, but it does not limit it to this. The Data Protection Directive does not include any definition for scientific research. The Proposal for a Data Protection Regulation prescribes that “scientific research for the purposes of this Regulation should include fundamental research, applied research, and privately funded research and in addition should take into account the Union’s objective under Article 179 (1) of the Treaty on the Functioning of the European Union of achieving a European Research Area”<sup>39</sup>. The Proposal for a Data Protection Regulation allows wide interpretations with clearly including privately funded research.

The Estonian Data Protection Inspectorate holds a view that processing of personal data for scientific and statistical purposes can be divided into three main categories:

- 1) with the consent of the data subject;
- 2) with the permission of the competent authority; or
- 3) in an anonymised (impersonalized) form<sup>40</sup>.

---

35 *Isikuandmete kaitse seaduse seletuskiri* [Explanatory Memorandum of the Personal Data Protection Act] (in Estonian). [interactive]. [accessed on 26-01-2014]. <<http://www.aki.ee/sites/www.aki.ee/files/elfinder/.../IKS%20SELETUSKIRI.rtf>>.

36 *Teadus- ja arendustegevuse korralduse seadus* [Organisation of Research and Development Act] (as entering into force on 02.05.1997). RT I 1997, 30, 471; RT I, 28.06.2012, 13 (in Estonian). [interactive]. [accessed on 26-01-2014]. Unofficial translation available via <http://www.legaltext.ee/>.

37 Organisation of Research and Development Act, § 2(7).

38 Explanatory Memorandum of the Personal Data Protection Act, *supra* note 36.

39 Recital 126 of the Proposal for the Data Protection Regulation.

40 *Andmekaitse Inspektsioon. Juhend: Isikuandmete Töötlemine Teadusuuringus* [Data Protection Inspectorate. Guideline: Processing of Personal Data in Scientific Research] (in Estonian). 2013 [interactive]. [accessed on 26-01-2014]. <[www.aki.ee/sites/www.aki.ee/.../teadusUuringute\\_juhis\\_veebr2013.rtf](http://www.aki.ee/sites/www.aki.ee/.../teadusUuringute_juhis_veebr2013.rtf)>.



The authors of this analysis are of the opinion that there is also a fourth category, which includes processing of disclosed personal data. The Estonian Data Protection Act prescribes that if a data subject has disclosed his or her personal data, has given the consent specified in paragraph 12 of this Act for the disclosure thereof or if such personal data have been disclosed on the basis of law, including subsection (2) of this section, then other sections of this Act do not apply to the processing of personal data. This means that, for instance, if the scientific researcher collects the “raw material” from a person’s blog and this blog also includes personal data that this person himself has disclosed to the public, then it can be used without the consent of the data subject. The law does give the data subject an option to demand, at all times, that the person processing disclosed personal data discontinue such activity unless otherwise provided by law and provided that this is technically possible and does not result in disproportionately high costs<sup>41</sup>. It should be noted that in case processing of data is part of scientific research, it might be technically impossible to stop use of this data or, if possible, result in disproportionately high costs. As a result, it is very hard for a data subject to stop the use of disclosed personal data. The situation becomes complicated when a person discloses someone else’s personal data without the proper consent of the concerned person. For instance, a person’s blog post or an article includes his friend’s personal data and he does not have a proper consent. Under these circumstances, the processor of data cannot rely on this clause, since the data has not been disclosed by the data subject himself or with his consent.

The authors of this article are of an opinion that the language research constitutes scientific research in the meaning of the data protection regulations. Language researchers need to follow all the rules provided above when processing personal data in development of digital language resources.

## 2. Specific aspects of processing personal data in the context of digital language resources

Linguists and developers of language technologies are facing different legal challenges. Data protection laws apply to the collection and use of both oral speech and written text. The authors of this paper start with addressing the specific issues relating to oral speech and then focus on problems common to both of them.

Gathering samples of oral speech also entails the collection of different personal data, such as name, age, nationality, place of residence, place of residence during first years of life of a sample provider. The collected personal data concerning a person providing a speech sample is essential for language researchers to understand and to analyse the speech and the spoken texts<sup>42</sup>.

---

41 Article 11(4) of the Data Protection Act.

42 The respective information has been disclosed to the authors by the experts at Laboratory of Phonetics and Speech Technology, Institute of Cybernetics at Tallinn University of Technology.

The central question here is whether a person's voice can be deemed her personal data? The Estonian Data Protection Inspectorate has in its different publications considered a person's voice to be his personal data<sup>43</sup>. It has even gone further and categorized a person's voice as dynamic biometric data<sup>44</sup>. Pursuant to the Estonian law, biometric data belong to the category of sensitive personal data<sup>45</sup>. Article 29 of the Data Protection Working Party<sup>46</sup> has also considered a person's voice to be biometric data of a person<sup>47</sup>. The categorization of a person's voice as sensitive personal data would require that the language researcher and everyone else, who uses voice recordings for whatever reason, should register their activities with the Data Protection Inspectorate. This would obviously be disproportionate obligation for the language researchers and language technology developers. This interpretation would also mean that anyone, who processes any oral material (e.g., radio broadcaster, who broadcasts any spoken texts), should register its activities with the Data Protection Inspectorate. The explanatory memorandum of the Personal Data Protection Act has emphasized that there are some cases, where in practice it is not realistically possible to apply the rules of processing of sensitive personal data. This is the case, when that data is used extensively in everyday life and registration of such processing would be clearly disproportionate. The explanatory memorandum has listed face image, signature and voice of a person as such data<sup>48</sup>. Therefore, a person's voice should definitely not be considered personal data in the context of the development digital language resources or any other context.

Both oral and written texts can include personal data. These texts can include personal data about the speaker or author himself or about a third person. It is more likely that spontaneous speech as a special type of oral speech contains third person's personal data. This is less likely if speakers follow previously prepared texts. Those texts usually do not contain any personal data.

---

43 See *Andmekaitse Inspektsiooni Juhend. Isikuandmete Töötlemine Töösuhetes. Abistav juhendmaterjal* [Data Protection Inspectorate's Guideline: Processing of Personal Data in an Employment Relationship. Assisting Guideline]. Tallinn, 2010 [interactive]. [accessed on 26-01-2014]. <<http://www.aki.ee/et/juhised/isikuandmete-tootlemine-toosuhetes>>.

44 *Andmekaitse Inspektsiooni Juhend. Biomeetriliste Isikuandmete Töötlemine*. [Data Protection Inspectorate's Guideline: Processing of Biometric Personal Data]. Approved 27.12.2007. Amended 01.03.2013. Tallinn, 2013 [interactive]. [accessed on 26-01-2014]. <[www.aki.ee/sites/.../Biomeetriliste%20isikuandmete%20töötlemine.rtf](http://www.aki.ee/sites/.../Biomeetriliste%20isikuandmete%20töötlemine.rtf)>.

45 PDPA, § 4(2), clause 5.

46 Article 29 of Data Protection Working Party was set up under the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Additional information available at: [http://ec.europa.eu/justice/data-protection/article-29/index\\_en.htm](http://ec.europa.eu/justice/data-protection/article-29/index_en.htm).

47 Opinion 4/2007 on the concept of personal data, *supra* note 29.

48 Explanatory Memorandum of the Personal Data Protection Act, *supra* note 36, p. 5.

Language researchers processing personal data in the course of development of digital language resources can basically choose from four different options:

- acquire the consent of the data subject;
- ask permission from the competent authority;
- use the data in an impersonalized (anonymised) form; or
- use the data that has been disclosed by the data subject or with its consent or if such personal data has been disclosed on the basis of law.

The safest approach is to acquire a sufficient consent from the data subject, whose data are being processed. This, however, is not always easy or even possible. It should also be emphasised that personal data obtained for a specific purpose (e.g., scientific research) cannot be used for any other purpose without the consent of the data subject. Therefore, if the personal data has been obtained just for research purposes, this data cannot be used for commercial purposes. This means that the consent of the data subject should also include the consent for any planned secondary uses of the personal data. The following of this suggestion is, unfortunately, quite complicated. It is especially so in cases, when the texts are collected not only for use by linguists themselves, but the texts are also made available at a language corpus. In this case, all possible uses of the texts that include the data cannot be predicted.

If the consent has been acquired to process data just for scientific research, then the researcher should only process personal data for research purposes. Other uses are not allowed. There are some ways how to overcome this obstacle. The CLARIN<sup>49</sup> report suggests that “[i]f personal data have been collected with insufficient rights for distribution or secondary use, there may still be some options, e.g., anonymisation for distribution or certain exemptions for scientific, historical or statistical research purposes”<sup>50</sup>.

Anonymisation of data is one possible solution, besides the scientific research exception. Once the personal data has been made anonymous, it can be freely communicated to any third party. It is no longer considered personal data anymore in the meaning of the data protection laws. At least in theory, this is a good solution, but, as discussed above, it is not always easy to determine if anonymous data is actually anonymous data or if the person is still identifiable. Therefore, in practice, this might not always provide legal certainty.

---

49 CLARIN is a short name for the Common Language Resources and Technology Infrastructure, which aims at providing easy and sustainable access for scholars in the humanities and social sciences to digital language data (in written, spoken, video or multimodal form) and advanced tools to discover, explore, exploit, annotate, analyse or combine them, independent of where they are located. Estonia is one of co-founders of CLARIN. More information available at: <http://www.clarin.eu/>. [accessed on 26-01-2014].

50 CLARIN. *A Report Including Model Licensing Templates and Authorization and Authentication Scheme*. EC FP7 project no. 212230, Deliverable: D7S-2.1 Version 5, final 2010, p. 45.

There is no regulation in the Estonian law concerning anonymous data or anonymization of data. The problem in small countries (with inhabitants around one million people as in Estonia) is whether anonymous data is not really anonymous. For example, when a sample of a child's speech is made available at a speech corpus with accompanying information about his school, age, place of residence at the first three years of his life, then only one child with those characteristics might be at this school. This means that he can be identified by this data without any unreasonable effort. If available data allow identification of a person, it is no longer anonymous data, but personal data, which are governed by the respective data protection regulations.

Some countries have defined anonymized data and adopted a regulation for anonymization in the research context. The German law requires depersonalization of personal data as soon as the research purpose allows it<sup>51</sup>. The Danish law prescribes that when a research project has been completed, the personal data must be destroyed or anonymized<sup>52</sup>. It should be noted that in practice this is also not always a good solution. It is not always possible to destroy the data right after its use or to make it anonymous as a whole because the consent for processing of personal data also usually includes a copyright licence to use the performance or work, and the data are retained for performance of the respective licence agreement. For this reason, there is a clear need for guidelines for making data anonymous and for use of this anonymous data. When the data is processed for research purposes, the liability of researchers should be limited if they have taken reasonable steps to make and retain the data anonymous. Once reasonable steps have been taken to make the data anonymous (even if the data has been collected only for scientific purposes) everyone should also be able to use the texts with anonymous data for any, *inter alia* commercial, purposes.

It should be emphasized that language researchers are not interested in personal data of an author or a third party contained in the text. The texts are intended only for the extraction of information about how language works, by both human researchers and machine learning algorithms in language-technology applications<sup>53</sup>. In the course of developing digital language resources, the texts containing personal data are collected, reproduced and stored. These activities constitute processing of personal data. It is so, despite the fact that researchers most likely do not even perceive the personal data in these texts.

The user of texts containing personal data can rely on the Personal Data Protection Act if the data has been disclosed by the data subject himself, with his consent or on the basis of law<sup>54</sup>. Problems arise when the data processed have in

---

51 De Cock Buning, M. *et al.* *The Legal Status of Research Data in the Knowledge Exchange Partner Countries*. The Netherlands: Centre for Intellectual Property Law (CIER), Molengraaff Institute for Private Law, Utrecht University, 2011, p. 20.

52 De Cock Buning, M. *et al.*, *supra* note 45, p. 17.

53 See Kelli, A.; Tavast, A.; Pisuke, H., *supra* note 3, p. 43.

54 PDPA, § 11.

fact been disclosed by a third person without the consent of the data subject or without any basis in law. Placing an obligation on language researchers to obtain consents for all such uses from all potential data subjects would clearly be overly burdensome and, in most cases, even impossible. In case texts containing personal data are processed with completely different objectives, such as research of language or creation of language technologies without even perceiving the data in the texts, then these activities should not fall under the personal data regulation or the law should prescribe a clear exemption that allows the processing without the consent of the data subject.

According to the law, when the data that have not been disclosed by the data subject himself and when it is impossible to obtain the consent of the data subject for processing of this data, processing is allowed only if the data are anonymous (it can no longer be considered personal data) or permission is asked from the competent authority. Acquisition of the permission from the competent authority is an option only in case of research and not in cases where the use is commercial (e.g., creation of commercially used language technologies). Obtaining the permission from the competent authority is quite burdensome and complicated for researchers, since due to massive amounts of used texts he is not aware of the personal data in these texts. As a result, it is impossible to disclose to the respective authority which data are processed. It should be emphasized again that language researchers and developers of language technologies are not interested in the personal data in texts. Their only interest is how the language works.

Digital language resources are generally created in public interest. They have an essential part in digitalization and development of national languages, which are at a heart of national identity and culture. This kind of activities could even be protected and enhanced by the European Convention on Human Rights<sup>55</sup> (ECHR). It has been said in the legal literature that “academic social investigation not only falls within the broad scope of art.10 protection but, due to its general concern to systematically contribute to improving societal knowledge about serious and important things, is part of that category of “high value” expression which the European Court of Human Rights has said should generally not be subject to legal restriction”<sup>56</sup>. A parallel could be drawn here with the development of digital language resources. The development of digital language resources encompasses linguist’s right to freedom of expression provided in Article 10 of the European Convention on Human Rights that can also be categorised as a “high value” expression. Additionally, it is also an important contribution to society as a whole. The same holds true for the use of

---

55 European Convention on Human Rights. [interactive]. [accessed on 16-02-2014]. <[http://www.echr.coe.int/Documents/Convention\\_ENG.pdf](http://www.echr.coe.int/Documents/Convention_ENG.pdf)>.

56 Erdos, D. Freedom of Expression Turned on Its Head? Academic Social Research and Journalism in the European Privacy Framework. *Public Law*. 2013, (1): 66.

the digital language resources for the creation of language technologies. It benefits not only developers of language technologies, but also society that gets valuable communication tools.

It should also be emphasised that the right to respect for private and family life is not an absolute right. This means that on certain legal basis interference is allowed<sup>57</sup>. Although the second sentence of section 26 of the Estonian Constitution should provide a comprehensive list of justifications for interference, additional justifications have been accepted in case of compelling values<sup>58</sup>. It should be considered whether development and use of digital language resources of the Estonian language involve such compelling values that take precedence over a person's right to private life. The Supreme Court has explained that, according to the preamble of the Constitution, the state must guarantee preservation of the Estonian people and the Estonian culture through the ages. It can be considered an objective justifying the interference with other rights<sup>59</sup>.

The constitutional guarantee of protection of the Estonian language could be a justification to develop digital language resources. It is very hard to predict whether the Estonian language would survive without digitalization. This means that preservation and development of the Estonian language can outweigh a person's right to private life (protection of personal data). It is important to bear in mind that the interference with the rights of the data subject is quite minimal (texts containing personal data are usually available in the form of books, newspaper article, blog posts, etc.).

Also, the Data Protection Directive does not rule out the possibility to use texts containing personal data by researchers without the consent of the data subject. The Data Protection Directive provides several grounds when the processing is legitimate. According to clause (f) of Article 7 of the Data Protection Directive, personal data may be processed only if “processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject [...]”. Consequently, Article 7(f) sets out two cumulative conditions that must be fulfilled in order for the processing of the personal data to be lawful: (1) the processing of the personal data must be

57 Article 26 of the Estonian Constitution provides that everyone is entitled to inviolability of his or her private and family life. Government agencies, local authorities, and their officials may not interfere with any person's private or family life, except in the cases and pursuant to a procedure provided by law to protect public health, public morality, public order or the rights and freedoms of others, to prevent a criminal offence, or to apprehend the offender.

58 Madise, Ü. *et al. Eesti Vabariigi põhiseadus. Kommenteeritud väljaanne. Kolmas, täiendatud väljaanne* (Constitution of the Republic of Estonia. Commented Edition. Third Edition). Tallinn: Juura (in Estonian), 2012 [interactive]. [accessed on 26-01-2014]. <<http://www.pohiseadus.ee/ptk-2/pg-26/>>.

59 Judgment of the Estonian Supreme Court of 3 May 2001 in Case 3-4-1-6-01.

necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed; (2) such interests must not be overridden by the fundamental rights and freedoms of the data subject. Digital language resources are created in public interest. Their use for the development of language technologies benefits the public, who will be able to use these technologies in their everyday life. The development and use of digital language resources have an essential part in the preservation of national languages.

The European Court of Justice has confirmed<sup>60</sup> that Article 7(f) of Directive 95/46 has a direct effect. It has asserted that “Article 7(f) of Directive 95/46 is a provision that is sufficiently precise to be relied on by an individual and applied by the national courts”<sup>61</sup>. This means that even though the Estonian data protection law does not include an equivalent of Article 7(f), this exception can still be relied on by individuals. In this decision, the European Court of Justice also found that “Article 7(f) of that directive precludes a Member State from excluding, in a categorical and generalised manner, the possibility of processing certain categories of personal data, without allowing the opposing rights and interests at issue to be balanced against each other in a particular case”<sup>62</sup>. This means that one should be able to balance the constitutional guarantee of preservation and development of the Estonian language in conjunction with language researchers’ and developers’ right to expression against the data subject’s right to private life. Striking a balance between these rights, it is crucial to bear in mind that language researchers use texts, which are already made available. Therefore, language researchers’ and developers’ rights to expression and public interest in preserving the Estonian language can overview data subjects’ right to privacy and private life. This could be an acceptable basis and justification for use of personal data without the consent of the data subject for development and use of digital language resources. This means that the development and use of digital language resources comply with conditions set forth with Article 7(f) of the Data Protection Directive. First, preservation and development of the Estonian language and freedom of expression constitute legitimate reasons for processing personal data. Second, these interests are not overridden by data subjects’ fundamental rights and freedoms, since the damage to these rights is negligible.

It is possible to rely directly on Article 7(f) of the Data Protection Directive. Still, the most clear cut solution could be the adoption of an additional exemption or extension of the existing ones in the personal data protection regulations. This new exemption should allow processing of oral and written texts, containing personal

---

60 Case C-468/10 *Asociación Nacional de Establecimientos Financieros de Crédito (ASNEF) v Administración del Estado* and Case C-469/10 *Federación de Comercio Electrónico y Marketing Directo (FECMD) v Administración del Estado* of the European Court of Justice. [2011] I-12181.

61 Cases C-468/10 and C-469/10 of the European Court of Justice, *supra* note 54, paragraph 52.

62 Cases C-468/10 and C-469/10 of the European Court of Justice, *supra* note 54, paragraph 48.

data for the purpose of development and use of digital language resources without the consent of the data subject and without any additional registration obligation. This kind of processing should be allowed at least for scientific purposes when there is no commercial objective. The authors would even go further and allow the same activity for the commercial purposes. This would facilitate the development of language technologies for commercial purposes. New language technologies clearly benefit society and enhance welfare.

The other option is to provide a more general exemption, covering unintentional use of data. An example of this exemption could be taken from the copyright regulation. Namely, the InfoSoc Directive provides an exception for incidental inclusion of a work or other subject-matter in other material and allows the use of this material without the right holder<sup>63</sup>. An analogue exemption that covers “incidental inclusion of a personal data” in any text processed would benefit language researchers. They have no interest in the personal data and usually do not even perceive the existence of such data. However, they are still processing it within the meaning of the data protection laws.

There could be an additional guarantee added to balance the interests of both parties. It could be provided that such use should not unreasonably undermine the fundamental rights and freedoms of the data subject.

## Conclusions

Language research offers tremendous opportunities for the European Union to address the complex issue of multilingualism in Europe. It is vitally important to small languages, such as the Estonian language, to survive in the internet era with a shrinking population.

It is crucial that the regulatory framework at the EU level supports and enhances the development of digital language resources and language technologies and thereby benefits the multilingual knowledge society.

Issues relating to personal data protection are amongst the biggest legal challenges in the context of development of digital language resources. Language researchers and developers of language technologies usually have to deal with personal data that they do not have any interest in or even knowledge of. Sometimes, data related to the data subject are needed to better understand recorded oral speech.

The authors are of the opinion that in cases, where the data allow identification of a person, but it is used only for scientific purposes, the rules should be less strict. Data protection law must allow language researchers to concentrate on their research activities without the fear of facing liability for unlawful use of personal data.

---

63 The InfoSoc Directive, Article 5(3)(i).



The issues with anonymous data and making data anonymous should also be addressed on the EU or at least national levels. Solid and clear guidelines for researchers on this topic are much desired.

The importance of protecting personal data and thereby privacy should not be underestimated. There are, however, other values, as well. It is crucial to protect and foster the preservation and development of national languages in the digital era on national and the EU level. The interpretation of the Estonian Constitution and the EU Personal Data Protection Directive allows us to say that language researchers can rely on these regulations and analyse different texts without the consent of the data subject. The interference with the privacy rights is minimal because usually these texts are published before or the data subject himself has provided this data during an interview. The authors suggest that there is a need to go even further and provided *expressis verbis* as an exception to the data subject rights allowing development of language resources. This would create legal certainty and help language researchers to come up with even better results.

## References

---

- Andmekaitse Inspektiooni Juhend. Biomeetriliste Isikuandmete Töötlemine [Data Protection Inspectorate's Guideline: Processing of Biometric Personal Data]. Approved 27.12.2007. Amended 01.03.2013. Tallinn, 2013 [interactive]. [accessed on 26-01-2014]. <[www.aki.ee/sites/.../Biomeetriliste%20isikuandmete%20töötlemine.rtf](http://www.aki.ee/sites/.../Biomeetriliste%20isikuandmete%20töötlemine.rtf)>.
- Andmekaitse Inspektioon. Juhend: Isikuandmete Töötlemine Teadusuuringus [Data Protection Inspectorate. Guideline: Processing of Personal Data in Scientific Research] (in Estonian). 2013 [interactive]. [accessed on 26-01-2014]. <[www.aki.ee/sites/www.aki.ee/.../teadusuuringute\\_juhis\\_veebr2013.rtf](http://www.aki.ee/sites/www.aki.ee/.../teadusuuringute_juhis_veebr2013.rtf)>.
- Andmekaitse Inspektiooni Juhend. Isikuandmete Töötlemine Töösuhetes. Abistav juhendmaterjal [Data Protection Inspectorate's Guideline: Processing of Personal Data in an Employment Relationship. Assisting Guideline]. Tallinn, 2010 [interactive]. [accessed on 26-01-2014]. <<http://www.aki.ee/et/juhised/isikuandmetootlemine-toosuhetes>>.
- An Inventory of Community Actions in the Field of Multilingualism – 2011 update. Commission Staff Working Paper. Brussels, July 7, 2011, SEC(2011) 926 final.
- Case C-468/10 *Asociación Nacional de Establecimientos Financieros de Crédito (ASNEF) v Administración del Estado* and Case C-469/10 *Federación de Comercio Electrónico y Marketing Directo (FECEMD) v Administración del Estado of the European Court of Justice*. [2011] I-12181.
- CLARIN. A Report Including Model Licensing Templates and Authorization and Authentication Scheme. EC FP7 project no. 212230, Deliverable: D7S-2.1 Version 5, final 2010.

- De Cock Buning, M. *et al.* *The Legal Status of Research Data in the Knowledge Exchange Partner Countries*. The Netherlands: Centre for Intellectual Property Law (CIER), Molengraaff Institute for Private Law, Utrecht University, 2011.
- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, 23/11/1995, p. 0031–0050.
- Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. OJ L 77, 27.3.1996, p. 20–28.
- 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 (InfoSoc Directive). OJ L 167, 22.6.2001.
- Erdos, D. Freedom of Expression Turned on Its Head? Academic Social Research and Journalism in the European Privacy Framework. *Public Law*. 2013, (1): 52–73.
- Eesti Vabariigi põhiseadus [The Constitution of the Republic of Estonia] (as entering into force on 3.7.1992). RT 1992, 26, 349; RT I, 27.4.2011, 1 [interactive] (in Estonian). [accessed on 26-01-2014]. English translation available via <http://www.legaltext.ee/>.
- Estonian Institute. *Estonian in a World Context*. [interactive]. [accessed on 26-01-2014]. <[http://www.estonica.org/en/Society/The\\_Estonian\\_Language/Estonian\\_in\\_a\\_world\\_context/](http://www.estonica.org/en/Society/The_Estonian_Language/Estonian_in_a_world_context/)>.
- European Convention on Human Rights. [interactive]. [accessed on 16-02-2014]. <[http://www.echr.coe.int/Documents/Convention\\_ENG.pdf](http://www.echr.coe.int/Documents/Convention_ENG.pdf)>. German Federal Data Protection Act. January 14, 2003 [interactive]. [accessed on 26-01-2014]. Translation available at: [http://www.gesetze-im-internet.de/englisch\\_bdsge/englisch\\_bdsge.html](http://www.gesetze-im-internet.de/englisch_bdsge/englisch_bdsge.html).
- Isikuandmete kaitse seadus [Data Protection Act] (as entering into force on 01.01.2008). RT I 2007, 24, 127; RT I, 30.12.2010, 2 (in Estonian). [interactive]. [accessed on 26-01-2014]. Unofficial translation available via <http://www.legaltext.ee/>.
- Isikuandmete kaitse seaduse seletuskiri [Explanatory Memorandum of the Personal Data Protection Act] (in Estonian). [interactive]. [accessed on 26-01-2014]. <<http://www.aki.ee/sites/www.aki.ee/files/elfinder/.../IKS%20SELETUSKIRI.rtf>>.
- Judgment of the Estonian Supreme Court of 3 May 2001 in Case 3-4-1-6-01.
- Kelli, A.; Tavast, A.; Pisuke, H. Copyright and Constitutional Aspects of Digital Language Resources: The Estonian Approach. Tallinn. *Juridica International*. 2012, 19: 40–48.
- Liin, K. *et al.* *The Estonian Language in the Digital Age/Eesti Keel Digiajastul*. White Paper Series/Valge Raamatu Sari, 2012.
- Madise, Ü. *et al.* *Eesti Vabariigi põhiseadus. Kommenteeritud väljaanne. Kolmas, täiendatud väljaanne* (Constitution of the Republic of Estonia. Commented Edition, Third Edition). Tallinn: Juura (in Estonian), 2012 [interactive]. [accessed on 26-01-2014]. <<http://www.pohiseadus.ee/ptk-2/pg-26/>>.

- Opinion 4/2007 on the concept of personal data. Article 29 of Data Protection Working Party. 01248/07/EN WP136. 2007 [interactive]. [accessed on 26-01-2014]. <<http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/>>.
- Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. COM(2012) 11 final.
- Recommendation 1/99 adopted on February 23rd 1999, on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware by the Working Party on the Protection of Individuals with regard to the Processing of Personal Data set up by Directive 95/46/EC of the European Parliament and of the Council of 24 October, 1995, 5093/98/EN/final.
- Rosier, K.; Vereecken, I. *Data Protection Aspects within the Framework of Socio-Economic Research*. Institute for Employment Studies, Report 415, 2004.
- Tavast, A.; Pisuke, H.; Kelli, A. Õiguslikud väljakutsed ja võimalikud lahendused keeleressursside arendamisel (Legal Challenges and Possible Solutions in Developing Language Resources). *Eesti Rakenduslingvistika Ühingu Aastaraamat*. 2013, (9): 317–332.
- Teadus- ja arendustegevuse korralduse seadus [Organisation of Research and Development Act] (as entering into force on 02-05-1997). RT I 1997, 30, 471; RT I, 28.06.2012, 13 (in Estonian). [interactive]. [accessed on 26-01-2014]. Unofficial translation available via <http://www.legaltext.ee/>.

## ASMENS DUOMENŲ TVARKYMO TEISINIAI ASPEKTAI VYSTANT IR NAUDOJANT SKAITMENINIUS KALBOS IŠTEKLIUS: ESTIJOS PERSPEKTYVA

Aleksei Kelli, Liina Jents

Tartu universitetas, Estija

**Anotacija.** Skaitmeniniai kalbos ištekliai yra pridėtinę vertę kuriančių kalbos technologijų pagrindas ir palaiko mažųjų kalbų išlikimą skaitmeniniame amžiuje. Straipsnio autorių dėmesio centre yra asmens duomenų apsauga plėtojant skaitmeninius kalbos išteklius. Teigiama, kad asmens duomenų tvarkymas be duomenų subjekto sutikimo plėtojant skaitmeninius kalbos išteklius gali būti teisėtas, kadangi yra atliekamas pagrindžiant viešuoju interesu.

**Reikšminiai žodžiai:** asmens duomenų apsauga, ypatingi asmens duomenys, skaitmeniniai kalbos ištekliai, kalbos technologijos.

## LEGAL ASPECTS OF PROCESSING PERSONAL DATA IN DEVELOPMENT AND USE OF DIGITAL LANGUAGE RESOURCES: THE ESTONIAN PERSPECTIVE

Aleksei Kelli, Liina Jents

University of Tartu, Estonia

***Summary.** Language research and development of digital language resources creates several legal challenges. The authors' main focus in this article relates to data protection issues in the context of the development of digital language resources. This article complements the research already done on copyright aspects of digital language resources.*

*In the first section of the article the authors analyse technical and legal nature of digital language resources and language technologies. The authors rely on the previous research and define digital language resources as a database containing oral and written texts. Since the choice of the texts constitutes a creative process then this database is protected as a work. Since the creation of language resources requires considerable investment then it can also be protected as sui generis database.*

*Digital language resources can be used for language research and creation of language technologies. The development of language technologies (e.g. spelling checkers, machine translation, speech synthesis and recognition) relies on the existing digital language resources. Language research and language technologies are crucial for the survival of a language in the internet era. The main legal issue here is that the texts used may contain personal data.*

*In the second section of the article the authors concentrate on the processing of personal data for scientific research and other purposes. The processing of the personal data can be based on the consent of data subject or exceptions allowed by law.*

*In the third section of the article the authors focus on specific aspects of processing personal data in context of digital language resources. The authors address the specific issues relating to oral speech and then focus on problems common to both of them.*

*The authors conclude that language research offers tremendous opportunities to address multilingualism issues in Europa. It is also a measure to fight against the extinction of a language. An adequate regulatory framework plays a key role in enhancement of language research.*

---

**Aleksei Kelli**, socialinių mokslų (teisė) daktaras, Tartu universiteto Teisės fakulteto docentas, Intelektinės nuosavybės teisės kodifikavimo ekspertų grupės vadovas (Estijos Respublikos teisingumo ministerija). Mokslinių tyrimų kryptys: intelektinės nuosavybės teisė, civilinė teisė.

**Aleksei Kelli**, University of Tartu, Faculty of Law, Associate Professor of Intellectual Property Law, the Head of an Expert Group on the Codification of the Intellectual Property Law (the Ministry of Justice of Estonia). Research interests: intellectual property law, civil law.

---

**Liina Jents**, Tartu universiteto Teisės fakulteto doktorantė, advokatė, Intelektinės nuosavybės teisės kodifikavimo ekspertų grupės narė (Estijos Respublikos teisingumo ministerija). Mokslinių tyrimų kryptys: intelektinės nuosavybės teisė, duomenų apsaugos teisė.

**Liina Jents**, University of Tartu, Faculty of Law, PhD student; Attorneys at Law Borenius, Attorney at Law; Member of the Expert Group on the Codification of the Intellectual Property Law (the Ministry of Justice of Estonia). Research interests: intellectual property law, data protection law.