

# A CLARIN Contractual Framework for Sharing Personal Data for Scientific Research

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

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

**Alexandros Nousias**  
Nousias/Linardos Business  
& Legal Consultants  
alexandros.nousias@  
gmail.com

## Abstract

The development and use of language resources often involve the processing of personal data. Processing has to have a legal ground. The General Data Protection Regulation (GDPR) provides several legal grounds. In the context of scientific research, *Consent*, *Public interest* and *Legitimate interest* are relevant. The main question is when researchers should rely on Consent and when on Public or Legitimate interest to conduct research. All three grounds have their advantages and challenges. For comparing Consent and Public interest, the Clinical Trial Regulation is used as an example. In addition, we study how this has been implemented in Finland based on the guidelines for research data from the Data Protection Ombudsman and suggest an update of the CLARIN Deposition License Agreement templates to accommodate data sets with personal data.

## 1 Introduction

Language resources (LRs) contain material subject to various legal regimes. For instance, they may contain copyright protected works, objects of related rights (performances) and personal data. This affects the way language resources are collected and used. Intellectual property issues relating to language resources have previously been addressed by Kelli et al. (2016). The general approach to dealing with personal data<sup>1</sup> in research is outlined by Kelli et al. (2019), where they discuss how processing<sup>2</sup> personal data without consent (see also Klavan et al. 2018) as the legal basis is possible in various EU countries.

Personal data issues are relevant for language resources, given that they potentially contain oral speech or written text, which relate to a natural person.<sup>3</sup> According to the Charter of Fundamental Rights of the European Union (Charter 2012) “*Everyone has the right to the protection of personal data concerning him or her*” (Art. 8 (1)). The general framework for personal data protection is provided in the General Data Protection Regulation<sup>4</sup> (GDPR). This paper primarily outlines the regulatory framework for processing personal data for research purposes. Special attention is given to the legal grounds for processing for research purposes.

---

<sup>1</sup> The General Data Protection Regulation defines personal data as “*any information relating to an identified or identifiable natural person ('data subject')*” (Art. 4 (1)).

<sup>2</sup> Processing is defined 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)).

<sup>3</sup> It should be emphasised that in case data is anonymised then data protection laws do not apply (see recital 26 of GDPR). For further discussion on anonymisation techniques, see WP29 2014.

<sup>4</sup> The GDPR is applicable in all EU Member States from 25 May 2018. It replaces the (Data Protection Directive).

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

The Charter of Fundamental Rights of the European Union foresees that personal data “*must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law*” (Art. 8 (2)). In addition to consent, the GDPR provides other legal grounds for processing personal data as well. We focus on two legal grounds relevant for processing personal data for research purposes, i.e. consent and the use of data for public interest research while also alluding to some conditions for using legitimate interest.

The choice of legal ground, however, does not affect the other obligations of the data controller<sup>5</sup> under the GDPR. For instance, when using data for research purposes in the public interest, we also need to consider whether and how it is feasible to inform the data subjects.

If we get data directly from the data subjects, the GDPR requires that the data subjects be informed (GDPR, Art. 13), i.e. we need to document that the data subjects have been aware of our activity. If on the other hand, we reuse personal data from large databases or publicly available sources, it is not always possible to inform the data subjects in person.

The focus of this study is to show how one can collect personal data directly from the data subjects while still using research in the public interest as a legal ground. We call this model confirmation to participate in research of public interest, and show that this is not just a hypothetical model but it is already in use in the Clinical Trial Regulation (CTR 2014). Framed in the terminology of the GDPR, the CTR model is based on processing data for a task carried out in the public interest. The data subject nevertheless has to be informed about the processing and must consent to participating in the trial, i.e., the data subject gives informed consent to participate in the research and may end participation at any time. However, according to the CTR, the data subject does not give specific consent to process personal data in the way in which consent is defined in the GDPR. The CTR consent has therefore also been labeled “broad consent”<sup>6</sup>, “ethical consent” or “consent to participate”, which has important consequences for the right of the data subject to limit processing of the data (e.g., the right to withdraw consent to process the data at any time).

Other types of research than clinical trials can also be carried out in the public interest using the same model for processing personal data, i.e., by *confirmation to participate in research carried out in the public interest*. Since key concepts of the data protection framework (personal data, data subject, etc.) are addressed in previous CLARIN publications (Kelli et al. 2016, 2019), they are not repeated here.

We first explore the consequences of GDPR enabling research in the public interest as a legal ground for doing scientific research vs. using GDPR consent as the legal ground in Sections 2-4. We then proceed to outline the documents that are needed to implement both of them in practice in Sections 5-6.

## 2 Processing personal data for research purposes according to the GDPR

Processing of personal data for research purposes can be illustrated with the graph in Fig. 1. The GDPR defines research broadly so that it covers “*technological development and demonstration, fundamental research, applied research and privately funded research*” (Recital 159). The GDPR provides the following requirements for processing data for research purposes (Art. 89):

- 1) processing for research purposes is subject to appropriate safeguards. The safeguards ensure that technical and organisational measures are in place in particular to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner;
- 2) the Member States may limit the following data subject’s rights for research purposes (optional limitations):
  - a) the right of access by the data subject (Art. 15);
  - b) the right to rectification (Art. 16);

---

<sup>5</sup> According to the GDPR the controller is defined as “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data” (Art. 4 (7)).

<sup>6</sup> Chassang (2017) discusses the difference between the “broad consent” in CTR and the specific consent in GDPR (as well as in the preceding directive from 1994) where he regards CTR as *lex specialis*. However, CTR predates GDPR, which now accommodates the CTR-style consent under research in the public interest allowing us to apply it also to other areas of research with public interest as the legal ground.

- c) the right to the restriction of processing (Art. 18);
- d) the right to object (Art. 21);

The implementation of the optional limitations varies by country and they are exemplified and discussed in Kelli et al. (2019). However, there are two mandatory limitations<sup>7</sup> to the rights of data subjects with regard to research data: 1) the right to be forgotten<sup>8</sup> and 2) the right to be informed about the processing.<sup>9</sup>

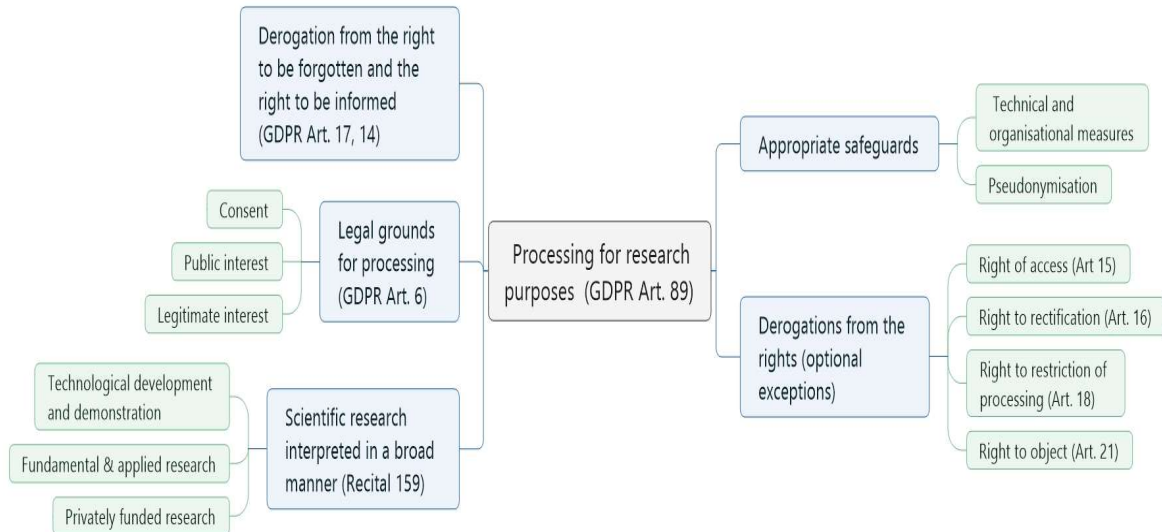


Figure 1. Processing personal data for research

When processing data for scientific research “*further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes*” (GDPR, Art. 5b). The European Data Protection Supervisor gave an opinion on how to interpret this in January 2020 (EDPS 2020): “*The presumption is not a general authorisation to further process data in all cases for historical, statistical or scientific purposes. Each case must be considered on its own merits and circumstances. But in principle personal data collected in the commercial or healthcare context, for example, may be further used for scientific research purposes, by the original or a new controller, if appropriate safeguards are in place.*”

However, there is an additional aspect, which needs to be kept in mind before we focus on the separate legal grounds for processing personal data. Namely, it is essential that the legal ground for processing is chosen before the processing starts and its subsequent replacement could be problematic. According to the guidelines on consent “*the controller cannot swap from consent to other lawful bases. For example, it is not allowed retrospectively to utilise the legitimate interest basis in order to justify processing, where problems have been encountered with the validity of consent. Because of the requirement to disclose the lawful basis which the controller is relying upon at the time of collection of personal data, controllers must have decided in advance of collection what the applicable lawful basis is*” (WP29 2018: 23). It therefore makes sense to use a legal ground, which is as amenable to scientific research as possible when collecting new data in order to avoid unnecessary difficulties for further processing.

The GDPR provides six legal grounds for processing personal data: 1) consent; 2) performance of a contract; 3) compliance with a legal obligation; 4) protection of the vital interests; 5) the public interest or in the exercise of official authority; 6) legitimate interest (Art. 6). The processing for research purposes is not an individual legal ground. The processing for research purposes can rely on consent, the

<sup>7</sup> Mandatory limitations are directly applicable. They do not need to be incorporated into the national laws.

<sup>8</sup> GDPR Art. 17 (3) d.

<sup>9</sup> GDPR Art. 14 (5) b.

performance of a task carried out in the public interest, or legitimate interest<sup>10</sup>. For comparison, we briefly outline the three legal grounds that can be used in scientific research: consent, public interest and legitimate interest.

## 2.1 Consent

The processing based on the data subject's consent (Art. 6 (1) a) offers the highest possible protection for the data subject through the following mechanism:

- 1) the consent has to be freely given, specific, informed and unambiguous (Art. 4 (11));
- 2) the data subject can withdraw the consent without any detriment (Art. 7 (3));
- 3) the burden of proof lies with the controller (Art. 7 (1))

The Article 29 Working Party<sup>11</sup> (WP29) explains that consent “*focuses on the self-determination of the data subject as a ground for legitimacy. All other grounds, in contrast, allow processing – subject to safeguards and measures – in situations where, irrespective of consent, it is appropriate and necessary to process the data within a certain context in pursuit of a specific legitimate interest*” (2014a: 13). In case the acquisition of consent is complicated or administratively burdensome (e.g., anonymous web posts, legacy resources, public videos and so forth), some other legal ground than consent is clearly needed.

## 2.2 Public interest

The GDPR names the performance of a task carried out in the public interest as a legal bases for processing personal data (Art. 6 (1) e)). According to WP29, the performance of a task carried out in the public interest is also a ground for processing personal data in the research context (2014a: 22).

The concept of research in the public interest can usually be invoked by research projects affiliated with universities or research institutions having a legal mandate to do research in the public interest, but it also allows for companies acting in the public interest on behalf of a Member State, e.g., to ascertain the safety and/or efficacy of a procedure performed in addition to normal clinical practice as outlined in the Clinical Trial Regulation (CTR 2014).

Since language data often contains both personal data and copyrighted works, within the framework of development of language technologies it is also relevant to take into consideration the concept of research as defined in the field of copyright law. The Digital Copyright Directive (DCD) provides a special regulation on the text and data mining (TDM) for the purposes of scientific research (Art. 3). According to the Digital Copyright Directive the concept of scientific research covers “*natural sciences and the human sciences*” (Recital 12). The TDM exception allows public-private partnerships (Recital 11). Research organisations are the main beneficiary of the TDM exception (DCD Art. 3). The Digital Copyright Directive defines them as follows (Art. 2 (1)): “*a ‘research organisation’ means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research:*

- (a) *on a not-for-profit basis or by reinvesting all the profits in its scientific research; or*
- (b) *pursuant to a public interest mission recognised by a Member State;*

*in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation”.*

After the harmonisation of the Digital Copyright Directive with national laws, the definition most likely becomes relevant for personal data processing as well since research organisations as such are defined generally and not only for copyright purposes.<sup>12</sup>

---

<sup>10</sup> Note that legitimate interest as a legal ground for research needs to be argued when one cannot claim to be acting with permission or in the interest of the data subject (legal grounds 1, 2 or 4 in Art. 6) or in the interest of the state (legal grounds 3 or 5 in Art. 6).

<sup>11</sup> The Article 29 Working Party (Art. 29 WP) is the independent European working party that dealt with issues relating to the protection of privacy and personal data until 25 May 2018, at which point it was succeeded by the European Data Protection Board.

<sup>12</sup> During the discussion concerning the implementation of the Digital Copyright Directive in the Ministry of Justice of Estonia (23-24 September 2019), it was indicated that the Estonian Organisation of Research and Develop-

## 2.3 Legitimate interest

Pursuant to the GDPR, the processing is lawful if it is “*necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data*” (Art. 6 (1) clause f). The guidelines provide several contexts where the legitimate interest could potentially serve as a legal ground for processing personal data. Such contexts are, e.g. “*conventional direct marketing and other forms of marketing or advertisement; unsolicited non-commercial messages, including for political campaigns or charitable fundraising; prevention of fraud, misuse of services, or money laundering; employee monitoring for safety or management purposes; whistle-blowing schemes; physical security, IT and network security, etc.*” (WP29 2014a: 25). As can be inferred from the examples, the legitimate interest covers purposes (including research purposes) that are typical for organisations (including commercial organisations) for which the interests of the organisation need to be explicitly argued to override the interests of the data subject. Fig. 2 illustrates the context for using legitimate interest.

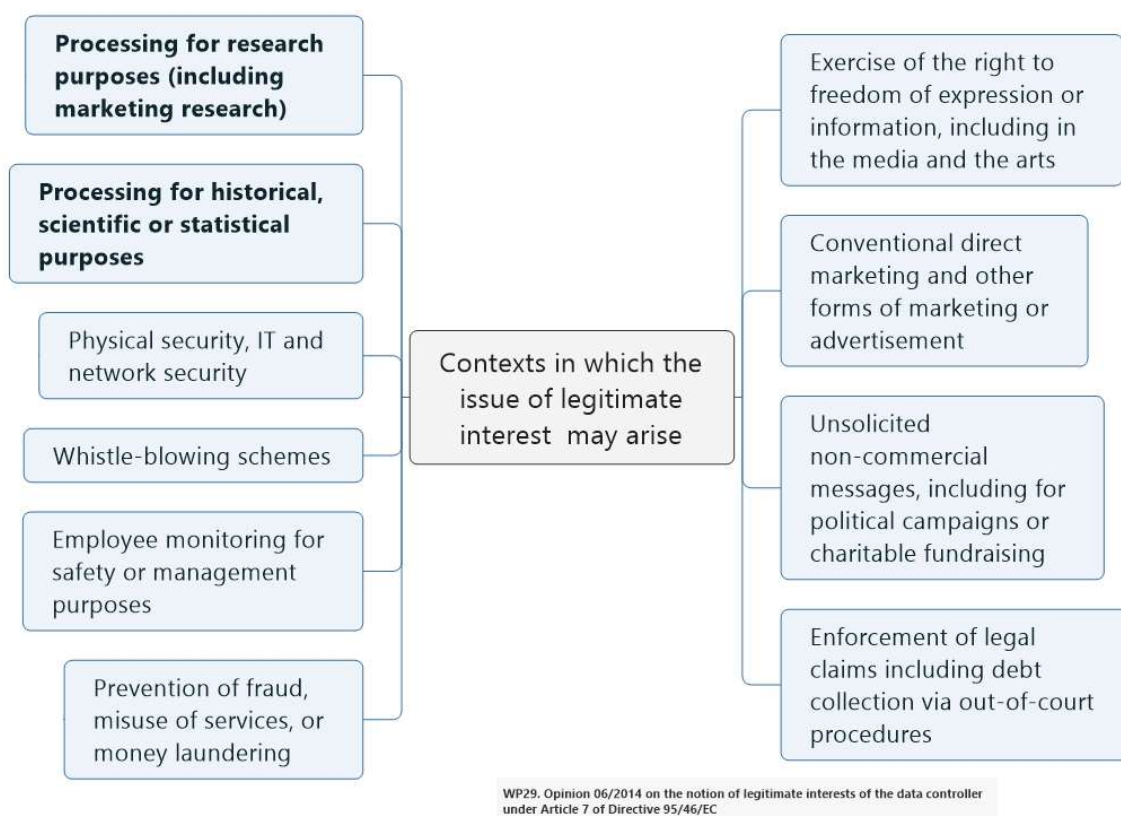


Figure 2. Legitimate interest contexts

## 3 Data Processing according to the Clinical Trial Regulation

The Clinical Trial Regulation (CTR) entered into force on 16 June 2014, but the timing of its application depends on the development of a fully functional EU clinical trials portal and database, which will be confirmed by an independent audit. The Regulation becomes applicable six months after the European Commission publishes a notice of this confirmation, which is likely to occur in 2020. The GDPR refers to CTR for special requirements for the consent to participate in scientific research activities in

---

ment Act is analysed and amended (if needed) to make it compatible with the Digital Copyright Directive. According to the Estonian Organisation of Research and Development Act (ORDA) a research and development institution is a legal person or an institution in the case of which the principal activity is carrying out basic research, applied research or development, or several of the aforementioned activities (§ 3 (1) clause 1).



clinical trials (Recital 161). The informed consent in CTR is often confused with GDPR Consent. In this paper, we try to highlight the difference between *GDPR Consent*, which is a legal ground, and *informed consent*, which is a protective measure when the legal ground is *GDPR Public interest*.

In CTR (Article 2(2)21), '*informed consent*' means a subject's free and voluntary expression of his or her willingness to participate in a particular clinical trial, after having been informed of all aspects of the clinical trial that are relevant to the subject's decision to participate or, in case of minors and of incapacitated subjects, an authorisation or agreement from their legally designated representative to include them in the clinical trial. This informed consent constitutes a confirmation to participate in a clinical trial. Special requirements for the informed consent are provided in Chapter V of the CTR.

About withdrawal of the informed consent, CTR says in its Recital (76) ..., *while safeguarding the robustness and reliability of data from clinical trials used for scientific purposes and the safety of subjects participating in clinical trials, it is appropriate to provide that, without prejudice to Directive 95/46/EC [now replaced by GDPR], the withdrawal of informed consent should not affect the results of activities already carried out, such as the storage and use of data obtained on the basis of informed consent before withdrawal*. This means that the withdrawal of the informed consent only implies that the data subject stops participating in the trial. Data collected during the participation can still be stored, e.g. for verifying the results of activities already carried out.<sup>13</sup>

The legal basis for a clinical trial is research on behalf of a Member State to ascertain the safety and efficacy of a procedure performed in addition to normal clinical practice, which is a prototypical case for research in the public interest. However, for clinical trials, the approval of an ethics committee is also needed, because a non-standard clinical procedure will be applied potentially affecting the well-being of the data subjects.

In GDPR terminology, the CTR policy can be restated as the data subject's informed consent to participate in research in the public interest. The requirement to provide information about the research is a protective measure so that the informant can decide whether he consents to participating in the research. He does not explicitly consent to any specific use of the provided data as the legal ground for collecting the data is the public interest in the potential outcome of the research, which has consequences for withdrawing provided data.<sup>14</sup> For further details on the interaction between GDPR and CTR when reusing data for scientific research, see Pormeister (2020).

#### 4 Discussion of consequences for other types of research

Other types of research than clinical trials can also be carried out in the public interest by research projects affiliated with institutions having a legal mandate to do research in the public interest. They can use the same legal ground for processing personal data. If research is carried out in the public interest and data is reused from large databases or public sources, it is not always possible to inform all data subjects about the processing, in which case a public record of processing activities may be deemed sufficient.

However, if data is collected directly from the data subjects, and the legal basis according to GDPR is *research in the public interest*, it may be useful to avoid confusion among data subjects and researchers alike by naming the verification for having provided mandatory information "*confirmation to participate in research carried out in the public interest*". The confirmation relies on informed consent to participate, so the consent is given only with regard to the participation and not with regard to the processing of the personal data, for which the legal basis is research in the public interest.

Since research is conducted in the public interest, the data subject's right to be forgotten is mandatorily limited by (GDPR Art. 17 (3)d) to the extent that processing is necessary for research purposes in so far as the right to be forgotten is likely to render impossible or seriously impair the achievement of the objectives of that processing. In many cases, the right to be forgotten could impair the replicability of the research.

---

<sup>13</sup> The GDPR has a similar approach. It provides that "*The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal*" (Art. 7 (3)). However, GDPR also provides for storage extension for research (Art. 5) and a research purpose extension (Art. 5).

<sup>14</sup> As data is collected for research purposes, there is also a limited right to be forgotten (Art. 17).

It should be noted that data collected in this way can only be used for research purposes. According to the GDPR<sup>15</sup>, such data can still be reused for other research purposes (GDPR Art. 5), but the transfer of the data to another research organisation needs to be protected to make sure that the data is processed only for research purposes<sup>16</sup>. If one wishes to make such personal data publicly available, e.g. as an illustrating example or a video clip potentially identifying the data subject on the internet or at a conference, *GDPR Consent* must be acquired from the data subject.

The legal framework for processing personal data for research purposes is based on the GDPR and national laws of the EU Member States. This means that in addition to the mandatory limitations of the rights of data subjects enforced by the GDPR, researchers that wish to develop language resources containing personal data may have further rights to maintain the research data integrity through nationally implemented limitations to the rights of data subjects.

## 5 Finland as a case study for research in the public interest as a legal ground

In Finland, the new Data Protection Act (FINLEX 1050/2018) entered into force on January 1, 2019. Collecting personal data to be used for scientific research purposes is currently shifting towards using research in the public interest as the predominant legal ground, while the preferred way to obtain personal data under the previous legislation was consent to use data. This is due to the fact that Finland was among the countries that took nearly full advantage of the leeway allowed in the GDPR to enact derogations to the rights of the data subjects when using data for scientific research in the public interest. As argued in the previous Sections of this paper, this is not something radically new introduced by the GDPR, it is merely extending a practice that has already been established in a clinical research setting to other domains of scientific research. To distinguish scientific research from other kinds of research, the Supreme Administrative Court (SAC) of Finland has established (FINLEX KHO:2013:181) that providing confidential information for scientific research requires: 1) an adequate research plan, 2) sufficient scientific qualifications of the participating researchers, 3) autonomous and public agency of the researchers and 4) scientific research objectives.

The Office of the Data Protection Ombudsman (DPO) of Finland published guidelines on January 30, 2020, for how to use personal data in research “The path to data protection in research” (Data Protection Ombudsman 2020), which for practical purposes are more detailed with regard to consent without prejudice to other available options such as scientific research in the public interest. The DPO points out that the GDPR takes a broad perspective on scientific research, i.e. scientific research purposes may include development and presentation of technology, basic research, applied research or privately funded research, but this does not extend the definition of scientific research beyond what has been prescribed by the SAC. In particular, the DPO emphasizes that scientific and historical research as well as statistical purposes entail increasing public knowledge in society, e.g. combining information from several data sources may provide valuable new information for disease prevention or treatment, which may serve as a basis for new policies improving the well-being of citizens and improve the efficacy of welfare services.

The guidelines recognize that the GDPR aims to facilitate access to data for scientific and historical research. As mentioned above, according to the GDPR all processing of personal data for scientific and historical research purposes is compatible with the original purpose if applying adequate protective and organisational measures, e.g. using data encryption with authorized access. However, further processing for scientific research purposes still requires taking all the other aspects of the data protection legislation into account, e.g. the data subjects need to be informed before the processing starts unless there are compelling and legitimate reasons. As it is not always possible to reach the data subjects personally, some other sufficiently public means may also be used to inform them about the further processing of the data. The DPO points out that a research purpose that is incompatible with the original purpose, for which the data was originally collected, is still possible but requires that the data subject be informed,

---

<sup>15</sup> GDPR is applicable only within the EU, but many countries have agreements with the EU, see e.g. <https://gdpr-info.eu/issues/third-countries/>

<sup>16</sup> The reuse still needs to answer to general GDPR requirements such as the data minimisation principle. According to Article 89(1) and Recital 156, for further processing, the controller should also assess the feasibility to fulfil the reuse purposes by processing anonymous or pseudonymous data.

and if the legal ground for collecting data was GDPR Consent, a new consent is needed for such incompatible further processing.<sup>17</sup> The fact that personal data collections can be reused if fulfilling the various criteria imposed by the GDPR is fundamental for a research infrastructure like CLARIN.

For a research infrastructure, the conditions on which the data can be communicated to others are key issues: the foremost among them are the purpose of use and the storage conditions of the personal data set. With regard to the definition of the original research purpose, the DPO guidelines are vague saying only that sweeping definitions such as “future research” or “general research purposes” are too broad and therefore inadequate. This still leaves us with more specific options, e.g. “scientific research for the purposes of humanities and social sciences” or “scientific research on language-aware technology” limiting the use to particular fields of scientific research. This may be problematic if we foresee multi-disciplinary research on the data. On the other hand, specifying one or more research questions such as “scientific research for enhancing the well-being of citizens” limits future research topics but allows multi-disciplinary research for such purposes.

Data should be stored only to the extent needed, i.e. neither the type of data nor the time-span should be excessive. However, it is usually necessary to keep data for verification of the research results and sometimes more data will be collected in longitudinal studies, in which case the original data needs to be kept for comparison. If an end-point is difficult to define, one may instead set up criteria for re-evaluating the need to keep the data at regular intervals. If the longitudinal study spans several decades, the original data subjects will eventually pass away, at which time the data is no longer personal.

## 6 CLARIN Deposition License Agreements

To facilitate the adoption of the requirements of the GDPR, the CLARIN License Agreements need to be updated with provisions for communicating data sets containing personal data. In general, this was the intention for the category of RES license agreements (for further discussion on categories, see Oksanen et al. 2010; Kelli et al. 2018), but very limited amounts of personal data were available for transfer and reuse at the time when the template was designed. With the GDPR, the definition of personal data has been considerably expanded and many more data sets may contain personal data so specific provisions for communicating such data needs to be adopted.

In addition, the GDPR assumes that an institution, which potentially has an appointed data protection officer, takes responsibility for the personal data set and that a record of processing activities is established by the organisation. An individual researcher may not have the authority to bind his organisation contractually to such activities, so for personal data sets the end-user is not primarily an individual researcher but a research organisation.

With an increased number of data sets identified as containing personal data, there is a rather motely practice for personal data set transfer within the EU. It may therefore be useful to agree on similar rules for personal data exchange at least within a national CLARIN consortium and further between organisations in the various CLARIN member states.

From a CLARIN perspective, the proposed agreement structure aims to establish a CLARIN B or C Centre as a Data Processor for the national CLARIN consortium with each of the consortium members, or some external party, as a Controller of its own personal data sets. To this end, we propose modifications that follow the suggestions proposed already by Kelli & al. (2016). The main agreement is renamed as the CLARIN Framework Deposition Agreement (FADA) with two appendices:

- 1) the Data Protection Agreement (DAPA), and
- 2) the Deposition License Agreements (DELA).

The CLARIN FADA establishes a framework of common deposition rules for data sets that can be communicated by a CLARIN Centre.

Individual data sets are added as attachments to the CLARIN FADA keeping only data set specific information in the DELA, which thereby reduces to a 1-page main document for each data set referring to the general conditions in the FADA and to four data set specific appendixes:

- 1) the data identification, description and citation texts,
- 2) the deposition license conditions with an end-user license agreement template,

---

<sup>17</sup> Note that further processing to verify previous research results is therefore allowed without renewal of consent, and such research may be varied and extensive as long as the final purpose is to verify previous results.



- 3) a list of third party copyrights or database rights, and
- 4) the personal data description and the purpose of use of the data set.

Appendices 3 or 4 may explicitly be left empty if there are no third party rights or no personal data in the data set.

In CLARIN, the RES licensing scheme is suitable for communicating copies of data sets with personal data. In the suggestions for how to implement the ethical intent of the GDPR in a research setting, Pormeister (2020) recommends that the original controller stays informed about all further use of a personal data set in order to inform the data subjects about such further use when necessary. The CLARIN RES license requires that data sets not be communicated to a third party by the end-user, because new end-users can obtain a copy directly from CLARIN. As CLARIN remains a mere processor of personal data for the purpose of communicating such data to research organisations, the original controller stays informed about all requests for further use of a data set. If there is a request for using a data set for a research purpose which is not sufficiently compatible with the original purpose of use, the data subjects need to be informed. From a CLARIN perspective, it is mostly a practical question whether the CLARIN Centre as a processor is commissioned to inform the data subjects or the original controller informs them, and how one goes about informing them, i.e. is personal communication possible or is a public announcement sufficient.

## 7 Conclusion

The development and use of language resources often involves the processing of personal data. To process data for research purposes according to the GDPR, it is possible to invoke research in the public interest as the legal basis for publicly funded research projects carried out at research institutions with a legal mandate to do research in the public interest. If data is collected directly from data subjects, they have to be informed about the data processing so that they can opt-in by confirming their willingness to participate. If they no longer wish to participate, they have a right to stop the processing of their data, but they do not automatically have a right to be forgotten and their personal data may be reused for other research purposes. This model for using and reusing personal data is already established in the CTR, but it can be extended to other domains of scientific research as well through the provisions of the GDPR.

In the paper, we discuss how to implement this in practice through a CLARIN Framework Deposition Agreement containing the general provisions of the CLARIN Deposition License Agreement, with an appendix containing general data protection conditions and another appendix containing data set specific information.

## References

- [Charter 2012] 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:12012P/TXT> (6.2.2020).
- [Chassang 2017] Gauthier Chassang (2017). The impact of the EU general data protection regulation on scientific research. *Ecancelmedicalscience*, 11, 709. doi:10.3332/ecancer.2017.709. Available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5243137/> (6.2.2020).
- [CTR 2014] Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use. Available at <https://ec.europa.eu/health/human-use/clinical-trials/regulation> (31.1.2020).
- [Data Protection Directive] 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. Available at <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046&qid=1522340616101&from=EN> (29.3.2020).
- [Data Protection Ombudsman 2020] Guidelines on Scientific Research published in Finnish on January 30, 2020 [Tieteellinen tutkimus] Available in Finnish at: <https://tietosuoja.fi/tieteellinen-tutkimus> (31.1.2020).
- [Digital Copyright Directive = DCD] 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> (26.1.2020).

- [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 <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1515793631105&uri=CELEX:32016R0679> (5.2.2020).
- [EDPS 2020] A Preliminary Opinion on data protection and scientific research by the European Data Protection Supervisor. 6.1.2020. Available at: [https://edps.europa.eu/sites/edp/files/publication/20-01-06\\_opinion\\_research\\_en.pdf](https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf) (31.1.2020).
- [FINLEX KHO:2013:181] Decisions by the Supreme Administrative Court in Finland concerning “Asiakirjajulkisuus - Tutkimuslupa - Tieteellinen tutkimus - Tutkimuksen tieteellisyys - Reseptitiedosto - Lääkeyritys - Kansaneläkelaitos” [Public access to documents – Research permit – Scientific Research – Conditions for scientific research – Medical prescription data files – Medical company – the Social Insurance Institution of Finland]. Available in Finnish at: <https://www.finlex.fi/fi/oikeus/kho/vuosikirjat/2013/201303651> (31.1.2020).
- [FINLEX 1050/2018] Data Protection Act enacted by the Parliament of Finland (English translation). Available at: <https://www.finlex.fi/en/laki/kaannokset/2018/en20181050> (31.1.2020).
- [Kelli et al. 2019] Aleksei Kelli, Krister Lindén, Kadri Vider, Pawel Kamocki, Ramunas Birštonas, Silvia Calamai, Penny Labropoulou, Maria Gavrilidou, Pavel Straňák (2019). Processing personal data without the consent of the data subject for the development and use of language resources. In: Inguna Skadina, Maria Eskevich (Ed.). Selected papers from the CLARIN Annual Conference 2018. Linköping University Electronic Press, 72-82. Available at <http://www.ep.liu.se/ecp/159/008/ecp18159008.pdf> (29.1.2020).
- [Kelli et al. 2018] Aleksei Kelli, Krister Lindén, Kadri Vider, Penny Labropoulou, Erik Ketzan, Pawel Kamocki, Pavel Straňák (2018). Implementation of an Open Science Policy in the context of management of CLARIN language resources: a need for changes? In: Maciej Piasecki (Ed.). Selected papers from the CLARIN Annual Conference 2017. Linköping University Electronic Press, 102-111. Available at <http://www.ep.liu.se/ecp/147/009/ecp17147009.pdf> (29.1.2020).
- [Kelli et al. 2016] Aleksei Kelli, Kadri Vider, Krister Lindén (2016). The Regulatory and Contractual Framework as an Integral Part of the CLARIN Infrastructure. In: Koenraad De Smedt (Ed.). Selected Papers from the CLARIN Annual Conference 2015. Linköping University Electronic Press, 13-24. Available at <http://www.ep.liu.se/ecp/article.asp?issue=123&article=002> (29.1.2020).
- [Klavan et al. 2018] Jane Klavan, Arvi Tavast, Aleksei Kelli (2018). The Legal Aspects of Using Data from Linguistic Experiments for Creating Language Resources. *Frontiers in Artificial Intelligence and Applications*, 307, 71-78. Available at <http://ebooks.iospress.nl/volumearicle/50306> (29.1.2020).
- [Oksanen et al. 2010] Ville Oksanen, Krister Lindén, Hanna Westerlund (2010). Laundry Symbols and License Management: Practical Considerations for the Distribution of LRs based on experiences from CLARIN' in Proceedings of LREC 2010: Workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management. Available at <https://helda.helsinki.fi/handle/10138/29359> (5.2.2020). [ORDA] Organisation of Research and Development Act. Entry into force 2.05.1997. English translation available at <https://www.riigiteataja.ee/en/eli/513042015012/consolide> (21.1.2019)
- [Pormeister 2020] Kärt Pormeister (2020). Transparency in relation to the data subject in genetic research – an analysis on the example of Estonia. Doctoral dissertation 76, School of Law, University of Tartu. p. 189. January 13, 2020. Available at: <https://dspace.ut.ee/handle/10062/66697> (31.1.2020)
- [VLO] CLARIN Virtual Language Observatory. Available at <https://vlo.clarin.eu/> (6.2.2020).
- [WP29 2018] Article 29 Working Party (WP29). Guidelines on consent under Regulation 2016/679. Adopted on 28 November 2017. As last Revised and Adopted on 10 April 2018. Available at [https://ec.europa.eu/newsroom/article29/item-detail.cfm?item\\_id=623051](https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051) (3.2.2020).
- [WP29 2014] Article 29 Working Party (WP29). Opinion 05/2014 on Anonymisation Techniques. Available at [https://iapp.org/media/pdf/resource\\_center/wp216\\_Anonymisation-Techniques\\_04-2014.pdf](https://iapp.org/media/pdf/resource_center/wp216_Anonymisation-Techniques_04-2014.pdf) (3.2.2020).
- [WP29 2014a] Article 29 Working Party (WP29). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. Available at [https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf) (3.2.2020).
- [WP29 2007] Article 29 Working Party (WP29). Opinion 4/2007 on the concept of personal data. Adopted on 20th June. Available at [http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136\\_en.pdf](http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf) (6.2.2020).