Published in: Navarretta, Costanza/Eskevich, Maria (Eds.): Proceedings of CLARIN Annual Conference 2020.
05 – 07 October 2020, Online Edition. - Utrecht: CLARIN, 2020. Pp. 171-177.
https://office.clarin.eu/v/CE-2020-1738-CLARIN2020_ConferenceProceedings.pdf

CLARIN contractual framework for sharing language data: the perspective of personal data protection

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

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

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

Pawel Kamocki
IDS Mannheim, Germany
pawel.kamocki@gmail.com

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

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

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

Mari Keskküla
University of Tartu, Estonia
mari.keskkula@gmail.com

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

Abstract

The article analyses the responsibility for ensuring compliance with the General Data Protection Regulation (GDPR) in research settings. As a general rule, organisations are considered the data controller (responsible party for the GDPR compliance). Research constitutes a unique setting influenced by academic freedom. This raises the question of whether academics could be considered the controller as well. However, there are some court cases and policy documents on this issue. It is not settled yet.
The analysis serves a preliminary analytical background for redesigning CLARIN contractual framework for sharing data.

1 Introduction

This paper focuses on sharing[1] language data (LD) containing personal data (PD).[2] A key issue here is to define obliged parties under the General Data Protection Regulation (GDPR). The GDPR identifies

[1] The General Data Protection Regulation (GDPR) Art. 4 (2) defines processing extensively so that it covers any operation which is performed on personal data (e.g., collection, storage, alteration and sharing).
[2] The article does not address transfer of PD outside the EU. This topic concerns special provisions of the GDPR and EU case law (e.g. C‑311/18).

the controller and the processor as responsible parties (Art. 4).[3] According to the accountability principle, the controller is responsible for the GDPR compliance (Art. 5 (2)). The GDPR defines the controller 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)). The processor is defined as "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*".[4]

The authors place this question into the context of language research and address practical questions such as 1) how to differentiate between obligations of the university and individual researchers (e.g., whether an individual researcher is the controller) and; 2) whether sharing LD results in joint controllership or separate controllership (whether the transferee of the data becomes the controller, the joint controller or the processor). The paper serves as a preliminary conceptual analysis behind redrafting CLARIN contractual framework for data sharing. It develops further the previous research (see Kelli et al. 2015; Kelli et al. 2018) focusing on personal data.

2 Duty-bearers for the GDPR compliance within research settings

The key feature of the controller is the **determination** of the **purposes** and **means** of the processing of PD.[5]

A relevant issue for the research community is to analyse whether an individual researcher and/or the university is the controller. Research settings is a unique context since academic freedom is defined as a fundamental right.[6] This gives researchers often more freedom compared with other employees and complicates matters as seen from the analysis.

Considering relations between companies and their employees WP29 (2010: 15) indicates that "preference should be given to consider as a controller the company or body as such rather than a specific person within the company or the body. It is the company or the body which shall be considered ultimately responsible for data processing and the obligations stemming from data protection legislation". According to the European Commission (EC) "if your company/organisation decides 'why' and 'how' the personal data should be processed, it is the data controller. Employees processing personal data within your organisation do so to fulfil your tasks as data controller". DP Handbook (2018: 102) similarly asserts that legal entity (not employees) is the controller.

WP29 (2010: 6) further explains that "*Controller and processor and their staff are therefore considered as the 'inner circle of data processing' and are not covered by special provisions on third parties*"". This is compatible with employment law. The employment law literature and case law also set forth that one of the characteristics of an employee is the fact that the employee is merged with the employer's team of other employees. The employee is employed and acts within the framework of the employer's economic activities (Risak & Dullinger 2018; C-22/98 para 26; C-66/85; C-256/01).

Although individual researchers conduct research with some freedom, their work is coordinated by the university. The university is responsible for the GDPR violations[7], and it has to implement appropriate technical and organisational measures (incl. data protection policies) to ensure the protection of PD (GDPR Art. 24) and to maintain a record of processing activities (Art. 30 GDPR).[8] For instance, to

[3] When it comes to liability, then as a general rule, any person who has suffered damage as a result of the GDPR violation has the right to receive compensation from the controller or processor (GDPR Art. 82 (1)).

[4] WP29 (2010: 1) explains that the processor has to meet two conditions: 1) a separate legal entity with respect to the controller; 2) it processes PD on behalf of the controller.

[5] According to WP29 "*Determination of the "means" includes both technical and organizational questions where the decision can be well delegated to processors (as e.g. "which hardware or software shall be used?") and essential elements which are traditionally and inherently reserved to the determination of the controller, such as "which data shall be processed?", "for how long shall they be processed?", "who shall have access to them?", and so on'*"" (2010: 14).

[6] According to the EU Charter of Fundamental Rights "*The arts and scientific research shall be free of constraint. Academic freedom shall be respected*" (Art. 13).

[7] Pursuant to WP 29 (2010: 15) "*In the strategic perspective of allocating responsibilities, and in order to provide data subjects with a more stable and reliable reference entity for the exercise of their rights under the Directive, preference should be given to consider as controller the company or body as such rather than a specific person within the company or the body. It is the company or the body which shall be considered ultimately responsible for data processing and the obligations stemming from data protection legislation*".

[8] There might be a practical problem when an individual researcher collects data himself and his host university is unwilling to be the controller. One approach could be that if the host university of a researcher is unwilling to become the controller,

ensure the GDPR compliance the University of Tartu (Estonia) has adopted the Data Protection Policy, which implies its position as the controller. In case academics (researchers, professors) infringe personal data rights, then they are liable before the university. The question is whether academics could be considered the controller as well. The issue is not settled yet. However, European Parliamentary Research Service in its study concerning research (EPRS study 2019: 34) suggests that "*researchers and universities should assume that when processing personal data, their activities render them data controllers*".

The analysis is likely to be different in countries (like Germany or France) where academics (or at least university professors) are not employees, but public servants appointed for life and independent in the exercise of their missions (not unlike, e.g. judges). It is due to this independence that in both Germany and France, e.g. copyright in the works created by academics belongs to them, and not to their university or institution.[9] In light of this rule (referred to as '*professors' privilege*'), it may be hard to argue that professors reap profits of their work (because they are free to decide how to do it), but the university should bear the responsibility for how they process personal data. Along these lines, the French National Centre for Academic Research (CNRS) in its guide on data processing for research purposes (CNRS guide 2019: 12) defines the director of a unit (so an individual, not an institution) as the data controller. Then, the director of a unit designates the CNRS' Data Protection Officer as DPO, thereby providing for a sort of 'bottom-up centralisation' of responsibility for data processing, providing of course that the DPO is regularly consulted and his advice followed.

Furthermore, it is worth noting that controllership is an element of fact. Any contractual or other arrangements made by the interested parties (e.g. mere designation of X as the controller in a consent form or a contract assigning controllership to B, whereas in fact, A determines the means and purposes of processing) is not binding on the data subject or data protection authorities.[10]

3 Legal grounds for sharing language data and freedom of contract to determine duties

Sharing data constitutes processing of personal data which requires the identification of a legal basis. The suitable legal grounds could be the data subject's consent, public interest research or legitimate interest (GDPR Art. 6 (1) (a), (e), (f)). Legal grounds for conducting research have been previously studied (see, Lindén et al. 2019; Kelli et al. 2019).

A key issue here is how much freedom parties have in determining who has which obligations under the GDPR. According to WP29 "*Being a controller is primarily the consequence of the factual circumstance that an entity has chosen to process personal data for its own purposes*" (2010: 8). WP29 clarifies further that the control could originate from the factual influence and the assessment of contractual relations is helpful since relevant actors often see themselves as facilitators rather than controllers. The contractual terms, however, are not decisive (2010: 11). WP29 (10/2006) found based on a functional influence test that an entity (SWIFT), despite presenting itself as a processor, it was a controller. This demonstrates that any designation of controller/processor which does not correspond to the facts is void.

The GDPR requires the controller to conclude a contract with the processor with several requirements (Art. 28). For instance, this contract is needed when an individual entity/person deposits language data with a CLARIN member. Special attention should be given to the content of this agreement. The GDPR requires that the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of GDPR and ensure the protection of the rights of the data subject (Art. 28 (1)) and the processor can process the personal data only on documented instructions from the controller (Art.28 (3) a). It means that the parties must not only refer to general obligations under the GDPR but sufficiently describe the nature of responsibilities, taking into account the real risks and activities carried out under the specific research project to ensure that all appropriate safeguards are provided (see, Art. 89 (1)). It is

the CLARIN Centre needs to do a due diligence to make sure that the data has been properly collected. If the due diligence is done correctly, the host of the CLARIN Centre may as well become co-controller of the data set.

[9] The solution is likely to differ when it comes to patents – e.g. in Germany, the professors' privilege in patent law was abolished in 2002, and the rights to a patentable invention developed by an academic now belong to his or her institution. Since in the field of language technologies university patents remain rather exceptional, we believe that in the context of this paper an analogy with copyright is more accurate.

[10] WP29 (2010: 9): "…even though the designation of a party as data controller or processor in a contract may reveal relevant information regarding the legal status of this party, such contractual designation is nonetheless not decisive in determining its actual status, which must be based on concrete circumstances".

essential to address the reimbursement of expenses incurred by the performance of the obligations and the payment of remuneration within the contract.

The agreement between the controller and the processor must be concluded in written form, including in electronic form (Art. 28 (9)).[11] The electronic form has been clarified by the European Parliament (2018): "*However, the rules for entering into contracts or other legal acts, including in electronic form, are not set forth in the GDPR but in other EU and/or national legislation. The e-commerce Directive (Directive 2000/31/EC) provides for the removal of legal obstacles to the use of electronic contracts. It does not harmonise the form electronic contracts can take. In principle, automated contract processes are lawful. It is not necessary to append an electronic signature to contracts for them to have legal effects. E-signatures are one of several means to prove their conclusion and terms*".

The joint controllers are not required to enter into such a contract as the controller and the processor. A transparent arrangement between the joint controllers must be agreed upon to comply with the GDPR (Art. 26 (1)), and the essence of this arrangement should be made available to the data subjects (Art. 26 (2)). WP29 (2010: 24) explains it as follows "*Parties acting jointly have a certain degree of flexibility in distributing and allocating obligations and responsibilities among them, as long as they ensure full compliance. Rules on how to exercise joint responsibilities should be determined in principle by controllers. However, factual circumstances should be considered also in this case, with a view to assessing whether the arrangements reflect the reality of the underlying data processing*". The controller's responsibilities must be clearly defined in accordance with actual data processing, and the arrangement must reflect the respective roles and relationships of the joint controllers *vis-à-vis* the data subjects (Art. 26 (2)). Otherwise, as indicated by WP29 (2010: 24), the processing is considered "*unlawful due to a lack of transparency and violates the principle of fair processing*".

4 Sharing language data within CLARIN framework

When it comes to data sharing, we can distinguish two different situations: 1) an external individual or entity deposits LD with a CLARIN member; 2) CLARIN member itself makes LD available.

The first case involves the conclusion of a deposition agreement between the depositor and the CLARIN member. The depositor determines the access and use conditions which makes the depositor the controller under the GDPR. The CLARIN member acts as the processor since it processes personal data on behalf of the depositor.

In the second scenario, the CLARIN member shares LD on its behalf and is the controller.

The main question in both scenarios is whether the sharing of LD leads to joint controllership. According to the GDPR joint controllers jointly determine the purposes and means of processing.[12] They need to determine their respective duties (Art. 26). Joint controllership could be relevant in the case of data sharing. The European Court of Justice (ECJ) has explained that "*a broad definition of the concept of 'controller', the effective and comprehensive protection of the persons concerned, the existence of joint liability does not necessarily imply equal responsibility of the various operators engaged in the processing of personal data.*

On the contrary, those operators may be involved at different stages of that processing of personal data and to different degrees" (C‑40/17 para 70). This means that processing at different stages can result in joint controllership. The court has also maintained that *"a religious community is a controller, jointly with its members who engage in preaching, of the processing of personal data carried out by the latter in the context of door-to-door preaching organised, coordinated and encouraged by that community, without it being necessary that the community has access to those data, or to establish that that community has given its members written guidelines or instructions in relation to the data processing*" (C‑25/17 para 75). It says that it is possible to be a joint controller even without having access to PD.

[11] According to Art. 28 (1). „*Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject.*"

[12] In practice, only 'purposes' are much more important than 'means' for determining the controller, cf. WP29 opinion: "*while determining the purpose of the processing would in any case trigger the qualification as controller, determining the means would imply control only when the determination concerns the essential elements of the means*" (2010: 14). It can be argued that in the CLARIN context, where the 'essential elements of the means' are generally similar and known to everyone (computational analysis), only purposes matter.

From the CLARIN perspective, the proposed agreement structure for transfer of personal data aims to establish a CLARIN Centre as a Data Processor serving the national CLARIN consortium with each of the consortium members, or an external party, as a controller of its data sets. To this end, Finland proposes a 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 standard 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 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,
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.

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

In the CLARIN infrastructure, there are three main licensing categories dividing language resources into three groups: 1) Publicly available (PUB); 2) For academic use (ACA) and; 3) For restricted use (RES) (for further discussions, see Oksanen et al. (2010) and Kelli et al. (2018). The CLARIN RES licensing category (for restricted use) is suitable for sharing 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 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 a new legitimate end-user can always obtain a copy directly from CLARIN. As the CLARIN Centre remains a mere processor of personal data to communicate such data to research organisations, the original controller will stay 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 a practical question whether the CLARIN Centre as a processor is commissioned to inform the data subjects or the original controller notifies them, and how one goes about informing them in practice, i.e. will personal communication be possible or is a public announcement sufficient.[13]

5 Conclusion

The determination of the controller is essential since the controller has to guarantee compliance with the GDPR. There could be some misconceptions whether individual researchers or the university is the controller. The authors suggest that the university is the controller. Based on the EU case law, it can be assumed that in addition to an organisation, its members can also be regarded as controllers (see C‑25/17 para 75). There is no established practice regarding universities.

Sharing language data within CLARIN requires specific arrangements depending on the nature of relationships. If a third party deposits data with a CLARIN member, then the party is the controller and the CLARIN member is the processor. If a CLARIN member shares data on its behalf, then it is considered the controller itself.

Several contractual arrangements are needed for sharing language data. However, the main reason for having a CLARIN consortium member become the controller of the PD is that researchers are often very mobile. Sometimes they no longer stay in academia, and sometimes researchers pass away. If someone needs access to the data several years down the road, a CLARIN Centre can still carry on with that role. Even if the researcher no longer is in a position personally to grant access to the data, the data is accessible also with a CLARIN Centre in a co-controller position where both a CLARIN Centre and the researcher separately have control of the data.

[13] It seems to depend on whether the data were collected from the data subject (Art. 13 has only one exception to the obligation to provide information) or not (Art. 14 considers impossibility or disproportionate effort).

References

[C‑311/18] Case C‑311/18. Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems (16 July 2020). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1598506855221&uri=CELEX:62018CJ0311 (27.8.2020).

[C‑40/17] Case C‑40/17. Fashion ID GmbH & Co. KG *vs*. Verbraucherzentrale NRW eV, interveners: Facebook Ireland Ltd, Landesbeauftragte für Datenschutz und Informationsfreiheit Nordrhein-Westfalen (29 July 2019). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1587057502926&uri=CELEX:62017CJ0040 (16.4.2020).

[C‑25/17] Case C‑25/17. Tietosuojavaltuutettu, intervening parties: Jehovan todistajat (10 July 2018). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1587066001018&uri=CELEX:62017CJ0025 (16.4.2020).

[C-22/98] Case C-22/98. Criminal proceedings against Jean Claude Becu, Annie Verweire, Smeg NV and Adia Interim NV (16 September 1999). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1587999262601&uri=CELEX:61998CJ0022 (27.4.2020).

[C-66/85] Case C-66/85. Deborah Lawrie-Blum *vs*. Land Baden-Württemberg (3 July 1986). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1587999733855&uri=CELEX:61985CJ0066 (27.4.2020).

[C-256/01] Case C-256/01. Debra Allonby v Accrington & Rossendale College, Education Lecturing Services, trading as Protocol Professional and Secretary of State for Education and Employment (13 January 2004). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1587999966374&uri=CELEX:62001CJ0256 (27.4.2020).

[CNRS guide 2019] CNRS Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte. GUIDE POUR LA RECHERCHE. Available at https://in-shs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf (26.8.2020).

[DP Handbook 2018] European Union Agency for Fundamental Rights and Council of Europe (2018). Handbook on European data protection law 2018 edition. Available at https://fra.europa.eu/sites/default/files/fra_uploads/fra-coe-edps-2018-handbook-data-protection_en.pdf (29.3.2020).

[Data Protection Policy] Data Protection Policy of the University of Tartu. Available at https://www.ut.ee/en/data-protection-policy (26.8.2020).

[EC] European Commission. What is a data controller or a data processor? Available at https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en (15.4.2020).

[EPRS study 2019] European Parliamentary Research Service. How the General Data Protection Regulation changes the rules for scientific research. July 2019. Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf (27.8.2020).

[European Parliament 2018] European Parliament (2018). Parliamentary questions. Available at https://www.europarl.europa.eu/doceo/document/E-8-2018-003163-ASW_EN.html (27.4.2020).

[EU Charter of Fundamental Rights] Charter of Fundamental Rights of the European Union. 2012/C 326/02. OJ C 326, 26.10.2012, p. 391-407. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (15.4.2020).

[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 https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1555312258399&uri=CELEX:32016R0679 (29.3.2020).

[Kelli et al. 2019] Aleksei Kelli, Krister Lindén, Kadri Vider, Pawel Kamocki, Ramunas Birštonas, Silvia Calamai, Penny Labrpolou, 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 (23.4.2020).

[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 (23.4.2020).

[Kelli et al. 2015] Aleksei Kelli, Kadri Vider, Krister Lindén (2015). 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 (23.4.2020).

[Lindén et al. 2019] Krister Lindén, Aleksei Kelli, Alexandros Nousias, (2019). To Ask or not to Ask: Informed Consent to Participate and Using Data in the Public Interest. Proceedings of CLARIN Annual Conference 2019: CLARIN Annual Conference, Leipzig, Germany, 30 September – 2 October 2019. Ed. K. Simov and M. Eskevich. CLARIN, 56-60. Available at https://office.clarin.eu/v/CE-2019-1512_CLARIN2019_ConferenceProceedings.pdf (23.4.2020).

[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 (24.4.2020).

[Pormeister 2020] Kärt Pormeister (2020). Transparency in Relation to the Data Subject in Genetic Research - an Analysis on the Example of Estonia. Doctoral dissertation. Irene Kull; Jaak Vilo; Katrin Õunap; Barbara Evans (sup). University of Tartu. Available at https://dspace.ut.ee/handle/10062/66697 (26.8.2020).

[Risak & Dullinger 2018] Martin Risak, Thomas Dullinger. The Concept of 'Worker' in EU Law: Status Quo and Potential for Change. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3190912 (27.4.2020).

[WP29 2010] Article 29 Working Party. Opinion 1/2010 on the concepts of "controller" and "processor". Adopted on 16 February 2010. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf (29.3.2020).

[WP29 2006] Article 29 Working Party. Opinion 10/2006 on the processing of personal data by the Society for Worldwide Interbank Financial Telecommunication (SWIFT). Adopted on 22 November 2006. Available at https://www.dataprotection.ro/servlet/ViewDocument?id=234 (27.4.2020).