

Toward a CLARIN Data Protection Code of Conduct

Pawel Kamocki

IDS Mannheim /
Université Paris
Descartes

Erik Ketzan

Birkbeck, University
of London

Julia Wildgans

IDS Mannheim /
Universität
Mannheim

Andreas Witt

Universität zu Köln /
IDS Mannheim /
Universität
Heidelberg

Abstract

This abstract discusses the possibility to adopt a CLARIN Data Protection Code of Conduct pursuant art. 40 of the General Data Protection Regulation. Such a code of conduct would have important benefits for the entire language research community. The final section of this abstract proposes a roadmap to the CLARIN Data Protection Code of Conduct, listing various stages of its drafting and approval procedures.

1. Overview of the General Data Protection Regulation (GDPR)

The General Data Protection Regulation (hereinafter: GDPR) is the EU Regulation 2016/69 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and the free movement of such data. On 25 May 2018 it became directly applicable in all the EU Member States and replaced the Personal Data Directive of 1995. Unlike a directive, which requires implementation in the national law of each Member State, a regulation applies directly and supersedes national rules.

The GDPR is built around the same concepts as the Personal Data Directive. The essential notion of personal data remains the same: it is defined as “any information relating to an identified or identifiable natural person (‘data subject’)” (art. 4 of the GDPR). According to art. 5 of the GDPR, processing of personal data (i.e. any operation performed on such data) shall comply with the principles of lawfulness, fairness and transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality and accountability. In principle, processing is lawful if the data subject has given his informed consent; exceptionally, other grounds for lawfulness are also possible (as listed in art. 6 of the GDPR). Special categories of personal data (i.e. revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health, sex life or sexual orientation) benefit from stricter protection (art. 9 of the GDPR).

The GDPR also reinforces the rights of the data subjects (such as right to information, access, rectification, erasure, restriction, right to data portability or right to object) and — correlatively — obligations of data controllers (e.g. implementing “privacy by design and by default”, keeping a record of processing activities or carrying out a data protection impact assessment) (Voigt, von dem Bussche, 2017).

2. Research Under the GDPR

The rules of the GDPR are meant to protect individual rights and freedoms, but their strict application to certain activities might have a chilling effect e.g. on freedom of research or artistic expression. This

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

is why the European legislator provided for some exceptions concerning processing of personal data in certain special situations, including research or archiving in the public interest. Art. 89 of the GDPR allows for significant derogations from the general framework in such context; however, for fear of an excessive intervention (and violating the subsidiarity principle), the European legislator left some leeway to the Member States concerning this matter. Therefore, the exact rules and procedures concerning processing of personal data for research purposes may vary slightly between Member States (Kelli et al., 2018). Nevertheless, for any derogations to apply, the processing must be “subject to appropriate safeguards for the rights and freedoms of the data subject”. The GDPR remains vague as to what can be construed as such an ‘appropriate safeguard’, providing only one example: pseudonymisation. It is obvious, however, that other ‘appropriate safeguards’ are also envisageable.

3. Bottom-up Standardisation under the GDPR

Given its huge scope, the multitude of interests at stake, and the rate at which technology is advancing, the GDPR necessarily remains vague and leaves room for interpretation in order not to quickly become obsolete. This is also why the European legislator created some instruments for bottom-up standardisation which will hopefully help ‘fill in the blanks’ and contribute to the creation of sector-specific best practices while at the same time guaranteeing a degree of flexibility. These instruments include certification (e.g. data protection marks and seals) and codes of conduct. In this paper, we would like to discuss how the latter could be implemented within CLARIN and the benefits this would achieve.

4. The Scope of a Code of Conduct

Art. 40 of the GDPR invites “associations and other bodies representing categories of data controllers or processors” to prepare codes of conduct “intended to contribute to the proper application of [the GDPR], taking into account the specific features of various processing sectors”. In particular, such codes of conduct could “calibrate the obligations of controllers and processors, taking into account the risk likely to result from the processing” (recital 98).

CLARIN is undoubtedly a body that is competent to draft a code of conduct concerning processing of personal data for the purposes of research in the field of linguistics and language technology. Such a code of conduct should address such questions as (cf. art. 40(2) a):

- fairness and transparency of processing (when shall the processing be considered “not fair”? To what extent the data subject should be informed about the processing and/or involved at the various stages of the processing?);
- the legitimate interests of the controllers, i.e. in which types of situations researchers can be exempted from the obligation to obtain the data subject’s consent?
- the pseudonymisation and anonymisation techniques: which are appropriate for various types of research projects?
- the collection of personal data: in what conditions shall personal data be collected from data subjects, particularly if they belong to a group that merits specific protection (e.g. children, the elderly, immigrants);
- the exercise of rights of data subjects (such as the right to be forgotten or the right of access and rectification);
- specific technical and organisational measures taken to guarantee the security of personal data processing, and to achieve “privacy by design and by default” (cf. art. 25 of the GDPR);
- out-of-court resolution of disputes related to the processing of personal data (e.g. establishing an independent expert arbitration board);
- transfer of personal data to non-EU countries and organisations.

These questions should be answered not only from the legal, but also — and perhaps more importantly — from the technical and the ethical perspectives. This is why in our opinion a multidisciplinary ad-hoc Working Group should be created to draft a CLARIN Data Protection Code of Conduct. This group could consist of delegated members of the Legal Issues Committee, the Committee on Technical Centres, the Standards Committee and possibly also external experts or representatives of data protection authorities.

5. Potential Advantages of a Code of Conduct

A CLARIN Data Protection Code of Conduct would have some obvious benefits for the entire community. First and foremost, it would allay doubts related to the application of the GDPR to the processing of personal data for language research purposes, especially with regards to the special framework applicable to research (art. 89 of the GDPR, cf. supra). This would provide for a greater degree of legal security among researchers and, in the long run, lower the costs of carrying out new projects. In addition, a CLARIN Data Protection Code of Conduct could provide for more consistency within and across national CLARIN consortia and thereby help achieve the GDPR's main purpose, while at the same time spreading and perpetuating good practices in the community, especially among young researchers. Furthermore, the Code would solve another problem related to cross-border sharing of research data, enabling transfer of personal data to partners in non-EU countries who adhere to the CLARIN Data Protection Code of Conduct (cf. art. 46(2)(e) of the GDPR). Last but not least, such a Code could be adhered to also outside of the CLARIN community, thereby increasing the visibility of the CLARIN infrastructure.

The benefits of a CLARIN Data Protection Code of Conduct, both immediate and long-term, are well worth the necessary effort invested in the drafting and approval of the Code.

6. Toward a CLARIN Code of Conduct: a Roadmap

In this section we outline various stages involved in the drafting and approval of the CLARIN Data Protection Code of Conduct. The procedure is outlined in art. 40 of the GDPR.

As mentioned above, we believe that a multidisciplinary ad-hoc Working Group (consisting of internal and external experts on legal, technical and ethical aspects of personal data processing) shall be established and assigned the task of drafting the Code of Conduct.

The first step shall consist of identifying internal and external stakeholders. Possibly, a questionnaire can be created and distributed among the stakeholders, covering all the issues that the Code of Conduct should address (cf. above). The Working Group should collect feedback from various CLARIN projects and institutions with experience in processing personal data both in the field of linguistics and social science. Whenever feasible, representatives of various groups of data subjects should also be included in the process.

A first draft of the Code should be based on the responses collected from the stakeholders. This draft could then be subject to a round of consultation among these stakeholders and other CLARIN bodies.

Once a final version of the draft Code is developed, it shall be submitted to a competent supervisory authority. In our opinion, the draft shall be simultaneously submitted to several national supervisory authorities in various CLARIN countries. This would increase the visibility of the Code of Conduct, as well as the chances for the draft to pass to the next step.

A supervisory authority can approve the draft Code if it finds that it provides for sufficient appropriate safeguards. An approved code of conduct can be adhered to by data controllers and processors. However, such an approval is only valid on the national level and has no cross-border effect (i.e. a code approved in the Netherlands is not automatically regarded as providing appropriate safeguards in Germany and vice versa). Nevertheless, if a supervisory authority decides that the draft code is relevant for many EU Member States, it shall, before approving the Code, submit it to the European Data Protection Board (a body composed of the head of one supervisory authority of each Member State and of the European Data Protection Supervisor, cf. art. 68 of the GDPR), which shall provide its opinion on the draft code.

Moreover, if the Board decides that the draft code provides for appropriate safeguards, it shall notify the European Commission. The EC may then decide to grant the code conduct a general validity in the EU; in other words, it would become a complement to the GDPR and could be adhered to by all data controllers and processors in the EU (i.e. even in those countries where the draft were not submitted to or approved by a supervisory authority). Such a generally valid code of conduct could be also adhered to by data controllers in non-EU member states (e.g. universities and research institutions in the US), which would facilitate the transfer of relevant personal data to these data controllers.

A code of conduct should contain mechanisms for monitoring compliance. This monitoring is mandatory and carried out by a body which is accredited by a supervisory authority and “has an appropriate level of expertise in relation to the subject-matter of the code”. It is advisable that once a CLARIN Data Protection Code of Conduct is adopted, the national CLARIN consortia or CLARIN centres apply for accreditation to monitor compliance with the Code.

The action plan described in this abstract will require time and effort by many CLARIN stakeholders; however, it is our belief that the approval of a CLARIN code of conduct in a single member country, or several slightly different codes in various member countries, would be a significant advancement for the entire CLARIN community.

Reference

Regulation 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) (articles 4, 5, 9, 25, 40, 41, 46, 68 and 89, recitals 98 and 99).

Voigt, P. and A. von dem Bussche (2017). “The EU General Data Protection Regulation (GDPR): A Practical Guide”, Springer 2017.

Kelli, A. et al. (2018). “Processing personal data without the consent of the data subject for the development and use of language resource”, to be presented at the CLARIN Annual Conference 2018.