Originally published in: Erjavec, Tomaž/Eskevich, Maria (Eds.):
Selected papers from the CLARIN Annual Conference 2022. Prague, Czechia, 2022, 10-12 October. -
Linköping: Linköping University Electronic Press, 2023. Pp. 57-65. (Linköping Electronic Conference Proceedings 198)
DOI: *https://doi.org/10.3384/ecp198006*

EU Data Governance Act: Outlining a Potential Role for CLARIN

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

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

Andrius Puksas
Mykolas Romeris University, Lithuania
andrius_puksas@mruni.eu

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

Abstract

The Data Governance Act was proposed in late 2020 as part of the European Strategy for Data, and adopted on 30 May 2022 (as Regulation 2022/868). It will enter into application on 24 September 2023. The Data governance Act is a major development in the legal framework affecting CLARIN and the whole language community. With its new rules on the re-use of data held by the public sector bodies and on the provision of data sharing services, and especially its encouragement of data altruism, the Data Governance Act creates new opportunities and new challenges for CLARIN ERIC. This paper analyses the provisions of the Data Governance Act, and aims at initiating the debate on how they will impact CLARIN and the whole language community.

1 Introduction

The third decade of the 21st century has started with some of the most perturbing events in generations: the COVID-19 pandemic and the Russian aggression against Ukraine, which – understandably so – overshadowed all other developments. Meanwhile, however, the European Union is dynamically modernising its legal framework concerning digital data, or, as the European Commission itself would put it, "shaping Europe's digital future"[1].

In recent years, the adoption of the General Data Protection Regulation (GDPR) and its entry into application have shown that the European Union is a true global "regulatory superpower". As one of the world's largest markets (also in terms of population, but especially in terms of purchasing power), the EU has the power to influence the policies of manufacturers of goods and providers of services worldwide, by imposing standards on goods and services that can enter its market. With the so-called "Brussels effect" (Bradford, 2020), the GDPR has become the global standard for personal data protection. This phenomenon is now likely to extend to other types of digital data, markets, and services. This is why the developments in the EU's legal framework affecting this domain are closely followed not only in Europe, but also on other continents, as they have the potential to – literally – change the digital world.

The Data Governance Act (DGA) can be considered one step further in data governance. The DGA aims to reduce the digital divide, ensure data access neutrality, portability and interoperability, and avoid lock-in effects (Recital 2 of DGA). Its objective is also to improve the conditions for data sharing in the internal market by creating a harmonised framework for data exchanges and laying down certain basic requirements for data governance, paying specific attention to facilitating cooperation between Member

[1] https://digital-strategy.ec.europa.eu/en/policies/strategy-data

Paweł Kamocki, Krister Linden, Andrius Puksas and Aleksei Kelli 2023. EU Data Governance Act: Outlining a Potential Role for CLARIN. *Selected papers from the CLARIN Annual Conference 2022. Ed. by Tomaž Erjavec and Maria Eskevich.* Linköping Electronic Conference Proceedings 198, pp. 57–65. DOI: https://doi.org/10.3384/ecp198006

States“ (Recital 3 of DGA). These aims are underlined by the idea that “data that has been generated or collected by public sector bodies or other entities at the expense of public budgets should benefit society” (Recital 6 of DGA).

This paper will present the content of the DGA from the perspective of CLARIN ERIC and the EU language community as a whole and discuss the impact that this Regulation may have on the functioning of CLARIN ERIC in the future.

In the second section, the authors analyse the concept of data covered by DGA and its interaction with personal data and intellectual property protection. The third section explains the process that led to the adoption of the DGA, and the fourth section explains its contents.

2 Rights in Re-usable Data

Article 2 (1) of the Data Governance Act defines data as “any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording”. The definition of data in DGA is wide and potentially encompasses different types of material. It is similar to definitions of research and language data and therefore entails similar challenges. For further discussion, see Kelli et al. (2020); Kelli et al. (2018).

Recital 12 of the Data Governance Act explains that the re-use[2] regime applies to data that *“the supply of which forms part of the public tasks of the public sector bodies”*. The Chapter of the DGA on the re-use of data provides that it applies to data protected on grounds of statistical confidentiality, intellectual property right (IPR) protection and personal data (Article 3 (1)). Issues related to re-use of data containing IP and personal data are discussed below.

Data defined by the DGA could be subject to intellectual property rights (IPR). According to Recital 17 of the Data Governance Act “This Regulation should neither affect the existence or ownership of intellectual property rights of public sector bodies nor limit the exercise of those rights in any way”. At the same time, Recitals 17 and 18 of the DGA require public sector bodies to exercise their IPRs in a way that facilitates re-use.

The DGA also has more detailed instructions on dealing with IPRs within the framework of conditions for re-use. Article 5 (3) (a) (ii) provides that public sector bodies grant access for the re-use of data only where it has ensured that data has been “modified, aggregated or treated by any other method of disclosure control, in the case of commercially confidential information, including trade secrets or content protected by intellectual property rights.” At the same time, Article 5 (7) of the DGA prescribes that reuse must be in compliance with intellectual property rights.[3]

Suppose the described methods (i.e. modifying, aggregating) are applied to trade secrets and confidential information. In that case, it should not pose any other legal problems than to need a guarantee that secret information is not revealed. When it comes to IP-protected content (especially copyright-protected works), the situation is different. For further discussion on copyright-protected works as language data, see Kelli et al. (2022).

The challenge with modifying copyright-protected works relates to the existence of moral rights (especially the right of integrity). According to Article 6*bis* of the Berne Convention, the author has the right “to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation”. Therefore, to avoid copyright violation, it is essential not to distort or mutilate works during the modification and aggregation process.

There are also further requirements for transmitting data protected by intellectual property rights to a re-user who intends to transfer those data to a third country (Article 5 (10)).

The DGA divides data into two main groups: personal data and non-personal data (Article 2 (3), (4)). According to Article 4 (1) of the General Data Protection Regulation, personal data means “any information relating to an identified or identifiable natural person (‘data subject’)”. Recital 4 of the Data

[2] Article 2 (2) of the DGA defines ‘re-use’ as “the use by natural or legal persons of data held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between public sector bodies purely in pursuit of their public tasks”.

[3] Recital 10 of the DGA emphasizes the same principle for trade secrets: “The re-use of data, which may contain trade secrets, should take place without prejudice to Directive (EU) 2016/943, which sets out the framework for the lawful acquisition, use or disclosure of trade secrets”.

Governance Act emphasises that "this Regulation should not be read as creating a new legal basis for the processing of personal data for any of the regulated activities [...] In the event of a conflict between this Regulation and Union law on the protection of personal data or national law adopted in accordance with such Union law, the relevant Union or national law on the protection of personal data should prevail". This means that processing personal data is regulated by the General Data Protection Regulation and not by the Data Governance Act. Recital 7 of the DGA repeats that legal bases for the processing of personal data is provided in the GDPR.

However, the DGA still foresees re-use of data containing personal data. Article 3 (1) (d) of the Data Governance Act *expressis verbis* provides that the re-use regulation "applies to data held by public sector bodies which are protected on grounds of the protection of personal data". At the same time, Article 5 (3) (a) (i) of the DGA obliges public sector bodies to ensure protection of personal data. Therefore they may grant access or the re-use of data only where public sector bodies have anonymised personal data.

Specific conditions for re-use of data containing personal data or third-party IPRs are discussed below.

3 Toward the Data Governance Act

On 19 February 2020, just a couple of weeks before the first COVID lockdowns, the European Commission launched the European Strategy for Data (European Commission, 2020a). This was followed by a large stakeholders consultation (lasting until 31 May 2020), in which 806 contributions were received, including 98 from academic/research institutions.[4] A series of proposals for Regulations (labelled, in the Anglo-Saxon way, "Acts") were adopted based on this consultation, including:

- Data Governance Act (25 November 2020);
- Digital Services Act (15 December 2020);
- Digital Markets Act (15 December 2020);
- Artificial Intelligence Act (21 April 2021);
- Data Act (23 February 2022).

So far (as of February 2023), the first three of these Acts were adopted: apart from the Data Governance Act (which, after a few modifications, has become the Regulation 2022/868 of 30 May 2022), also the Digital Services Act (Regulation (EU) 2022/2065 of 19 October 2022) and the Digital Markets Act (Regulation (EU) 2022/1925 of 14 September 2022). Both the Artificial Intelligence Act and the Data Act are in advanced stages of their legislative processes.

The DGA will enter into force on 24 September 2023.

4 The Content of the Data Governance Act

As expressly stated in the explanatory memorandum accompanying the Commission's proposal for the DGA (European Commission, 2020b), the Act draws inspiration from the principles for data management and re-use developed for research data, namely the FAIR data principles stipulating that data should, in principle, be findable, accessible, interoperable and re-usable. This principle with a particular attention to a high level of cybersecurity is emphasised in Recital 2 of DGA as well. The Act's announced aim is to address a number of rather diverse practical issues, i.e.:

Making public sector data available for reuse in situations where such data is subject to rights of others.

- Sharing of data among businesses, against remuneration in any form.

[4] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12271-European-Strategy-for-data/public-consultation_en

- Allowing personal data to be used with the help of a 'personal data-sharing intermediary', designed to help individuals exercise their rights under the General Data Protection Regulation (GDPR).
- Allowing data use on altruistic grounds.

From the perspective of the language community, the most important changes introduced by the DGA can be summarised in four points:

- The Act enables wider re-use of 'protected' data held by public sector bodies;
- The Act introduces a mechanism for supervising providers of data sharing services;
- The Act promotes a new paradigm in sharing data for purposes of general interest, referred to as 'data altruism'.
- The Act establishes a European Data Innovation Board.

Each of these points is developed in a relevant sub-section below.

4.1 Wider Re-usability of 'Protected' Data Held by Public Sector Bodies

Public sector bodies (i.e., according to the legal definition [Art. 2(1) of the Open Data Directive], States, regional and local authorities, and bodies established for general interest purposes and financed by the State or regional and local authorities) produce and collect massive amounts of data, which represent considerable economic value. Since the collection and/or production of such data is indirectly financed by the taxpayers, it is justifiable to make these data available for anyone, and re-usable for any purpose, be it commercial or not. This was the general logic behind the Public Sector Information (PSI) Directive (2003/98/EU, amended by Directive 2013/37/EU), which was recently amended by the Open Data Directive (EU) 2019/1024.

Under this framework, Public Sector Information seemed an interesting source of research data for the language community; it was used as one of the main sources of data in the European Language Resources Coordination (ELRC) initiative (Lösch et al., 2018), aimed at collecting Language Resources for the development of automated translation solutions. The amount of public sector data collected within the ELRC initiative (in which Kamocki, one of the co-authors of this paper, was temporarily involved as a legal expert), was rather significantly limited by the fact that a large portion of language data, mono- or multilingual, is effectively excluded from the scope of the rules on the re-use of PSI. Most importantly, under PSI/Open Data Directive documents in which third parties (i.e. anyone else than the public body holding the document) have intellectual property rights, as well as documents containing personal data, are not considered reusable (cf. Article 1(2)(c) and (h) of the Open Data Directive). For example, if a translator (or a translation provider) holds copyright in a document, this document is not covered by the rules on re-use of public sector information under the Open Data Directive. The DGA aims at mitigating this situation.

Under the DGA, public sector bodies have a general obligation to make their datasets available for re-use, even if they contain personal data or data protected by third-party copyright. It does not mean, however, that copyright or personal data protection no longer applies to documents held by public sector bodies; rather, these bodies are obliged to find solutions that enable re-use of concerned datasets in a way that respects the rights of third parties. This can be achieved by sharing pre-processed data (e.g., anonymised, aggregated or otherwise modified, possibly also in a 'derived text format', cf. Article 5(3)(a) of the DGA). Another possibility is to limit access to the data to a secure processing environment provided or controlled by the public sector body (Article 5(3)(b) and (c) of the DGA). Such secure processing environments are already in use for the processing for research purposes of statistical microdata held by Eurostat (on the basis of EU Regulation No 557/2013 No 557/2013), or in the Leibniz Institute for the Social Sciences, also for the processing of statistical microdata (Kamocki et al., 2016). It is to be expected that most public sector bodies will not be able to provide sophisticated virtual data rooms, such as those used in some merger and acquisition transactions, but a simple dedicated room with no Internet access, which cannot be entered with any electronic device, can probably be a sufficient

solution in most cases. Finally, if providing access to the data in secure conditions proves impossible, the public sector body should "make best efforts" to help potential re-users obtain necessary consents or permissions from rightholders (Article 5(7) of the DGA).

For making data available for reuse, public sector bodies may charge fees, but these fees should not exceed the costs incurred; they should also be transparent and non-discriminatory (Article 6 DGA). Moreover, public sector bodies may apply reduced fees, or even allow re-use free of charge, for certain categories of re-users, such as start-ups, small and medium-sized enterprises (SMEs) and educational establishments.

This may sound extremely promising in terms of opening vast amounts of data for re-use; however, the direct impact of the new rules on the research community will be limited due to the fact that, unlike the Open Data Directive, the relevant provisions of the DGA do not include data held by educational and cultural establishments in its scope (Article 3(2)(c) of the DGA). This means that 'protected' data held by museums, archives or libraries are not concerned by the abovementioned rules, and are not to be made available for reuse. The situation of 'protected' research data (for example, corpora of copyright-protected language data, or speech recordings) held by universities is in fact quite uncertain. On the one hand, universities qualify as 'education establishments', and as such should be excluded from the relevant provisions of the DGA. Recital 12 of the DGA, however, suggests the contrary; it reads:

"(...) Research-performing organisations and research-funding organisations could also be organised as public sector bodies or bodies governed by public law.

This Regulation should apply to such hybrid organisations only in their capacity as research-performing organisations."

This seems to indicate that research data held by universities are in fact concerned by the rules on re-use, which has the potential of transforming the role of CLARIN and its centres as providers of (mostly 'protected') research data. One could imagine that in a world where universities and public research organisations are obliged by law to make their research data available for re-use, CLARIN could effectively become an essential intermediary, a one-stop-shop for researchers EU-wide, where they can at least learn which dataset is available at what place, and what are its re-use conditions. Such an intermediary, referred to as "sectoral information point", is mentioned in Article 8 of the DGA, the Article providing that Single Information Points shall be established in every Member State, and a Single European Information Point should be established by the Commission. These Single Information Points shall make available a searchable list of datasets available for re-use in the relevant area (together with information on their size and conditions for re-use). Some of these datasets can be available through Sectoral Information Points. Information Points should work as interfaces for re-users – they shall receive requests for re-use of datasets which are present on their lists, and transmit these requests to the public sector bodies that hold the data. In this approach, a CLARIN centre could become a Sectoral Information Point for Language Data, and maintain a list of all the relevant datasets available in their area (held by universities, but also, e.g., by ministries or local authorities). A user would consult the list, file a request for re-use with the centre, which will then transmit it to the body (e.g., a university) who holds the data, in order to enable the user to access the data through the body's own secure processing environment. This role of a Sectoral Information Point could complement the data hosting task.

Moreover, in order to assist public sector bodies in fulfilling their obligations under the DGA (including to provide them with technical support, e.g. in the field of data anonymisation), Member States should create 'competent bodies' with adequate legal and technical capabilities and expertise (Article 7 of the DGA). This can be done either by establishing new bodies, or entrusting existing bodies (like, for example, National Data Protection Authorities) with these missions. This has a twofold benefit for CLARIN – firstly, CLARIN centres could share their expertise in the field of data sharing with those competent bodies, thereby contributing to making (language) data more widely accessible while also increasing the visibility of the CLARIN network. Secondly, these competent bodies will likely issue guidelines and recommendations regarding sharing of protected data, which will shed light on some of the grey areas of the field, and facilitate CLARIN's own mission.

To sum up, the provisions of the DGA related to the re-use of 'protected' data held by public sector bodies (Chapter II) create many opportunities for CLARIN and the EU language community as a whole – first of all, a wealth of new data, including language data, will be made available for reuse, including for such purposes as developing language resources, and training language models. Secondly, CLARIN

could act as a Sectoral Information Point, thereby assisting researchers in accessing datasets that cannot be hosted in a CLARIN centre due to copyright- or data-protection-related constraints. Thirdly, the DGA will stimulate the development of anonymisation and pseudonymisation techniques and standards, as well as secure solutions for data sharing, which will also open new possibilities for language data. Fourthly and finally, CLARIN's expertise in processing language data can potentially be interesting for the 'competent bodies', which will create new possibilities for liaising with other actors of the data economy, and increase the visibility of CLARIN and its activities.

4.2 New Framework for Data Intermediation Services

Chapter III of the DGA contains some requirements applicable to data intermediation services. These services are defined to include, among others (Article 10 of the DGA):

intermediation services between data holders and potential data users, including making available the technical or other means to enable such services; those services may include bilateral or multilateral exchanges of data or the creation of platforms or databases enabling the exchange or joint use of data, as well as the establishment of other specific infrastructure for the interconnection of data holders with data users.

The activities of CLARIN ERIC (and/or CLARIN B-centres) undoubtedly fall within the scope of this definition. However, it seems that in principle CLARIN and its centres are covered by the exception of Article 15 of the DGA, according to which:

[the relevant provisions] shall not apply to recognised data altruism organisations (see below – authors) *or other not-for-profit entities insofar as their activities consist of seeking to collect data for objectives of general interest, made available by natural or legal persons on the basis of data altruism, unless those organisations and entities aim to establish commercial relationships between an undetermined number of data subjects and data holders on the one hand and data users on the other.*

This suggests that Chapter III of the DGA applies only to those CLARIN centres which decided – or will decide in the future – to get involved in commercial relationships ('with an undetermined number of partners', i.e., probably, making data commercially available to anyone interested). Since such a turn may be considered by some centres, or even the ERIC as a whole, it is worthwhile to discuss here the impact of the DGA on commercial data sharing.

Under Articles 10 and 11, the provision of 'data sharing services' is subject to notification (Articles 10 and 11). This notification is to be made to a 'competent authority' (such authorities are to be designated by every Member State), whose task is to monitor compliance with a set of obligations listed in Article 12 of the DGA. Those include, e.g.,

- the prohibition of re-using data for other purposes than providing them to the users and the obligation to place data sharing services under a separate legal entity (Article 12(a));
- the prohibition to use metadata collected from the provision of the service for other purposes than developing the service (Article 12(c));
- the general obligation to keep the data in the format in which they were provided by the user (Article 12(d)) and
- the obligation to ensure continuity of service and access to the data in case of insolvency (Article 12(h)).

Failing to meet these conditions may result in 'dissuasive' financial penalties and/or cessation of the provision of the service (Article 14(4)).

This strict framework, intended to instil trust in data sharing within the European Data Spaces, may in fact dissuade CLARIN and its centres from involvement in commercial data sharing.

4.3 Promotion of Data Altruism

It has been noted that many individuals and companies are willing to 'donate' their data for a general interest purpose. This phenomenon is generally observed in the domain of medical research, but it can also affect other areas of research: for example, one can imagine writers or publishers who agree to have

their books and other publications analysed by scientists "for the greater good". Chapter IV of the DGA introduces a legal framework for such 'data donations', which it calls 'data altruism'.

'Data altruism' is defined as 'the voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward (...) where they make their data available for objectives of general interest (...) [such as] scientific research purposes (...)' (Article 2(16) of the DGA).

The data can be 'donated' by signing a European data altruism consent form (Article 25 of the DGA), which will be adopted by the European Commission at a later date.

The key role in this new framework is played by so-called 'certified data altruism organisations', trusted organisations that function as intermediaries between the data donators and the users. Their task is to ensure that the data are only used for the purposes of general interest for which they were 'donated'.

It may seem that CLARIN could consider becoming such a 'certified data altruism organisation', although it is not an easy task to accomplish. In order to be certified, an organisation needs to meet objectives of general interest and operate on a non-for profit basis, independently from any for-profit entities (such as private research companies), and through a legally independent structure (Article 18 of the DGA). Moreover, it needs to respect a Rulebook which will be adopted by the European Commission at a later date (Article 22 of the DGA); the Rulebook will specify such aspects as technical requirements related to the storage of the data (regarding security and interoperability standards), or information to be provided to data donors.

Data altruism organisations which meet these requirements shall be subject to registration with a competent authority (which are to be designated by each Member State); they are bound to respect transparency requirements *vis-à-vis* the data holders (Article 20(1) of the DGA; e.g., a full up-to-date list of entities granted access to the data should be provided together with the purpose of processing declared by each of those entities). Furthermore, these organisations will be obliged to submit an annual activity report to the national competent authority (Article 20(2)); if applicable the report shall include a summary of results of the data use allowed by the organisation.

The mechanism of 'data altruism consent' could of course be extremely beneficial – it could potentially make most important legal hurdles in access to language data go away – but the status of a 'data altruism organisation' would require changes in how CLARIN centres operate. In particular, the data obtained through data altruism cannot be made available for re-use to everyone and for every purpose (i.e., under 'open' conditions), but only to researchers, possibly with an obligation to report back on the results. It remains unclear if a 'data altruism organisation' can also provide access to data obtained through other means than the 'data altruism consent', e.g. via open licences or directly from the public domain. Finally, it is not clear whether ERICs are to be granted any form of special treatment with regards to this aspect of the DGA (such as automatic recognition as 'data altruism organisations'). It could be envisaged to split the responsibilities between the various CLARIN entities – some of them (e.g. one per consortium,) could indeed become certified data altruism organisations, while others could operate with commercial actors as 'regular' data intermediaries (see above). It is also envisageable to assign the role of a certified data altruism organisation only to CLARIN ERIC, which will then grant access to the data to the CLARIN centres.

In sum, for CLARIN ERIC, 'data altruism' is a central element of DGA, one which will certainly have to be discussed at the highest level. Unfortunately, as for today many elements of this framework remain unclear – but it is not too early to start a debate.

4.4 Establishment of the European Data Innovation Board

Last but not least, the DGA also provides that the European Commission shall establish a European Data Innovation Board (Chapter VI of the DGA). It will consist of representatives of national competent authorities for data intermediation (see above) and for data altruism (see above), as well as, inter alia, the European Data Protection Board and the European Union Agency for Cybersecurity. In addition, the Commission will launch a call for experts to invite additional members.

The European Data Innovation Board will be divided into at least three sub-groups, one of which will be expressly devoted to standardisation, portability and interoperability of data. Apart from acting as an

advisor to the Commission, the Board will also issue guidelines for Common European Data Spaces, addressing such issues as data interoperability and cross-sectoral standardisation.

In the future, CLARIN will certainly benefit from the guidelines issued by the European Data Innovation Board, or even contribute, directly or indirectly, to their adoption.

5 Conclusion

The Data Governance Act, after its entry into application in September 2023, will be a major development in the legal framework affecting CLARIN and the whole language community. With its new rules on the re-use of data held by the public sector bodies and on the provision of data sharing services, and especially its encouragement of data altruism, the Data Governance Act creates new opportunities and new challenges for CLARIN ERIC.

References

Berne Convention. Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886. Available at https://wipolex.wipo.int/en/text/283698 (26.1.2023).

Bradford, A. 2020. *The The Brussels Effect: How the European Union Rules the World.* Oxford University Press, USA.

DGA. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). OJ L 152, 3.6.2022, p. 1–44. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868&qid=1674669989345 (25.1.2023).

European Commission. 2020a. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions. A European strategy for data.* COM/2020/66 final.

European Commission. 2020b. *Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*. COM(2020) 767 final.

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 (25.1.2023).

Kamocki, Paweł, Katharina Kinder-Kurlanda, Marc Kupietz (2016). „One Does Not Simply Share Data". Organisational and Technical Remedies to Legal Constraints in Research Data Sharing – building bridges between Digital Humanities and the Social Sciences. In: When DH Meets Law: Problems, Solutions, Perspectives. Multiple Paper Session at Digital Humanities 2016, Krakow, Poland, 12-16 July 2016.

Kelli, Aleksei, Arvi Tavast, Krister Lindén. 2022. Building a Chatbot: Challenges under Copyright and Data Protection Law. In: Martin Ebers, Cristina Poncibò, Mimi Zou (Ed.). Contracting and Contract Law in the Age of Artificial Intelligence. (115−134). Hart Publishing. DOI: http://dx.doi.org/10.5040/9781509950713.ch-007 (26.1.2023).

Kelli, Aleksei, Arvi Tavast, Krister Lindén, Kadri Vider, Ramunas Birštonas, Penny Labropoulou, Irene Kull, Gaabriel Tavits, Age Värv, Pavel Stranák, Jan Hajic. 2020. The Impact of Copyright and Personal Data Laws on the Creation and Use of Models for Language Technologies. In: Kiril Simov, Maria Eskevich (Ed.). Selected Papers from the CLARIN Annual Conference 2019 (53−65). Linköping University Electronic Press. Available at https://doi.org/10.3384/ecp2020172008 (25.1.2023).

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

Lösch, Andrea, Valérie Mapelli, Stelios Piperidis, Andrejs Vasiljevs, Lilli Smal, Thierry Declerck, Eileen Schnur, Khalid Choukri and Josef van Genabith. 2018. European Language Resource Coordination: Collecting Language Resources for Public Sector Multilingual Information Management. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pp. 1339-1343.

Open Data Directive. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). - ELT L 172, 26.6.2019, p. 56-83.