EU Data Governance Act : New Opportunities and New Challenges for CLARIN

þÿ K a m o c k i , P a w e B

CLARIN ERIC

2022

Kamocki , P & Linden , K 2022 , EU Data Governance Act : New Opportunities and New Challenges for CLARIN . in T Erjavec & M Eskevich (eds) , CLARIN Annual Conference Proceedings, 2022 . CLARIN Annual Conference Proceeedings , CLARIN ERIC , Utrecht , pp. 44-47 , CLARIN Annual Conference , Prague , Czech Republic , 10/10/2022 .

http://hdl.handle.net/10138/349985

EU Data Governance Act: New Opportunities and New Challenges for CLARIN

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

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

Abstract

The Data Governance Act was proposed in late 2020 as part of the European Strategy for Data, 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, its new limits on international transfers of non-personal data, and especially its encouragement of data altruism, the Data Governance Act creates new opportunities and new challenges for CLARIN ERIC. This abstracts briefly 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 war in Ukraine, which – understandably so – overshadowed all other developments. Meanwhile, however, the European Union is dynamically modernizing its legal framework concerning digital data.

The GDPR and its application has shown that the EU is a global regulatory superpower. As one of the world's largest markets (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.

2 Towards 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 (until 31 May 2020), in which 806 contributions were received, including 98 from academic/research institutions. 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).

As of August 2022, the Digital Services Act and the Digital Markets Act in the final stages of their legislative process; the legislative processes regarding the Data Act and the Artificial Intelligence Act are ongoing; and the only one of these Acts that has already been adopted and published is the Data Governance Act (DGA). The Commission's proposal has, with few modifications, become the Regulation 2022/868 of 30 May 2022 on European data governance.

This abstract will briefly present the content of the DGA from the perspective of CLARIN ERIC and the EU language community as a whole.

3 The Content of the Data Governance Act

As expressly stated in the explanatory memorandum, the DGA was inspired by the FAIR principles. It has a variety of goals, which can be summarized as follows:

- To encourage wider re-use of personal data and data protected by IP, held by public sector bodies.

Under DGA, public sector bodies (such as public administrations – data held by cultural and educational establishments are expressly excluded from the scope of the DGA – Article 3(2)(c) of the DGA) are under a general obligation to make their datasets available for re-use, even if they contain personal data or data protected by third-party copyright (such data are currently excluded from the rules on re-use of public sector information, cf. Article 1(2)(c) and (h) of the Open Data Directive 2019/1024). This can be achieved e.g., by sharing pre-processed (anonymised) data (Article 5(3)(a)(i) of the DGA Proposal), or by limiting data sharing to secure processing environments (Article 5(3)(b) and (c) of the DGA). Where no other solution is available, the DGA imposes an obligation for public sector bodies to 'provide assistance' to potential re-users in seeking consent from data subjects, or permission from rightholders (Article 5(6) of the DGA). In order to assist public sector bodies in fulfilling these obligations (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 creates 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 re-use, including for such purposes as developing language resources, and training language models. Secondly, 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. Thirdly, 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 data economy, and increase the visibility of CLARIN and its activities.

- To introduce a mechanism for supervising providers of data sharing services.

Under the DGA, the provision of 'data sharing services' is subject to notification (Articles 10 and 11). 'Data sharing services' are defined to include, among others (Article 10(a)):

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 notification is to be made to a 'competent authority' (which 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, including, 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)).

CLARIN ERIC (and/or CLARIN B-centres) undoubtedly fall within the definition of a data sharing service provider, and the obligations imposed by the DGA might be very difficult to comply with. However, the regulations mentioned above do not apply to "(…) 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" (Article 15). It may therefore be necessary for CLARIN ERIC and for CLARIN B-centres to strictly limit their activity to collecting data made available on the basis of data altruism (see below).

- To promote 'data altruism'

'Data altruism' is defined as 'the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services' (Article 2(16) of the DGA).

DGA recognises 'data altruism organisations' as legal entities constituted to meet objectives of general interest (such as language research) which operate on a non-for profit basis independently from any for-profit entities (such as private research companies), and through a legally independent structure (separate from other activities). Data altruism organisations which meet these requirements are 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 are 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 uses allowed by the organisation. Article 22 of the DGA obliges the European Commission to adopt (in close cooperation with stakeholders) a rulebook laying down specific rules applicable to data altruism organisations, which shall include *inter alia* appropriate technical and security requirements for data storage, and recommendations on relevant interoperability standards.

The advantage of being granted the statute of a 'data altruism organisation' lie in the mechanism described in Article 25 and referred to as the 'data altruism consent form'. The form will be adopted by the European Commission in a specific procedure and its exact content is impossible to predict; however, it seems that it will be possible for natural persons and legal entities to 'donate' their data to a 'data altruism organisation' in a way that 'circumvents' the some of the current regulations on data protection, and possibly also on intellectual property.

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 licenses 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').

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.

- To restrict transfer of non-personal data to third countries.

The transfer of personal data to third countries is already restricted by the GDPR (Articles 44-50). Many provisions of the DGA aim at restricting the transfer of non-personal data to third countries (i.e., outside the European Economic Area). This applies to non-personal data made available by public sector bodies, as well as by data altruism organisations (see above), which can only be transferred to third countries if certain obligations are met (see esp. Article 31 of the DGA).

This de-globalisation measure may turn out to be an obstacle for CLARIN's collaborations with countries from beyond the European Economic Area (EU, Iceland, Lichtenstein and Norway).

4 Conclusion

The Data Governance Act, after its entry into application in mid-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, its new limits on international transfers of non-personal data, and especially its encouragement of data altruism, the Data Governance Act will create new opportunities and new challenges for CLARIN ERIC.

References

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

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.