Semantifying the Governance of Data in Europe

Beatriz Esteves[1,*], Víctor Rodríguez-Doncel[1]

[1]Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain

Abstract

A new wave of regulations regarding the use and governance of data is being proposed and discussed for adoption in the European Union countries. In this context, on May 2022, the Data Governance Act, a regulation focused on legislating the activity of data intermediation services and data altruism organisations, was approved and should now be applied in all EU member states. This paper describes a set of requirements established by the DGA to protect data subjects and data holders and legislate the area of activity of the competent authorities, as well as a set of scenarios where the application of Semantic Web technologies could help these stakeholders in the fulfilment of their rights and obligations.

Keywords

Data Governance Act, Semantic Web, Solid, Data Intermediaries

1. Introduction

In November 2020, the European Commission (EC) announced a package of new regulation proposals to legislate the European strategy for data [1]. Among them, the proposal for a Regulation of the European Parliament and of the Council on European data governance, the Data Governance Act (DGA), was proposed to improve data availability and promote trust in data intermediation services across the European Union. After the approval by the European Parliament and by the European Council, this new law will now be applicable 15 months after its entry into force date, on May 30th, 2022 [2].

Already an active developer in the data protection field, the Semantic Web community can play an important role in the enforcement of such a law. By promoting an interoperable Web of Linked Data, Semantic Web technologies can be leveraged to model conditions for re-use of public data, to declare the permissions and use requests of data holders and data users (entities that have the right to grant access to data or to lawful access said data for commercial or non-commercial purposes, respectively) in a machine-readable format or to keep records of the processing activities performed by the new organisations introduced by the DGA.

Therefore, the main research objectives of this contribution can be found below:

RO1. Identifying a set of requirements established by the new data governance regulation for the modelling of a DGA Linked Data vocabulary.

SEMANTICS 2022 EU: 18th International Conference on Semantic Systems, September 13-15, 2022, Vienna, Austria

*Corresponding author.

beatriz.gesteves@upm.es (B. Esteves); vrodriguez@fi.upm.es (V. Rodríguez-Doncel)

0000-0003-0259-7560 (B. Esteves); 0000-0003-1076-2511 (V. Rodríguez-Doncel)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3235/paper17.pdf

RO2. Outlining scenarios where Semantic Web technologies can be used to assist stakeholders in representing data-related policy preferences and to keep machine-readable records for the competent authorities.

This paper is organized as follows: Section 2 describes in detail the Data Governance Act and its core regulation goals, while Section 3 introduces a set of scenarios where Linked Data vocabularies and decentralised storage systems, such as Solid, could be leveraged by data altruism organizations and data intermediaries in the implementation of their services and the last section presents conclusions and future lines of work.

2. Data Governance Act

Similar to other data regulations in the European Union, the DGA provides new rights and duties to entities that hold personal and non-personal data and regulates the activity of data users and two types of services, related with data intermediation and altruism. Moreover, the main goals of this regulation are:

1. Facilitating the re-use of protected public-sector data, while preserving its privacy and confidentiality, in situations where such data is subject to the rights of others, including trade secrets, personal data and data protected by intellectual property rights.
2. Regulating and keeping a register of data intermediation services, which allow the sharing of data among businesses and provide individuals with the help of a 'personal data-sharing intermediary', designed to help them exercise their rights under the General Data Protection Regulation (GDPR).
3. Allowing companies and individuals to voluntarily donate data for altruistic purposes, such as medical research.
4. Establishing a new structure, the European Data Innovation Board, to oversee the activities of data intermediation services and data altruism organisations.

In the following subsections, each regulation objective is further detailed.

2.1. Re-use of protected data held by public sector bodies

DGA's Chapter II is dedicated to the re-use of categories of data held by public sector bodies which are safeguarded on the grounds of commercial and statistical confidentiality, the protection of intellectual property rights of third parties or the protection of personal data. These public bodies have to make the conditions of the re-use and the request procedure transparent and publicly available. This right should be regulated by a contract between the involved parties that cannot exceed 12 months and that must include the nature and categories of data and the purposes for re-use. Each EU member state has to designate at least one competent body to provide guidance and technical support to the public sector bodies on the formatting and storage of the data, implementation of privacy-preserving methods to preserve the integrity of personal data and obtaining consent from data subjects and permission from data holders for the data re-use.

2.2. Data intermediation services

One of the novelty inclusions on the DGA is the regulation of data intermediation services. According to this law, such a service "aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means". In particular, in Article 12, a list of 15 conditions for the provision of this type of service is provided, such as the conversion of data into specific formats, using international or European data standards to promote interoperability across sectors, the provision of tools to gather data subjects' consent terms and data holders' permissions, as well as tools to update or withdraw these terms and the maintenance of records of their activity. To provide such a service, the competent authority for data intermediation services must be notified with a set of information details including the identity and contact details of the data intermediation services provider and a description of its activities and a public register of all data intermediation services providers will be kept by the EC. Each EU member state has to designate at least one competent authority to collect the notifications for data intermediation services and supervise their activity.

2.3. Data altruism

Another new concept introduced by DGA involves the concept of 'data altruism' – in essence this term relates to the sharing of both personal data, based on data subjects' consent, and non-personal data, based on data holders' permissions, for 'common good' purposes such as healthcare, combating climate change or scientific research. At a national level, the EU member states can establish their national policies for data altruism and, as with data intermediation services, the competent authority of each country must keep a public registry of the recognised data altruism organisations – these registries should include at least details regarding the identity, legal status and contact details of the organisation as well as information regarding its main goals and nature of the data. In addition, a recognised data altruism organisation has to keep transparent records of its activity, including information regarding the identity of the organisms using the data held by the organisation and the duration and purpose of the processing, as well as to produce an annual activity report for the relevant competent authority. In order to facilitate data collection by these organisations, a European data altruism consent form will be developed to "allow the collection of consent or permission across Member States in a uniform format".

2.4. European Data Innovation Board

In line with the three previously described regulation goals, the DGA also describes the establishment of a new European data structure, the European Data Innovation Board (EDIB), to supervise the activities of data intermediation services and data altruism organisations. This Board will have representatives of both competent authorities for data intermediation services and for the registration of data altruism organisations of all EU members, as well as representatives with specific expertise on standardisation, portability and interoperability and other relevant stakeholders. In particular, in Article 30, a list of 13 EDIB tasks is described, including

guidance for a consistent practice of data altruism across the EU or the proposal of guidelines for the creation of common European data spaces.

3. Semantic Web meets the DGA

This section introduces a set of use cases where Semantic Web technologies can be used by the new entities described in the DGA for the implementation of their services.

Policies for the re-use of public data Standardised policy languages, such as the W3C Open Digital Rights Language (ODRL) [3], can be leveraged to define permissions and duties for the processing of public-sector data as they allow for the drafting of fine-grained policies which can be constrained to particular recipients or purposes. Expressing such policies in a common format would also allow the development and deployment of services to do policy conformance checking.

Solid Pod providers as intermediaries for personal data Solid[1], a decentralised storage initiative based on open and interoperable Web specifications such as HTTP or the Linked Data Platform standards, provides its users with personal online datastores, 'Pods'. Different Pod providers can provide data intermediation as a service, allowing users to select which data intermediation provider relates more to their preferences for the processing of personal data.

Records of data altruism activities Similar to the GDPR's Register of Processing Activities (ROPA), related to the handling of personal data, the DGA also includes duties on the competent data altruism authorities to keep an up-to-date record of the activities of such organisations. There is already published work for a common semantic model for GDPR's ROPA[4], which can be assessed and extended to deal with the requirements of the registries mandated by the DGA.

Machine-readable data altruism consent form Vocabularies such as the Data Privacy Vocabulary (DPV) [5] or GConsent [6], which already model purposes for processing, duration and entity identity data, as well as information regarding data subjects' consent, can be extended to create a unique machine-readable consent form to be used in all European countries by data altruism organisations. This extension should include altruism as a purpose for processing data as well as additional legal bases and other taxonomies to model specific concepts such as benefits or detriments.

Interoperability of data among different competent bodies The usage of common Semantic Web vocabularies also allows to have linked records of the different data-related services accessible by the different competent authorities on data protection regulation, as well as by supervisory bodies such as the European Data Protection Board or the newly-founded European Data Innovation Board.

[1] https://solidproject.org/

4. Conclusions and Future Work

In this contribution, we described the requirements brought on by the adoption of a new data protection regulation in Europe. The Data Governance Act improves on the previous 2019 Open Data Directive[2] by regulating the activity of two new data economy services for the sharing of personal and non-personal data. Alongside the analysis of these requirements stated in this regulation articles, a set of different use cases were described where Semantic Web vocabularies and data storage solutions can be used to promote interoperability of data formats and to assist in the linkage of data generated by the different records of activities mandated by the regulation.

In future lines of work, the development of a linked data vocabulary for the DGA, including the new introduced stakeholders, the conditions for the establishment of data intermediation service providers and data altruism organisations and the information to be included in their registries of activities, will provide the concepts needed to deal with the scenarios drafted in the previous section.

Acknowledgments

This research has been supported by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497 (PROTECT).

References

[1] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - A European strategy for data, 2020. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066.

[2] 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), 2022. URL: http://data.europa.eu/eli/reg/2022/868/oj/eng.

[3] R. Iannella, M. Steidl, S. Myles, V. Rodríguez-Doncel, ODRL Vocabulary & Expression 2.2, 2018. URL: https://www.w3.org/TR/odrl-vocab/.

[4] P. Ryan, H. J. Pandit, R. Brennan, A Common Semantic Model of the GDPR Register of Processing Activities (2020) 251–254. doi:10.3233/FAIA200876.

[5] H. J. Pandit, A. Polleres, B. Bos, R. Brennan, B. Bruegger, F. J. Ekaputra, J. D. Fernández, R. G. Hamed, E. Kiesling, M. Lizar, E. Schlehahn, S. Steyskal, R. Wenning, Creating a Vocabulary for Data Privacy: The First-Year Report of Data Privacy Vocabularies and Controls Community Group (DPVCG), in: OTM 2019 Conferences, volume 11877, Springer International Publishing, 2019, pp. 714–730. doi:10.1007/978-3-030-33246-4_44.

[6] H. J. Pandit, C. Debruyne, D. O'Sullivan, D. Lewis, GConsent - A Consent Ontology Based on the GDPR, in: The Semantic Web, volume 11503 of *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 270–282. doi:10.1007/978-3-030-21348-0_18.

[2] https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data