Knowledge Graphs: Semantics, Machine Learning, and Languages
M. Acosta et al. (Eds.)
© 2023 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SSW230015

Semantics for Implementing Data Reuse and Altruism Under EU's Data Governance Act

Beatriz ESTEVES [a], Víctor RODRÍGUEZ DONCEL [a]
Harshvardhan J. PANDIT [b] Dave LEWIS [c]
[a] *Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*
[b] *ADAPT Centre, Dublin City University, Ireland*
[c] *ADAPT Centre, Trinity College Dublin, Ireland*

Abstract. *Purpose:* Following the impact of the GDPR on the regulation of the use of personal data of European citizens, the European Commission is now focused on implementing a common data strategy to promote the (re)use and sharing of data between citizens, companies and governments while maintaining it under the control of the entities that generated it. In this context, the Data Governance Act (DGA) emphasizes the altruistic reuse of data and the emergence of data intermediaries as trusted entities that do not have an interest in analysing the data itself and act only as enablers of the sharing of data between data holders and data users. *Methodology:* In order to address DGA's new requirements, this work investigates how to apply existing Semantic Web vocabularies to (1) generate machine-readable policies for the reuse of public data, (2) specify data altruism consent terms and (3) create uniform registers of data altruism organisations and intermediation services' providers. *Findings:* In addition to promoting machine-readability and interoperability, the use of the identified semantic vocabularies eases the modelling of data-sharing policies and consent forms across different use cases and provides a common semantic model to keep a public register of data intermediaries and altruism organisations, as well as records of their activities. Since these vocabularies are openly accessible and easily extendable, the modelling of new terms that cater to DGA-specific requirements is also facilitated. *Value:* The main results are an ad-hoc vocabulary with the new terms and examples of usage, which are available at https://w3id.org/dgaterms. In future research, this work can be used to automate the generation of documentation for the new DGA data-sharing entities and be extended to deal with requirements from other data-related regulations.

Keywords. Data Governance Act, Semantic Web, Machine-readable policies, Data intermediaries, Data altruism, Registers of Activities

1. Introduction

In February 2020, following the impact of the General Data Protection Regulation (GDPR) [1] in the specification of new data subject rights and in the implemen-

tation of new obligations on the entities processing personal data, the European Commission published a document establishing its *strategy for data*, including a package of new regulation proposals to legislate the usage of non-personal and public data, the activity of digital services and digital markets and the development of *common European data spaces* [2]. While putting the data in the centre of this transformation, by making it available to all and facilitating its flow between sectors, the interests of data subjects and data holders will be kept by having clear data-sharing policies to govern the usage and access to data and trusted entities that enable said sharing while enforcing compliance with the new regulations.

In particular, the Regulation of the European Parliament and of the Council on European data governance, the Data Governance Act (DGA), was proposed to improve the availability of public data, promote trust in data intermediation service providers and data altruism organisations as enablers of data-sharing between data holders and data users for purposes of general interest, and to establish a new supra-national authority charged with overseeing the activities of such entities, the European Data Innovation Board [3]. The DGA, along with the proposed visions for an European Health Data Space [4] and the Data Act [5], put an emphasis on the altruistic reuse of data – in the Health Data Spaces case to address the challenges of access and sharing to electronic health data – currently trapped within various institutions in the EU and unavailable to be used by all. The key challenges to be addressed to realise these visions are related to the:

Ch1. *Availability / Discovery of datasets*: without the promotion and technical support for the development of common data spaces and trusted data sharing entities, data subjects and data holders will not have automated tools to share their data to be reused for common good purposes, nor solutions to support them in the exercising of their rights, and data users will not have tools to search for the data they seek.

Ch2. *Establishment of conditions for usage and access to data*: without standards and metadata vocabularies to express interoperable, machine-readable policies, the establishment of conditions for usage and access to personal, non-personal and public-sector data, based not only on legal but also on ethical, organisational and social norms, will provoke interoperability issues between entities providing and seeking access to data.

Ch3. *Production of Documentation*: without keeping records of their activities in a structured format, data intermediation service providers and data altruism organisations will rely on manual processes to produce documentation that demonstrates their accountable and responsible practices.

Therefore, Semantic Web vocabularies, such as the W3C's Data Privacy Vocabulary (DPV)[1], Open Digital Rights Language (ODRL)[2] or Data Catalog Vocabulary (DCAT)[3], have an important role to play in these processes as they are interoperable and form common standards that enable machine-readable tools to be used for the automation of tasks. DCAT and ODRL are W3C Recommendations to describe data catalogues published on the Web and to express usage rules

[1] https://w3id.org/dpv
[2] https://www.w3.org/TR/odrl-model/, https://www.w3.org/TR/odrl-vocab/
[3] https://www.w3.org/TR/vocab-dcat-2/

over digital datasets, respectively. DPV is a W3C Community Group Report, which has recently published a stable version 1, aimed at providing a complete, open-access set of taxonomies to express machine-readable metadata about the use and processing of personal data, such as taxonomies for legal entities, purposes, types of processing activities, legal basis, types of data, rights or technical and organisational measures. By combining the usage of these standards and specifications to automate the discovery of datasets (Ch1), specify policies for the reuse and sharing of data (Ch2), and comply with legal obligations (Ch3), such as sending notifications to the competent authorities under the DGA, this work will enable organisations to gradually move from completely manual processes to ones based on utilising automation and technologies to assist in ensuring correctness and scalable architectures on the data-sharing services ecosystem.

In order to address the identified challenges, we determined what reuse conditions are necessary to specify *how* to share data, *who* are the new involved stakeholders, and *what* documents are required to comply with the new law. Therefore, the following research objectives are presented as the basis of this work:

RO1. Identify the stakeholders, information items and information flows relevant for the sharing of data compliant with the DGA.
RO2. Identify terms missing from W3C's specifications for representing data-sharing policies and consent terms.
RO3. Generate registers of altruistic and data intermediary activities which can be queried by the competent authorities.

Moreover, the principal contributions of this paper are summarised as follows:

C1. Identification of DGA entities and how data flows between them.
C2. Identification of Use Cases where the usage of semantic vocabularies will assist in the automation of tasks.
C3. Development of ad-hoc vocabulary for representing data-sharing policies, consent and permission terms and registries of activities.
C4. Demonstration of representation of DGA-related information using the mentioned semantic web technologies and the developed vocabulary.

This paper is organized as follows: Section 2 describes the entities and flows of data between entities defined by the DGA in which the usage of semantic technologies can promote the automation of tasks, while Section 3 discusses the state of the art in semantic models for data governance. Section 4 provides an identification of vocabulary terms that can be reused for the purposes of providing examples for policies for the reuse and sharing of data, querying registries of activities and specifying data altruism form terms. Section 5 discusses the impact of our approach on compliance with DGA and its limitations and the last section presents conclusions and future lines of work.

2. Information Flows in the DGA

As the DGA promotes the availability and regulates the sharing of data, a set of information flows, related to the information that needs to be exchanged between

data-sharing entities, can be described. In this context, an information flow specifies the information that has to be transmitted from one entity to another or that needs to be kept in a document, such as a record of activity or a public register, to fulfil a certain DGA requirement. Figure 1 displays a diagram of the identified entities and information flows.

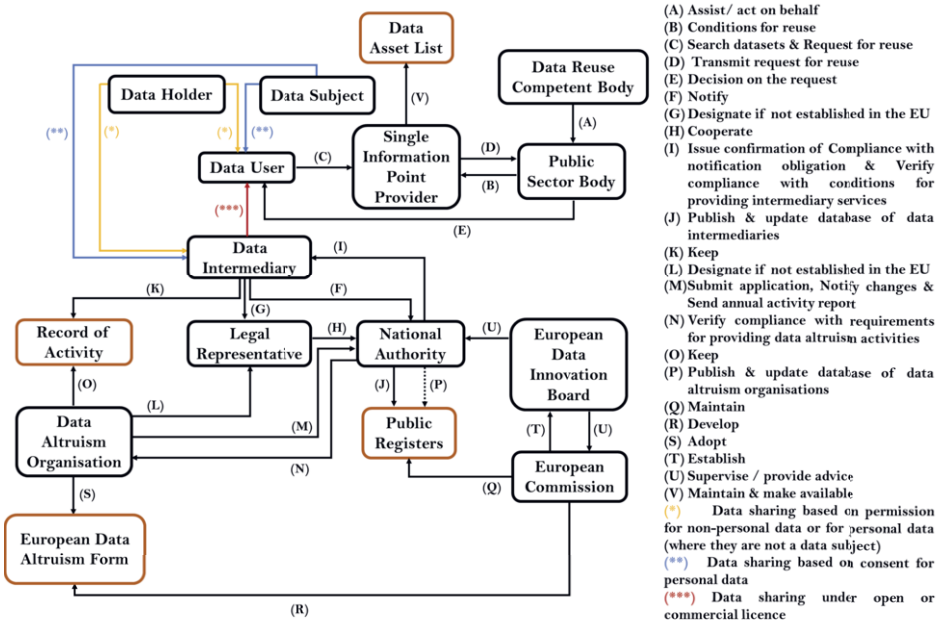

Figure 1. Flows of information between DGA-specified stakeholders. The concepts surrounded by a black box represent (legal) entities and the ones surrounded by an orange box represent newly introduced documents, to be created and maintained by the identified entities. The direction of the arrows represents the direction of the information flow between entities. A simple description of each information flow is provided on the right side of the Figure.

The identified entities can be classified as a data holder, data subject, data user, data intermediation service provider – or data intermediary –, data altruism organisation, legal representative, national, or competent, authority, single information point provider, public sector body, or competent body. Information flows including the soon-to-be-created European Data Innovation Board and European Commission are also displayed in Figure 1. Their definitions are presented below:

Data Subject Individual whose personal data is being processed

Data Holder An entity who has the right to grant access to or to share certain personal data or non-personal data

Data User An entity who has the right to use personal or non-personal data for commercial or non-commercial purposes

Data Intermediation Service Provider An entity who establishes commercial relationships for the data sharing between data subjects and data holders on the one hand and data users on the other

Data Altruism Organisation An non-profit organisation that collects and shares data for altruistic purposes

Public Sector Body An entity or association of entities governed by public law formed by one or more State, regional or local authorities

Legal Representative A representative of a legal entity designated to act on behalf of a data intermediation service provider or altruism organisation

Competent Body An entity designated by a public sector body to provide legal and technical support on the access and reuse of public sector data

Single Information Point Provider An entity who is responsible for receiving and transmitting requests for the re-use of public data

Competent Authorities Authorities in charge of supervising the activity of data intermediation service providers and data altruism organisations and maintaining a public register of said entities

European Data Innovation Board An authority tasked with overseeing the activities of data intermediaries and data altruism organisations

For instance, the data intermediary shares the conditions for data access under an open or commercial licence with the data user (flow represented in Figure 1 with a red arrow) and keeps a record of its activities (flow represented in Figure 1 with the (K) arrow). This diagram is derived from an analysis of Chapters II ('Re-use of certain categories of protected data held by public sector bodies'), III ('Requirements applicable to data intermediation services'), IV ('Data altruism') and VI ('European Data Innovation Board') of the DGA. Each article in these chapters was manually studied to search for interactions between the identified entities and, when a flow of information was identified between more than one entity, the respective interaction was recorded in the diagram. In addition, requirements related to the production of compliance documentation are also recorded in the diagram as they require the recording of information which can be automated with the usage of semantic technologies.

In the next three subsections, we focus on the information flows related to the conditions for the re-use of public data (subsection 2.1), with keeping registers of altruistic and intermediary activities (subsection 2.2), and with data altruism activities (subsection 2.3), where the usage of semantic technologies can best assist the involved entities in automating their flow-related tasks. For each example use case, a methodical study of the involved information flows, and respective items of information that need to be exchanged, was manually performed for each identified flow and systematised in the following subsections.

2.1. Use Case U1: Conditions for the Reuse of Public Data

DGA's Chapter II is dedicated to the 'Re-use of certain categories of protected data held by public sector bodies', including the specification of what categories of data it refers to (Article 3), the information conditions that public sector bodies need to document in order to provide such services (Article 5 and 6) and the description of single information point providers and how they are used by data users to search for and request datasets for re-use (Article 8 and 9). The information that public sector bodies need to provide, and a list of the DGA's source articles where it is mentioned, is available in Table 1. This information,

which can be specified with the assistance of a competent body (as represented by the (A) arrow in Figure 1), needs to be shared with the single information point provider (as represented by the (B) arrow in Figure 1), so that data users can search datasets ((C) arrow) and send a request for re-use of data through the single information point ((D) arrow). Single information point providers must also maintain and make available a data asset list (represented by the (V) arrow), including information on available resources and the conditions for their re-use.

Table 1. Information items about public sector bodies' services.

Article	Information items
2.9	Data user/categories of users
5.1	Public sector body information
5.1	Competent body information
5.2	Categories of data
5.2	Purposes for usage and access
5.2, 5.3(a)	Nature of data
5.3(b), 5.3(c)	Processing environment
5.5	Measures to prevent re-identification of data holders/subjects
5.9	Third party recipients
6.2	Fees
8.2	Data format
8.2	Data size
9	Procedure to request reuse

2.2. Use Case U2: Registers of Altruistic and Intermediation Activities

DGA's Chapter III and IV is dedicated to the requirements applicable to entities who wish to provide data intermediation or data altruism services. As for the former, and as is defined in DGA's Article 11, the entities who wish to provide data intermediation as a service need to notify their competent national authority of said intentions (as represented by the (F) arrow in Figure 1), which in turn must publish and maintain an updated public register of intermediaries (as represented by the (J) arrow). The conditions required to perform such service are depicted in Article 12, such as the requirement to appoint a legal representative if the data intermediation entity is not established in the EU (as represented by the (G) arrow), to provide information about the commercial terms of the service, including pricing, date and time of the creation of the data and its geolocation, the format of the data and which formats it can be converted, and about the tools and measures used by the intermediary to facilitate the exchange of data, to protect and ensure its interoperability, and to ease the exercising of data holders and data subjects' rights, including the tools to obtain and withdraw permissions and consent, respectively. In addition to these conditions, the data intermediation service provider must also keep a log record of its activities (as represented by the (K) arrow), which in addition to the previously mentioned conditions, must also contain the entity-related information which is made available in the public register of intermediation providers, including name, public website, legal status,

form, ownership structure, subsidiaries, registration number and address of the provider, as well as information regarding the type of provided service.

As for the requirements to open activity as a data altruism organisation, and as is defined in DGA's Article 19, the entities who wish to provide data altruism as a service need to submit an application to their competent national authority (can be the same authority as the one who regulates the national data intermediation service providers) of said intentions (as represented by the (M) arrow in Figure 1). If approved, the national authority must include information about the organisation on a public register of data altruism organisations (as represented by the (P) arrow). Such register includes information regarding the name, public website, legal status, form, registration number of the entity, and the entity's, and its representative if applicable, contact details, as well as information regarding the altruistic purposes behind the activity of the organisation. Moreover, the organisation has to publish and update a uniform and structured record of data altruism activity (as represented by the (O) arrow), which is sent annually to the national authority for verification of compliance (as represented by the (M) and (N) arrows). This record must log the activity of the data altruism organisation and provide information regarding the nature and categories of data that it works with. In addition, such records need to keep logs regarding the users of data, their contact details, the date and duration of the processing, the altruistic purpose for which the data was used, the fees paid by data users or any other sources of income, the technical means used for the processing, as well as a summary of the results of said processing.

2.3. Use Case U3: Allowing Data Altruism

DGA's Article 25 discusses the implementation and development of a "European data altruism consent form", which shall be developed by the European Commission (as represented by the (R) arrow in Figure 1), after consulting with GDPR's watchdog European Data Protection Board, with the soon-to-be-created European Data Innovation Board and with other interested stakeholders. This form should be adopted by the data altruism organisations (as represented by the (S) arrow) to record both the consent given by data subjects to share their personal data and the permissions given by data holders to share their non-personal data. These forms should be kept in both a human and machine-readable format and, as such, are the focus of Use Case U3 as semantic technologies, such as ODRL and DPV, can be used to create an electronic rendition of these documents.

3. Related Work

Jurisdictional data-related laws, such as the DGA or the Data Act, specify obligations and requirements based on the context, purpose, and entities involved in how the data is obtained, used, and shared. For a system to conduct, document, and verify compliance-related activities, such as the maintenance of public registers and records of activities, different types of information need to be represented: (i) the obligations and requirements; (ii) the personal, non-personal and public

data and (iii) the data use. Previous work has been performed and published within the general fields of 'regulation compliance' and 'legal metadata expression using vocabularies' [6], to specify how jurisdictional laws can be translated into semantic models for data governance. In the context of this work, we focus on the existing research and solutions, limited to addressing the requirements presented in the Use Cases specified in Section 2, and present the state of the art across the areas of (i) vocabularies to express policies and data activities-related metadata, and (ii) vocabularies to specify information about (personal) data and metadata processing, further described in the next two subsections.

3.1. Vocabularies to Express Policies and Metadata

A recent survey [7] has been published where a set of vocabularies and policy languages are analysed in terms of their capacity to represent the information required to comply with the obligations and rights of GDPR-related entities. In particular, this survey concludes that ODRL is a mature resource, ready to be used for representing policies related to data protection law requirements, which is open source, supported by good documentation and continues to be developed and maintained by a W3C Community Group. The ODRL Information Model [8] is a W3C Standard Recommendation that allows the expression of flexible and complex digital policies, including the possibility to represent permissions and prohibitions to perform certain actions over assets and further restrict said policies using constraints and duties. ODRL also supports the development of extensions, the so-called ODRL profiles[4], that provide a way to add further terms for specific domains which are not present in the core ODRL vocabulary. Though other solutions, such as XACML [9] or LegalRuleML [10], provide a richer expressivity and formal semantics to utilise such resources, ODRL has a convenient extension mechanism and has been proven to work as a policy language to deal with GDPR requirements [11,12]. Other general vocabularies, such as the W3C DCAT Recommendation [13] or the DCMI Metadata Terms (DCT) specification [14], will also be used as they provide terms to describe metadata related with datasets including information about the entities who create and maintain data or temporal and spatial assertions regarding the usage and access to data.

3.2. Legal Vocabularies to Specify Data and its Processing

A vocabulary specifying legal concepts is required for expressing policies aligned with data-related regulations and, in the case of this work, one that can easily complement and be integrated with ODRL, and the other previously mentioned vocabularies, to express examples related to the Use Cases specified in Section 2. While no work within the state of the art provides concepts to deal with DGA requirements, several vocabularies have been developed to cover GDPR concepts that can be reused. In particular, and as confirmed by the previously cited survey on data protection vocabularies [7], DPV's [15] set of taxonomies provides the most complete set of vocabularies to express information regarding data, entities, processing activities, purposes, legal basis, rights, risks and consequences, tech-

[4]ODRL Profile Best Practices - https://w3c.github.io/odrl/profile-bp/

nical and organisational measures, rules or technologies. Moreover, there already exists published work that uses DPV to create a semantic model for the representation of information related to GDPR's Register of Processing Activities [16]. As such, DPV will be the base vocabulary upon which this work will be developed.

4. Extending W3C vocabularies to cover DGA requirements

As covered by the previous sections, there is a gap in the representation of information brought by the DGA requirements, in particular, to specify conditions for the reuse of public data (further developed in Section 4.1), to populate public registers of data intermediation service providers and data altruism organisations and record their activities (further developed in Section 4.2) and to create a common data altruism form for data subjects' consent and data holders' permissions (further developed in Section 4.3). In the following subsections, we discuss terms of existing standards and specifications that can be used to represent some of the information items described in Section 2 and define the terms that are missing in an open-source ad-hoc vocabulary, to cover the identified Use Cases. In addition, for each Use Case, we also provide examples to demonstrate their applicability.

4.1. Policies for the Reuse and Sharing of Public Data

As described in Section 2.1, public sector bodies need to provide single information point providers information regarding the data resources they own and the conditions for their usage, so that these providers can make available and maintain a searchable asset list, which data users can use to search and request datasets for re-use. Table 2 contains the DPV, DCAT and DCT's terms that can be reused to model some of the concepts identified in Table 1.

Table 2. Information items that need to be modelled to express the conditions of re-use of public sector bodies datasets and respective terms from existing vocabularies that can be reused.

Article	Information items	Terms from existing vocabularies
5.1	Public sector body information	`dpv:hasName`, `dpv:hasContact`
5.1	Competent body information	`dpv:hasName`, `dpv:hasContact`
5.2	Categories of data	`dpv:hasData`, `dpv:Data`
5.2	Purposes for usage and access	`dpv:hasPurpose`, `dpv:Purpose`
5.3(a)	Nature of data	`dpv:hasData`, `dpv:AnonymisedData`, `dpv:PseudonymisedData`
5.3(b), 5.3(c)	Processing environment	`dpv:ProcessingContext`, `dpv:hasLocation` `dpv:WithinVirtualEnvironment`, `dpv:WithinPhysicalEnvironment`
5.5	Technical and operational measures to prevent re-identification of data holders/subjects	`dpv:Deidentification`
5.9	Third party recipients	`dpv:ThirdParty`
8.2	Data format	`dcat:mediaType`, `dct:format`
8.2	Data size	`dct:extent`

In addition to these, to specify data users, public sector bodies, competent bodies and single information point providers, we added four new classes of entities (as subclasses of `dpv:LegalEntity`) to our vocabulary to represent these terms, `DataUser`, `PublicSectorBody`, `DataReuseCompetentBody`, and `SingleInformationPointProvider`, respectively. EU, national, regional, local and sectorial-level single information point providers are also modelled as subclasses of `SingleInformationPointProvider`s, as depicted in DGA's Article 8. To be able to classify the nature of the data held by public sector bodies, as specified in Article 3.1, we also added four new subclasses of `dpv:Data`, `ConfidentialData`, `CommerciallyConfidentialData`, `StatisticallyConfidentialData` and `IntellectualProperty` to represent data protected through `CommercialConfidentialityAgreement`s or through `StatisticalConfidentialityAgreement`s and data protected by intellectual property rights.

Moreover, the following legal basis for the transfer of public sector body-held data, as specified in Article 5, are also included in our vocabulary, as subclasses of `dpv:DataTransferLegalBasis`: `A5-9` for permissions to transfer, `A5-11` for model contractual clauses, and `A5-12` for adequacy decisions. `DataReusePolicy`, `DataTransferNotice` and `ThirdCountryDataRequestNotice` concepts were also added, as subclasses of DPV's policy and notice concepts, to represent the conditions for reuse of data and the notice provided to the owners of said data. As there were no concepts identified to model the searchable asset list maintained by the `SingleInformationPointProvider`s and the procedure to request datasets, both concepts were modelled as `DataAssetList` and as `DataReuseRequestProcedure` and as subclasses of `dpv:OrganisationalMeasure`.

To showcase the usage of existing and newly created terms, an example `DataReusePolicy` for reusing the *http://example.com/dataset_001* dataset, that can be used until the end of 2023 for the purpose of `ScientificResearch`, can be found in Listing 1[5]. It is modelled as an ODRL offer as it proposes the terms of usage of the dataset, but does not grant any privileges to the data user. Said policy can be used by single information point providers to maintain an updated list of available assets and the conditions for their usage. Listing 2 provides an example of a `DataAssetList` published by a `SingleInformationPointProvider`, using the existing and the newly created terms. This list contains the previously mentioned dataset, *http://example.com/dataset_001*, with additional metadata regarding the category of data it contains, the policy that governs its usage, *http://example.com/policy_001*, the format and size of the data and the fees charged by the publisher of the dataset.

4.2. Querying public registers of data intermediaries

As described in Section 2.2, data intermediation service providers and data altruism organisations need to submit information about their activity to a public register of such entities in order to have a centralised database of entities, which can be used by data users, data holders or data subjects to retrieve or publish data, for instance for altruistic purposes.

[5]The prefixes and namespaces described in Listing 1 are valid for all Listings.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX odrl: <http://www.w3.org/ns/odrl/2/>
PREFIX dpv: <https://w3id.org/dpv#>
PREFIX dpv-pd: <https://w3id.org/dpv/dpv-pd#>
PREFIX dpv-gdpr: <https://w3id.org/dpv/dpv-gdpr#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://example.com/>
PREFIX : <http://anon/dgaterms#>

ex:policy_001 a odrl:Offer, :DataReusePolicy ;
    odrl:permission [
        odrl:target ex:dataset_001 ; odrl:action :Reuse ;
        odrl:assigner ex:publicsectorbodyX ;
        odrl:constraint [
            odrl:and [
                odrl:leftOperand odrl:dateTime ;
                odrl:operator odrl:lteq ;
                odrl:rightOperand "2023-12-31"^^xsd:date ], [
                odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:isA ;
                odrl:rightOperand :ScientificResearch ] ] ] .
ex:publicsectorbodyX a :PublicSectorBody ;
    dpv:hasName "Public Sector Body X" ;
    dpv:hasContact "mailto:publicsectorbodyX@email.com" ;
    :hasCompetentBody [
        a :DataReuseCompetentBody ; dpv:hasName "Competent Body X" ;
        dpv:hasContact "mailto:competentbodyX@email.com" ] .
```

Listing 1: ODRL Offer policy set by the Public Sector Body X that permits the re-use of a dataset until the end of 2023 for scientific research.

```
ex:SIPPA_assets a :DataAssetList, dcat:Catalog ;
    dct:description "Asset list maintained by SIPPA" ;
    dct:created "2022-12-10"^^xsd:date ;
    dct:publisher ex:SIPPA ; dcat:dataset ex:dataset_001 .
ex:SIPPA a :SingleInformationPointProvider .
ex:dataset_001 a dcat:Dataset ; dct:publisher ex:publicsectorbodyX ;
    dpv:hasData :StatisticallyConfidentialData ;
    dct:description "Dataset with statistically confidential data" ;
    dct:created "2022-12-04"^^xsd:date ;
    odrl:hasPolicy ex:policy_001 ; :hasFee "0€"^^xsd:string ;
    dcat:mediaType <iana.org/assignments/media-types/text/csv> ;
    dct:extent "5.6MB"^^xsd:string .
```

Listing 2: Data asset list maintained by the Single Information Point Provider A.

```
ex:publicregistry_DI_PT a :RegisterOfDataIntermediationServiceProviders ;
    dct:description "Public register of intermediaries working in PT" ;
    dct:created "2023-12-15"^^xsd:date ;
    dct:modified "2023-12-23"^^xsd:date ;
    dct:publisher ex:nationalauthority_PT ;
    :hasDataIntermediationServiceProvider ex:DISP_Y .
ex:nationalauthority_PT a :DataIntermediationAuthority ;
    dpv:hasName "Data Intermediation Authority of Portugal" ;
    dpv:hasContact "mailto:nationalauthority_PT@email.com" ;
    dpv:hasJurisdiction "PT" .
ex:DISP_Y a :DataCooperative ;
    dpv:hasName "Data Cooperative Y" ; dpv:hasAddress "Lisboa, Portugal" ;
    dct:description "Provider of anonymised geolocation data" ;
    dcat:landingPage <http://cooperativeA.com/> ;
    dct:date "2023-12-23"^^xsd:date .
```

Listing 3: Example of a public register of data intermediation service providers.

Using the existing and the newly created terms, an example of a register of data intermediation service providers can be found in Listing 3. Due to restrictions in the size of this publication, we do not provide an example of a public register of a data altruism organisation, as both types of public registers contain similar information and will have similar semantic representations. The register `ex:publicregistry_DI_PT` will have a complete list of intermediaries operating in Portugal. Beyond the stored metadata regarding the national authority `ex:nationalauthority_PT` and creation dates, the register has already a registered `DataCooperative` company: `ex:DISP_Y`.

DPV's `hasName`, `hasContact` and `hasAddress` and DCAT's `landingPage` can be used to provide information about the providers of data intermediation or data altruism services, while DCT's `description`, `created`, and `publisher`, can be used to describe metadata about the register, including its creation date and its publisher. In addition to these terms that can be reused from existing standards and specifications, to specify a data intermediation service provider (as a subclass of `dpv:LegalEntity`), or one of its types, we added four new classes of entities to our vocabulary to represent these terms, `DataIntermediationServiceProvider`, `DataCooperative`, `DataIntermediationServiceProviderForDataHolder`, and `DataIntermediationServiceProviderForDataSubject`. Moreover, data altruism organisations are modelled as a subclass of `dpv:NonProfitOrganisation`. Information related to the nature of the entity, as specified in Article 11.6(b), to represent the legal status, form, ownership structure, subsidiary and registration number of an entity, is out of the scope of this contribution as it refers to organisational details. However, as a future contribution, upper ontologies such as GIST [17] or Schema.org [18] can be explored, and if necessary extended, to include such concepts.

Furthermore, a `PublicRegister` class was also added to our vocabulary, and its respective subclasses `RegisterOfDataIntermediationServiceProviders` and `RegisterOfDataAltruismOrganisations` to represent public registers of

```
SELECT DISTINCT ?Provider ?Name ?Web WHERE {
    ?Provider a :DataCooperative .
    ?Provider dpv:hasName ?Name .
    ?Provider dcat:landingPage ?Web . }
```

Listing 4: SPARQL query to retrieve data cooperatives.

```
ex:altruism_logs a :RegisterOfDataIAltruismActivity ;
    dct:description "Activity logs of the Data Altruism Organisation A" ;
    dct:created "2023-11-04"^^xsd:date ;
    dct:modified "2023-11-13"^^xsd:date ;
    dct:publisher ex:altruism_A ; dcat:record ex:log_001 .
ex:altruism_A a :DataAltruismOrganisation ;
    dpv:hasName "Data Altruism Organisation A" ;
    dpv:hasAddress "Lisboa, Portugal" ;
    dcat:landingPage <http://example.com/altruism_A> .
ex:log_001 a dcat:CatalogRecord ;
    dct:created "2023-11-13"^^xsd:date ;
    :hasDataUser ex:userZ ; :hasFee "1000€"^^xsd:string ;
    dpv:hasPersonalDataHandling [
        dct:description "Download and reuse anonymised health records to
        ↪ improve healthcare" ;
        dpv:hasProcessing :Download, :Reuse ; dpv:hasDuration 6226453 ;
        dpv:hasPurpose :DataAltruism, :ImproveHealthcare ;
        dpv:hasPersonalData dpv-pd:HealthRecord ;
        dpv:hasTechnicalMeasure dpv:Anonymisation ] .
ex:userZ a :DataUser ; dpv:hasName "Data User Z" ;
    dpv:hasContact "mailto:user_z@email.com" .
```

Listing 5: Example of a register of data altruism activity logs.

data intermediaries and of altruistic organisations, respectively. By having the public register stored in RDF using the identified and developed semantic vocabularies, such register can then be easily queried, using a query language such as SPARQL to automate the retrieval of information regarding data intermediation service providers. An example of a query for data cooperatives is provided in Listing 4, which will return a list of data intermediation service providers that offer the services of data cooperatives, including their names and public websites.

Listing 5 provides an example of a register of data altruism activity, represented through the newly created concept `RegisterOfDataIAltruismActivity`. Activity logs should be associated with the entities using the data and can be recorded using DPV's `hasPersonalDataHandling` to provide information about the processing of data, including its duration, purpose and (personal) data categories.

```
ex:consentForm_001 a :EuropeanDataAltruismConsentForm ;
    dpv:hasIdentifier <http://example.com/consentForm_001> ;
    dpv:hasDataSubject ex:Anne ; dpv:isIndicatedBy ex:Anne ;
    dpv:isIndicatedAtTime "2022-12-14" ;
    dpv:hasPersonalDataHandling [
        dpv:hasPurpose :DataAltruism, :ImproveTransportMobility ;
        dpv:hasLegalBasis dpv-gdpr:A6-1-a ;
        dpv:hasPersonalData dpv-pd:Location ;
        dpv:hasProcessing dpv:Use, dpv:Store ;
        dpv:hasDataController [
            a dpv:DataController, :DataAltruismOrganisation ;
            dpv:hasName "Company A" ] ] .
```

Listing 6: Data altruism form where data subject Anne consents to the usage of their location data for the altruistic purpose of improving mobility.

4.3. Uniform, Machine-readable Data Altruism Form

As already proven by the examples provided in the former two subsections, the identified vocabularies, as well as the one we developed ourselves, can also be used to automate the production of consent forms for data subjects and permission forms for data holders. By relying on such technologies by design, the European data altruism forms will promote interoperability and can be reused throughout the EU. An example of a consent form, using the term, `EuropeanDataAltruismConsentForm`, by a data subject is provided in Listing 6. In this example, we use an altruistic purpose for processing defined in our vocabulary. As such, we define `DataAltruism` as a subclass or `dpv:Purpose` and we specify seven new purposes that can be used in a data altruism setting: `ImproveHealthcare`, `CombatClimateChange`, `ImproveTransportMobility` (used in Listing 6), `ProvideOfficialStatistics`, `ImprovePublicServices`, `ScientificResearch` and `PublicPolicyMaking`. Additional purposes, mentioned throughout the DGA, are also provided in the ad-hoc vocabulary. Similarly, in Listing 7, we provide an example of a permission form of a data holder.

5. Discussion

A complete list of all defined terms is available at *https://w3id.org/dgaterms*, under an open and permissive licence. The analysis of how semantic technologies can be used to operationalise the DGA yields some promising applications, however, a number of hindrances can be identified. Among the advantages, the following ones should be carefully noted:

- Semantic technologies can help forge a common understanding of the provisions in the regulation.
- Machine-readable policies can be effectively represented in RDF, and executed with appropriate reasoners.

```
ex:permissionForm_001 a dpv:Permission ;
    dpv:hasIdentifier <http://example.com/permissionForm_001> ;
    :hasDataHolder ex:dataHolderA ; dpv:isIndicatedBy ex:dataHolderA ;
    dpv:isIndicatedAtTime "2022-12-15" ;
    dpv:hasPersonalDataHandling [
        dpv:hasPurpose :DataAltruism, :ProvideOfficialStatistics ;
        dpv:hasLegalBasis :A2-6 ; dpv:hasData dpv:AnonymisedData ;
        dpv:hasProcessing dpv:Use, dpv:Store ;
        dpv:hasDataController [
            a dpv:DataController, :DataAltruismOrganisation ;
            dpv:hasName "Company A" ] ] .
```

Listing 7: Permission for data altruism where data holder A allows the usage of their anonymised data for the altruistic purpose of providing official statistics.

- Trust technologies certifying a data altruism consent expression provide legal certainty and encourage data reuse, in the very spirit of the DGA.
- Semantic Web technologies excel at publishing policies on the Web, with JSON-LD serializations easily consumed by Web developers. In addition, RDF can effectively act as a bridge between future expression languages.
- Data altruism may be rewarded in non-economic forms, encouraging in turn further data sharing.

However, the following limitations can be identified:

- Most of the conditions declared in the policies will not be able to be automatically enforced and the declarative nature of the policies will hopelessly lead to data misuse.
- The agreement may not be such if no ontology gains hegemonic spread, if it is not sanctioned by a public authority, or if heavy discrepancies prevent reaching a consensus.

6. Conclusions and Future Work

While powerful, the European strategy for data brings many interoperability challenges that need to be surpassed if we are to implement common spaces to share data between individuals, companies and governments. As such, the effort we made in this work on analysing the requirements of the DGA and providing a common semantic model to record the activities of public sector bodies, data intermediation service providers and altruism organisations are a first step towards conquering this interoperability challenge. As proposed, we identified the stakeholders, information and requirements necessary to model the activities of the new data-sharing entities brought by the DGA and provided a semantic vocabulary, and examples of usage of such vocabulary, that can be used to automate the documentation tasks of these new entities. As for future work, the Data Act and Data Spaces proposals should be explored to improve the quality of this work and promote the interoperability envisioned by the common European data spaces.

Acknowledgments

This work was funded partially by the project Knowledge Spaces: Técnicas y herramientas para la gestión de grafos de conocimientos para dar soporte a espacios de datos (Grant PID2020-118274RB-I00, funded by MCIN/AEI/10.13039/501100011033). This research has also been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497 (PROTECT) and by the ADAPT SFI Research Centre funded by Science Foundation Ireland and co-funded under the European Regional Development Fund (ERDF) through Grant#13/RC/2106_P2.

This is ongoing work which has been contributed to the Data Privacy Vocabularies and Controls Community Group (DPVCG) as a proposal to integrate the DGA within its outputs.

References

[1] 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); 2018. Available from: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[2] 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. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066.

[3] 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. Available from: http://data.europa.eu/eli/reg/2022/868/oj/eng.

[4] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space; 2022. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197.

[5] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act); 2022. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN.

[6] de Oliveira Rodrigues CM, de Freitas FLG, Barreiros EFS, de Azevedo RR, de Almeida Filho AT. Legal ontologies over time: A systematic mapping study. Expert Systems with Applications. 2019;130:12-30.

[7] Esteves B, Rodríguez-Doncel V. Analysis of Ontologies and Policy Languages to Represent Information Flows in GDPR. Semantic Web Journal. 2022.

[8] Iannella R, Villata S. ODRL Information Model 2.2. W3C Recommendation. 2018.

[9] Parducci B, Lockhart H, Rissanen E. eXtensible Access Control Markup Language (XACML) Version 3.0 [OASIS Standard]; 2013. Available from: http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html.

[10] Palmirani M, Governatori G, Athan T, Boley H, Paschke A, Wyner A. LegalRuleML Core Specification Version 1.0; 2021. Available from: https://docs.oasis-open.org/legalruleml/legalruleml-core-spec/v1.0/os/legalruleml-core-spec-v1.0-os.html.

[11] Esteves B, Pandit HJ, Rodríguez-Doncel V. ODRL Profile for Expressing Consent through Granular Access Control Policies in Solid. In: 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS PW); 2021. p. 298-306. ISSN: 2768-0657.

[12] Agarwal S, Steyskal S, Antunovic F, Kirrane S. Legislative compliance assessment: framework, model and GDPR instantiation. In: Annual Privacy Forum. Springer; 2018. p. 131-49.

[13] Albertoni R, Browning D, Cox S, Beltran AG, Perego A, Winstanley P. Data Catalog Vocabulary (DCAT) - Version 2; 2020. Available from: https://www.w3.org/TR/vocab-dcat-2/.
[14] DCMI Metadata Terms; 2008. Available from: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.
[15] Pandit HJ, Polleres A, Bos B, Brennan R, Bruegger B, Ekaputra FJ, et al. Creating a Vocabulary for Data Privacy: The First-Year Report of Data Privacy Vocabularies and Controls Community Group (DPVCG). In: Panetto H, Debruyne C, Hepp M, Lewis D, Ardagna CA, Meersman R, editors. On the Move to Meaningful Internet Systems: OTM 2019 Conferences. vol. 11877. Springer International Publishing; 2019. p. 714-30. Available from: http://link.springer.com/10.1007/978-3-030-33246-4_44.
[16] Ryan P, Brennan R, Pandit HJ. DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue Based on GDPR. Information. 2022;13(5). Available from: https://www.mdpi.com/2078-2489/13/5/244.
[17] gist ontology v12.0.0; 2023. Available from: https://w3id.org/semanticarts/ontology/gistCore.
[18] Schema.org v22.0; 2023. Available from: https://schema.org/.