GDPRtEXT – GDPR as a Linked Data Resource

4 authors, including:

Kaniz Fatema
Trinity College Dublin
26 PUBLICATIONS   **227** CITATIONS

SEE PROFILE

Declan O'Sullivan
Trinity College Dublin
187 PUBLICATIONS   **1,116** CITATIONS

SEE PROFILE

David Lewis
Trinity College Dublin
195 PUBLICATIONS   **1,007** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Mobile Health (mHealth) & ICT4D in Sierra Leone View project

EC FP7 TAS3 View project

GDPRtEXT - GDPR as a Linked Data Resource

Harshvardhan J. Pandit[1], Kaniz Fatema[2], Declan O'Sullivan[1], and Dave Lewis[1]

[1] ADAPT Centre, Trinity College Dublin, Dublin, Ireland,
`firstname.lastname@adaptcentre.ie`
[2] University of Derby, Derby, United Kingdom,
`k.fatema@derby.ac.uk`

Abstract. The General Data Protection Regulation (GDPR) is the new European data protection law whose compliance affects organisations in several aspects related to the use of consent and personal data. With emerging research and innovation in data management solutions claiming assistance with various provisions of the GDPR, the task of comparing the degree and scope of such solutions is a challenge without a way to consolidate them. With GDPR as a linked data resource, it is possible to link together information and approaches addressing specific articles and thereby compare them. Organisations can take advantage of this by linking queries and results directly to the relevant text, thereby making it possible to record and measure their solutions for compliance towards specific obligations. GDPR text extensions (GDPRtEXT) uses the European Legislation Identifier (ELI) ontology published by the European Publications Office for exposing the GDPR as linked data. The dataset is published using DCAT and includes an online webpage with HTML id attributes for each article and its subpoints. A SKOS vocabulary is provided that links concepts with the relevant text in GDPR. To demonstrate how related legislations can be linked to highlight changes between them for reusing existing approaches, we provide a mapping from Data Protection Directive (DPD), which was the previous data protection law, to GDPR showing the nature of changes between the two legislations. We also discuss in brief the existing corpora of research that can benefit from the adoption of this resource.

Keywords: GDPR, DPD, linked resource, regulatory technology, legal compliance, SKOS, DCAT, e-governance

1 Introduction

The General Data Protection Regulation (GDPR) is the new European data protection legislation that enters into force on 25th May 2018. It is an important legislation in terms of changes to the organisational measures required for compliance. In particular, GDPR focuses on the use of consent and personal data and provides the data subject with several rights. These new changes have spurred innovation within the community - both in the industry as well as in academia [3,5,6,21] - that targets compliance with the various obligations of the

GDPR. While such solutions claim assistance to the various provisions of the GDPR, comparing and collecting such solutions is a difficult undertaking due to the inability of consolidating them in an efficient manner.

A lack of method to address specific sections of the GDPR prevents related methods from being linked together in a uniform and consistent manner. This creates challenges in the comparison of their degree and scopes, especially regarding metrics of solutions that claim to assist in compliance. This hampers progress as it limits the findability of information related to a particular resource - in this case a concept or a point within the GDPR text.

Since particular points and concepts within a legislation are referred to using a form of numbering system, for example - Article 5(1) refers to the first point within Article 5, this can be used to create URIs that refer to each distinct resource within the GDPR. Using these URIs it is possible to define information targetting or referring to specific points or concepts. In this manner, legal resources and concepts within GDPR can be linked to each other as well to other resources.

Organisations and researchers gain the advantage through this of linking queries and results directly to the relevant text of GDPR, thereby making it possible to record and measure solutions for compliance towards specific obligations. The same approach can also be used in linking related legislations to the GDPR. Such links between associated information offer the benefit of being machine-readable and therefore can be queried and acted upon in an automated manner.

Through this paper and its associated resources, we aim to alleviate the above discussed limitation of legal text and demonstrate the application of FAIR principles to link together information associated with the GDPR and its compliance. To this effect, we present GDPR text extensions (GDPRtEXT), which consists of GDPR text defined as a DCAT catalog containing the official text as well as created RDF resources as datasets, and a SKOS ontology defining concepts related to GDPR. An application of GDPRtEXT is presented in the form of linking GDPR with its predecessor - the Data Protection Directive (DPD) - to facilitate the adoption of existing solutions in addressing GDPR. All resources described are documented and available online[3] under the CC-by-4.0 license[4].

The rest of the paper is structured in the following manner - Section 2 discusses the motivation and creation of the GDPR linked data resource and GDPRtEXT ontology. Section 3 contains the mapping between DPD and GDPR obligations. Section 4 describes the related work. Section 5 discusses potential applications and benefits to the community. Section 6 concludes the paper with a discussion of future work.

---

[3] https://openscience.adaptcentre.ie/projects/GDPRtEXT/
[4] https://creativecommons.org/licenses/by/4.0/

2   Creation of resource

2.1   Motivation

In our previous work involving the GDPR [10, 13, 15, 19], we faced challenges in referring to GDPR concepts and obligations as well as in consolidating information referring to particular obligations. In light of this, we initially used the solution of assigning a permanent URI to each point in the text using HTML fragments so as to refer to them as linkable resources. Later, we extended this approach to create a RDF dataset that could be used in other works and queried using SPARQL. In light of its usefulness and benefits to the community, we removed the project specific artefacts to provide it as an open resource.

2.2   Scope

The scope of this work is explicitly limited only to the GDPR. Although other legislations are explicitly mentioned within GDPR, they are not addressed in this resource. At the time of creation of this resource, the GDPR is considered to be final, i.e. published as a complete version as opposed to being a draft. Updates (related legislations or future laws) to the GDPR are published as separate legal documents that can be similarly annotated and added to this dataset, whereby they can then be published together through the DCAT catalog. Where updating the ontology is necessary, it should be approached as incorporating new legal laws into existing knowledge bases.

2.3   GDPR as linked data

The GDPR document [7] is structured as three types of statements in the following order - 173 Recitals, 99 Articles, and 21 Citations. The articles are structured within Chapters (numbered from I to X) and Sections, where each Chapter has zero or more sections. An article may contain several points which may have sub-points that may or may not be numbered. Citations appear at the end of the document and are numbered by their order of reference within the text.

To define legal resources, we use the European Legislation Identifier (ELI)[5] ontology published by the European Publications Office. *ELI* allows identification of legislation through URI templates at the European, national and regional levels based on a defined set of related attributes or terms. It provides properties describing each legislative act as a set of metadata definitions and its expression in a formal ontology. Serialisation of *ELI* metadata elements provides integration of metadata into the legislative websites using RDFa.

We use the latest iteration of the ELI OWL ontology (v1.1, released 2016-09-19) to define the classes *Chapter, Section, Article, Point, SubPoint, Recital,* and *Citation* as subclasses of *eli:LegalResourceSubdivision* (*LRS*), which is itself a subclass of *eli:LegalResource* (*LR*). *LR* is used to define resources at the

---

[5] http://publications.europa.eu/mdr/eli/index.html

document level, whereas *LRS* is used to define the resources contained within a *LR*. *ELI* defines the properties *has_part* and its inverse *is_part_of* to connect two *LR*s. GDPRtEXT extends these with additional properties to connect the various chapters, sections, articles, and points with each other. As we extend *ELI*, GDPRtEXT can be used in a manner that is compatible with the intended use of *ELI*, including its use in websites as RDFa.

The official text of the GDPR available online[6] as a HTML page was used to extract its text using Javascript[7]. This was then exported as a JSON document with metadata such as resource type (article, chapter, etc.) and numbering (roman, arabic, or unnumbered). The RDF dataset was generated through a python script using rdflib[8] that created the triples iteratively using the ELI extension described above.

DCAT[9] and VOID[10] were used to define the GDPR catalog containing three datasets as described in Table 1. The first dataset describes official documents published by the European Publication Office as distributions with their canonical urls. The second dataset describes distributions containing a copy of the GDPR text hosted by the ADAPT Centre and contains HTML and JSON representations of the GDPR. The HTML version uses the id-attribute to define fragment identifiers for each legal resource within the text. This makes it possible to directly refer (or navigate) to a particular article, point, or sub-point within the text. This is similar to the RDF distribution of the GDPR text where each legal resource has a distinct URI. The third dataset contains RDF serialisations of the GDPR in the Turtle, N-Triples, N3, and JSON-LD formats.

The catalog describes a SPARQL endpoint hosted using the OpenLink Virtuoso[11] triple-store for exposing the RDF dataset containing the GDPR text along with an online front-end using Pubby[12] as shown in Fig. 1. All URIs for the dataset and distributions use the permanent-url (purl.org) scheme to refer to resources.

Table 1: Datasets and Distributions in the GDPRtEXT DCAT Catalog

Dataset	Distributions	Comment
canonical_dataset	HTML, PDF, XML	official distributions
textid_dataset	HTML, JSON, text	with IDs
annotated_dataset	XML, N3, NT, Turtle, JSON-LD	RDF datasets

---

[6] http://eur-lex.europa.eu/eli/reg/2016/679/oj

[7] https://opengogs.adaptcentre.ie/harsh/GDPRtEXT/

[8] https://rdflib.readthedocs.io/en/stable/

[9] https://www.w3.org/TR/vocab-dcat/

[10] https://www.w3.org/TR/void/

[11] https://virtuoso.openlinksw.com/

[12] http://wifo5-03.informatik.uni-mannheim.de/pubby/

article12-3 at **GDPRtEXT**

https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3

Property	Value
eli:description	▪ The controller shall provide information on action taken on a request under Articles 15 to 22 to the data subject without undue delay and in any event within one month of receipt of the request. That period may be extended by two further months where necessary, taking into account the complexity and number of the requests. The controller shall inform the data subject of any such extension within one month of receipt of the request, together with the reasons for the delay. Where the data subject makes the request by electronic form means, the information shall be provided by electronic means where possible, unless otherwise requested by the data subject. (xsd:string)
is gdprtext:hasPoint of	▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12>
gdprtext:isPartOfArticle	▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12>
gdprtext:isPartOfChapter	▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapterIII>
gdprtext:isPartOfSection	▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapterIII-1>
eli:is_part_of	▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/GDPR> ▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12> ▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapterIII> ▪ <https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapterIII-1>
eli:number	▪ 3 (xsd:string)
eli:title_alternative	▪ Article12(3) (xsd:string)
rdf:type	▪ eli:LegalResourceSubdivision ▪ gdprtext:Point

Metadata

Anon_0		
rdf:type	prv:DataItem	
rdf:type	<http://www.w3.org/2004/03/trix/rdfg-1/Graph>	
foaf:primaryTopic	<https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3>	
foaf:topic	Anon_0	
ir:realizes	<https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/data/article12-3>	
prv:createdBy	Anon_1 (more)	

expand all

This page shows information obtained from the SPARQL endpoint at https://openscience.adaptcentre.ie/sparql.
As N3 | As RDF/XML | Browse in Disco | Browse in Tabulator | Browse in OpenLink Browser

Fig. 1: Article 12(3) in GDPRtEXT as RDF displayed using Pubby

2.4 GDPRtEXT Ontology

The aim of the GDPRtEXT ontology is to provide a way to refer to the concepts and terms expressed within the GDPR. The ontology does not aim to provide an interpretation of compliance obligations using methods such as inference. It is intended to be an open resource for addressing modeling solutions for interoperability related to GDPR compliance.

The development of the ontology followed the seminal guide "*Ontology Development 101*" [18]. The development started by deciding on the scope which led to the aim of providing a way to refer to the various concepts defined in the GDPR. The SKOS[13] vocabulary was chosen to provide descriptions of GDPR concepts as it is a W3C Recommendation for defining terms.

The defined terms were linked to the relevant points in the GDPR text using the URI scheme described in the previous section through the *rdfs:isDefinedBy* property. Additional links between the terms were created using the additionally created annotation property termed '*involves*'. The development of the ontology was done using the Protégé[14] v5.2.0 ontology development environment.

Following the preliminary collection of terms, additional terms were added to the ontology in order to create a hierarchy of concepts based on the requirements of the GDPR. For example, identified data types such as personal data and anonymous data are defined as subclasses of the common term Data. This allows a representation of the encapsulation of concepts and provides a way to refer to the term in its abstract form. Note that the ontology does not use the

---

[13] https://www.w3.org/TR/skos-reference/
[14] https://protege.stanford.edu/

broader/narrower concepts present in SKOS, but uses subclasses as a simple means of collecting related concepts in a hierarchical manner.

Similar undertakings were taken for concepts related to the various obligations and entities. Additional concepts mentioned or inferred from the GDPR text such as *Entity, Principle,* and *Unlawful Processing* were also added but not linked with the GDPR text. The collection was undertaken in order to define all necessary terms required in the documentation of compliance. At the time of writing this paper, GDPRtEXT contains 200+ classes defined using the SKOS vocabulary based on terms occurring within the text of the GDPR.

The vocabulary, as visualised in Figures. 3(a) and 3(b), defines terms broadly using classes for *Activity, Compliance, Consent, Data, Entity, Exclusion & Exceptions, Obligations, Principles, Rights*, and *Seals / Certifications*. Each of these contains several terms in a hierarchy of subclasses that further defines the concept and its scope. Certain terms come under more than one concept, and are therefore defined as subclasses of multiple concepts. For example, the term *ProvideCopyOfData*, which refers to the right of data portability where a copy of the personal data is provided to the data subject, is a concept under *DataActivity* as well as *RightOfDataPortability*.

Various resources other than the official text of the GDPR were used to understand the terms and concepts. We used the resources made available by official sources such as Data Protection Commissioner of Ireland[15] and Information Commissioner's Office of UK[16] as well as industry-based sources such as Nymity's GDPR handbook[17].

2.5 Documentation

The documentation of the ontology presented a challenge due to the large number of terms. To automate part of the process, we used the Wizard for Documenting Ontologies (Widoco)[18] [12], which uses LODE[19] [20] for generating ontology documentation and WebVOWL[20] [16] for its visualisation. The resulting HTML documentation is available online[21] as depicted in Fig. 2. The documentation groups terms based on core concepts such as *Consent, Data, Activity,* and *Compliance* to make the documentation more readable. Each term in the ontology was defined with the metadata as specified by WIDOCO's best practices document[22] to generate its comprehensive documentation.

The documentation contains two example use-cases of GDPRtEXT. The first shows its use in compliance reporting by linking the relevant tests and results

---

[15] https://www.dataprotection.ie/

[16] https://ico.org.uk/

[17] https://www.nymity.com/

[18] https://doi.org/10.5281/zenodo.591294

[19] http://www.essepuntato.it/lode

[20] http://vowl.visualdataweb.org/webvowl.html

[21] http://purl.org/adaptcentre/openscience/ontologies/GDPRtEXT/docs

[22] https://dgarijo.github.io/Widoco/doc/bestPractices/index-en.html

Fig. 2: Documentation for GDPRtEXT Ontology generated using Widoco

with the GDPR articles they represent. This is shown using the EARL[23] vocabulary for expressing test results. The second example shows the linking of GDPR obligations with the previous data protection law, which is described in more detail in the next section. We have also used GDPRtEXT to link provenance terms with their definitions in relevant GDPR articles [19].

3 Linking DPD obligations with GDPR

The Data Protection Directive (DPD) is the previous data protection legislation which was adopted in 1995 and is superseded by the GDPR. As a large number of solutions and approaches already exist that address compliance with the DPD,

---

[23] https://www.w3.org/TR/EARL10-Schema/

it would be beneficial to look into reusing these existing solutions for GDPR. To that end, we provide a mapping from DPD obligations to GDPR obligations containing annotations that describe the nature of change between the two. The annotation also describes changes required in our previous work in using XACML rules to model DPD obligations [8, 9].

The annotations are available online in the form of a HTML table, a CSV file, and a RDF dataset. Each row (HTML table, CSV) has 5 columns that contain a reference from a point in DPD to its corresponding point in the GDPR, the nature of change between the two, whether the corresponding XACML rule needs to be extended, and a description comment. The nature of change is represented as one of the following - *same, reduced, slightly changed, completely changed*, and *extended*. For XACML rules, the notation '*N/A*' is used in the case where there were no XACML rules for DPD but the corresponding point in GDPR has changed. The value '*No*' is used where there is no change in the GDPR obligation or the existing XACML rule is sufficient to handle the change, whereas '*Yes*' is used to indicate a change required in the XACML rule to handle the obligation.

To model the annotations as a RDF resource using GDPRtEXT, we created a linked data version of the DPD which assigned URIs for every resource in the legislation similar to the GDPR linked data resource. The annotations are represented as instances of the class *DPDToGDPRAnnotation*. The property *resourceInDPD* is used to refer to the particular resource within DPD through its URI. Similarly, the property *resourceInGDPR* is used to refer to the corresponding resource in GDPR. The nature of change is defined using the property *hasChange* whose value is an instances of the class *ChangeInObligation*, with defined instances for *Extended, Same, Reduced, CompletelyChanged*, and *SlightlyChanged*. Similarly, the change in the XACML rules is defined as a property whose values are one of *Yes, No*, and *N/A* defined as instances of the class *ChangeInXACMLRule*. Comments are defined using the *rdfs:comment* property. An example of such linking between the DPD and the GDPR obligations can be seen below in Listing. 3-1.

```
@prefix gdpr:
<http://purl.org/adaptcentre/openscience/resources/GDPRtEXT#>.
@prefix dpd:
<http://purl.org/adaptcentre/openscience/resources/DPD#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.

dpd:mappingrule6 a dpd:DPDToGDPR_Annotation ;
    dpd:hasChange dpd:ChangeExtended ;
    dpd:hasXACMLChange dpd:XACMLNoChange ;
    dpd:resourceInDPD dpd:Article7-a ;
    dpd:resourceInGDPR gdpr:Article6-1-a ;
    rdfs:comment "added consent given to ..." .
```

Listing 3-1: Linking obligation between DPD and GDPR related to processing of personal data and consent

4 Related Work

The creation of the GDPRtEXT vocabulary was influenced by the work done by Bartolini and Muthuri in the creation of their data protection ontology based on the GDPR [2,3]. The purpose of their ontology was to model the requirements and duties of the controller to be compliant with the GDPR. To this end, their ontology defines several properties and 'rules' as well as equivalency of classes for compliance concepts. By contrast, GDPRtEXT does not contain any inference but provides a definition of useful terms using SKOS. This demonstrates the difference of scope between the two approaches and the varying complexities of the work involved.

Bartolini's ontology is based on a draft of the GDPR, and as such, contains some minor inconsistencies with the published (final) version. Additionally, the authors term the ontology to be a preliminary work in its early stages. While their use of classes and inferences can prove to be useful in determining compliance, GDPRtEXT takes a more generic approach of trying to provide a way to refer to specific concepts within the GDPR, making it possible to combine the approaches.

Expressing legal obligations as a set of rules that are machine-processable allows automated systems to be modelled for compliance. One such recent approach models GDPR obligations as machine-readable rules using ODRL[24] [1]. The model is based on using 31 articles to create a graph containing 313 nodes and 810 defined edges to express relations between obligations. The aim of this work is to create a compliance checking tool for GDPR. Using the relations, a hierarchy of obligations is expressed as dependencies between them. The obligations are categorised into core obligations and those depending on them, called as sub-obligations. A total of 48 obligations have been identified in this manner along with 105 dependant sub-obligations. The published paper contains a diagram of the relations between the articles, but does not provide a way to access the data regarding obligations.

The approach of linking GDPR articles together based on obligations is similar in principle to the approach described in this paper, and shows the benefits of creating a linked data version of the GDPR. Using a collection of concepts as provided by GDPRtEXT, the defined obligations can be linked to their respective GDPR text, which allows consolidation of information in the usage and documentation.

5 Applications & Benefit to Community

Example applications of legal ontologies in the areas of information and knowledge systems and compliance solutions are described by [11,14,17]. The publication describing the data protection ontology [3] described in section 4 provides applications in the domains of information retrieval, transition from DPD to

---

[24] https://www.w3.org/TR/odrl-model/

GDPR, automated classification and summarisation, question answering, decision support and decision making, and autonomous agent systems.

The advantages of linked data have been documented before [4, 22], and are equally applicable to legislations such as DPD and GDPR. Here, we discuss the benefits provided by GDPRtEXT in consolidating related research with a focus on compliance solutions.

We take the example of the work described in the previous section of modelling the GDPR obligations as machine-processable ODRL rules intended for a compliance based system. The work describes dependencies between the GDPR obligations, which extends to the same dependencies between the ODRL rules modelled from these obligations. This results in a hierarchical model of obligations where compliance of a rule depends on compliance of all of its sub-rules.

Consider an organisation or a researcher that wishes to use these rules to augment the compliance system they have in place. It is more than likely that both solutions in this case refer to the GDPR via text or some internal form of reference. In order to consolidate both approaches such that they are compatible with each other, modifications will have to be made to either or both, the degree and scope of which varies with the complexity of the system. Additionally, comparing the compliance for a particular obligation from different systems again might involve the use of a text-based or internal form of reference which is isolated to that particular system and might not be machine-readable.

If instead, a common form existed that could be used to refer to the required concepts or obligations, then the task of consolidation of information becomes significantly simpler. In the above example use-case, the ODRL rules and other systems can refer to specific GDPR obligations through the same set of URIs as provided by GDPRtEXT. This makes it possible to link related information in a consistent manner between systems that act on the same obligations to provide consolidated compliance results.

This use-case can be extended to communications between entities such as data subjects, data controllers, data processors, and data protection authorities. The nature of information exchanged between these entities becomes increasingly complex due to the involvement of multiple systems that interact with different entities. Consolidating compliance related information between these entities can benefit from adoption of a common baseline for describing the GDPR such as that provided by GDPRtEXT.

GDPRtEXT relates to and exposes GDPR which is an important legal document. Therefore, ensuring the resource stays alive and available is of essence with respect to its use by the community. Currently, the resource is hosted by Trinity College Dublin on its VM cluster which are managed by a dedicated team of system administrators. Additionally, since GDPRtEXT uses permanent urls (via `purl.org`) for resolving the actual URL of the resource, it is possible to move the resource to a different location without changing how it is exposed to the world. While we plan to continue maintaining the resource, this provides a way to move the resource to another maintainer or a community hosting service such as Github if the need arises in the future. A copy of the resource is mir-

rored on hosting sites such as Datahub and Zenodo, providing alternate means of access should the need arise.

6   Conclusion & Future Work

In this paper, we presented GDPR text extensions (GDPRtEXT), which uses the European Legislation Identifier (ELI) ontology published by the European Publications Office for exposing the GDPR as linked data. The resulting dataset is published using DCAT and VOID vocabularies with distributions for official publication, an annotated online copy of the GDPR that uses fragment identifiers to refer to individual points, and its serialisations in several RDF formats. The dataset is published using the CC-by-4.0 license and is available with a DOI at Zenodo[25] and at Datahub[26]. We have submitted GDPRtEXT to *data.gov.ie*[27] as a suggested dataset which would allow it to be indexed by the EU Open Data Portal[28]. The dataset contains a SPARQL endpoint with a front-end using Pubby. GDPRtEXT also provides a SKOS vocabulary for defining terms and concepts in GDPR. An example of linking related legislations was provided using Data Protection Directive (DPD) that reused existing work in modelling obligations using XACML. The applications and benefits of GDPRtEXT to the community were also discussed.

GDPRtEXT is an ongoing effort and we actively seek suggestions as well as support and guidance from the community[29]. In terms of future work, we plan to work on the documentation so as to make it more accessible to the community outside of the semantic web domain. To this end, we are looking into conceptualising a model for the documentation and dissemination of GDPR metadata related to consent, provenance, data sharing, and compliance.

The online webpage displaying the linked data version of GDPR currently only displays its text. Using Javascript along with Web Annotation Data Model[30], it is possible to display additional information associated with particular articles or points. This can be used to create a rich interface for the GDPR that filters or displays information in an interactive manner. Examples of annotations displayed alongside GDPR are compliance status reports and legal notes showing information related to fulfilment of a particular obligation.

The principle of linked data works only when there are links to both (or all) involved resources. Therefore, we plan on using GDPRtEXT with GDPR related work such as European Privacy Seal (EuroPriSe)[31], which is an organisation that provides certifications and seals for GDPR compliance. The criteria used by EuroPriSe is based on translating the requirements of GDPR into questions that

---

[25] https://doi.org/10.5281/zenodo.1146351

[26] https://datahub.ckan.io/dataset/gdprtext

[27] https://data.gov.ie/dataset/suggest/a5f365f6-719f-4eab-9bfe-4d494aeb31ee

[28] http://data.europa.eu/euodp/

[29] https://opengogs.adaptcentre.ie/harsh/GDPRtEXT/issues

[30] https://www.w3.org/TR/annotation-model/

[31] https://www.european-privacy-seal.eu

can be answered in the context of an audit or a certification. These questions, therefore, are based on obligations specified by the GDPR, and their answers determine compliance to these obligations. This is an area where GDPRtEXT can be used to link the related information in EuroPriSe's certifications and compare them to similar research undertaken in the domain of GDPR compliance.

Acknowledgement

References

1. Agarwal, S., Kirrane, S., Scharf, J.: Modelling the general data protection regulation. In: Internationales Rechtsinformatik Symposion (IRIS 2017) (2017)
2. Bartolini, C., Muthuri, R.: Reconciling data protection rights and obligations: An ontology of the forthcoming eu regulation (2015)
3. Bartolini, C., Muthuri, R., Cristiana, S.: Using ontologies to model data protection requirements in workflows. In: Ninth International Workshop on Juris-informatics (JURISIN 2015). http://orbilu.uni.lu/handle/10993/22383 (2015)
4. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts pp. 205–227 (2009)
5. Bonatti, P., Kirrane, S., Polleres, A., Wenning, R.: Transparent personal data processing: The road ahead. In: International Conference on Computer Safety, Reliability, and Security. pp. 337–349. Springer, Cham (2017)
6. Chassang, G.: The impact of the eu general data protection regulation on scientific research. ecancermedicalscience 11 (2017)
7. 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). Official Journal of the European Union L119, 1–88 (May 2016), `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC`
8. Fatema, K., Chadwick, D.W., Van Alsenoy, B.: Extracting access control and conflict resolution policies from european data protection law. In: IFIP PrimeLife International Summer School on Privacy and Identity Management for Life. pp. 59–72. Springer (2011)
9. Fatema, K., Debruyne, C., Lewis, D., OSullivan, D., Morrison, J.P., Mazed, A.A.: A semi-automated methodology for extracting access control rules from the european data protection directive. In: Security and Privacy Workshops (SPW), 2016 IEEE. pp. 25–32. IEEE (2016)
10. Fatema, K., Hadziselimovic, E., Pandit, H.J., Debruyne, C., Lewis, D., O'Sullivan, D.: Compliance through informed consent: Semantic based consent permission and data management model. In: 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2017). No. 1951 in CEUR Workshop Proceedings (2017), `http://ceur-ws.org/Vol-1951/#paper-05`

11. Gangemi, A., Prisco, A., Sagri, M.T., Steve, G., Tiscornia, D.: Some ontological tools to support legal regulatory compliance, with a case study. In: OTM Workshops. pp. 607–620. Springer (2003)
12. Garijo, D.: Widoco: a wizard for documenting ontologies. In: International Semantic Web Conference. pp. 94–102. Springer (2017)
13. Hadziselimovic, E., Fatema, K., Pandit, H.J., Lewis, D.: Linked data contracts to support data protection and data ethics in the sharing of scientific data. In: Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci). pp. 55–62. No. 1931 in CEUR Workshop Proceedings (2017), `http://ceur-ws.org/Vol-1931/#paper-08`
14. Hoekstra, R., Breuker, J., Di Bello, M., Boer, A., et al.: The lkif core ontology of basic legal concepts. LOAIT 321, 43–63 (2007)
15. Lewis, D., Moorkens, J., Fatema, K.: Integrating the management of personal data protection and open science with research ethics. In: Ethics in NLP Workshop, EACL. pp. 60–65 (2017)
16. Lohmann, S., Link, V., Marbach, E., Negru, S.: Webvowl: Web-based visualization of ontologies. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 154–158. Springer (2014)
17. Mommers, L.: A knowledge-based ontology of the legal domain. In: Second International Workshop on Legal Ontologies, JURIX (2001)
18. Noy, N.F., McGuinness, D.L., et al.: Ontology development 101: A guide to creating your first ontology (2001)
19. Pandit, H.J., Lewis, D.: Modelling provenance for gdpr compliance using linked open data vocabularies. In: 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2017). No. 1951 in CEUR Workshop Proceedings (2017), `http://ceur-ws.org/Vol-1951/#paper-06`
20. Peroni, S., Shotton, D., Vitali, F.: The live owl documentation environment: a tool for the automatic generation of ontology documentation. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 398–412. Springer (2012)
21. Steyskal, S., Kirrane, S.: If you can't enforce it, contract it: Enforceability in policy-driven (linked) data markets. In: SEMANTiCS (Posters & Demos). pp. 63–66 (2015)
22. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al.: The fair guiding principles for scientific data management and stewardship. Scientific data 3, 160018 (2016)
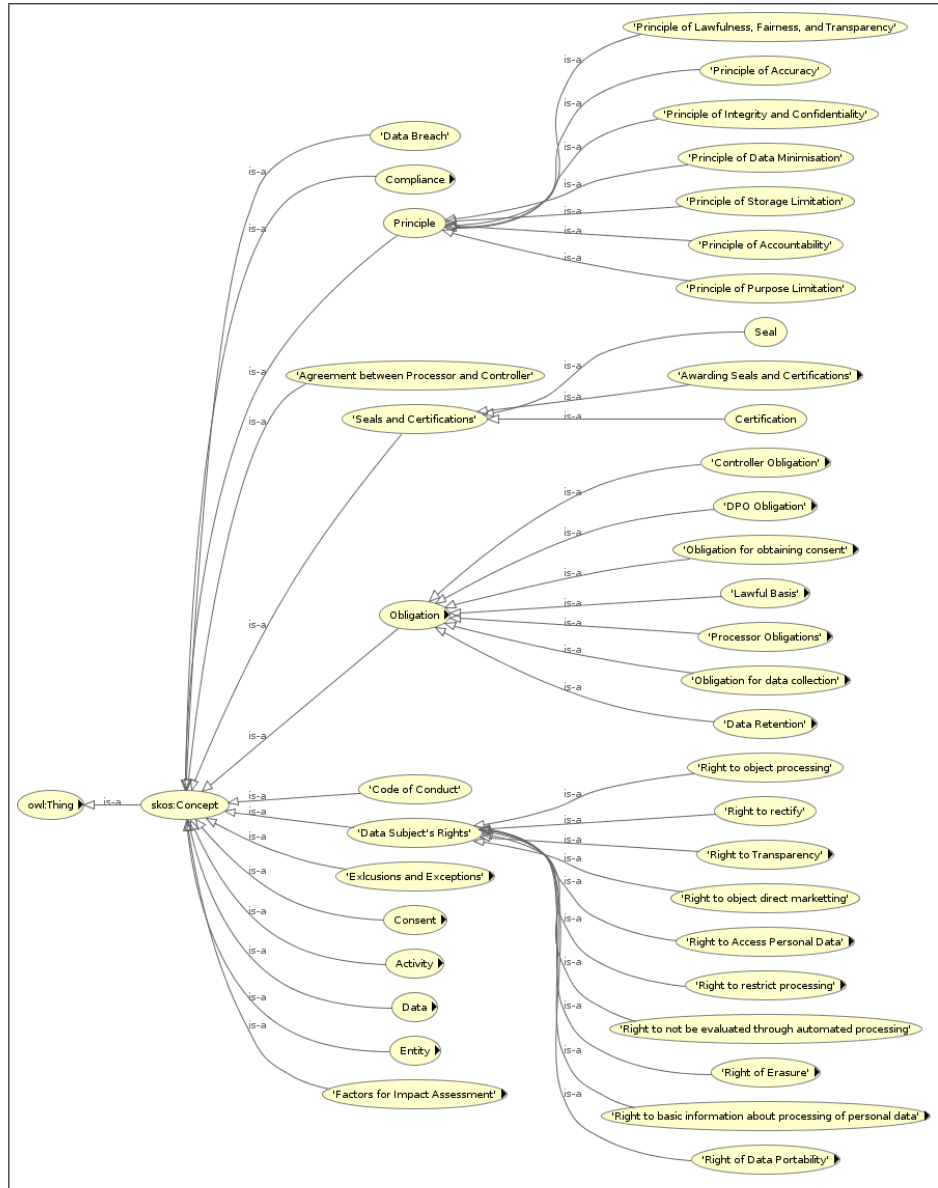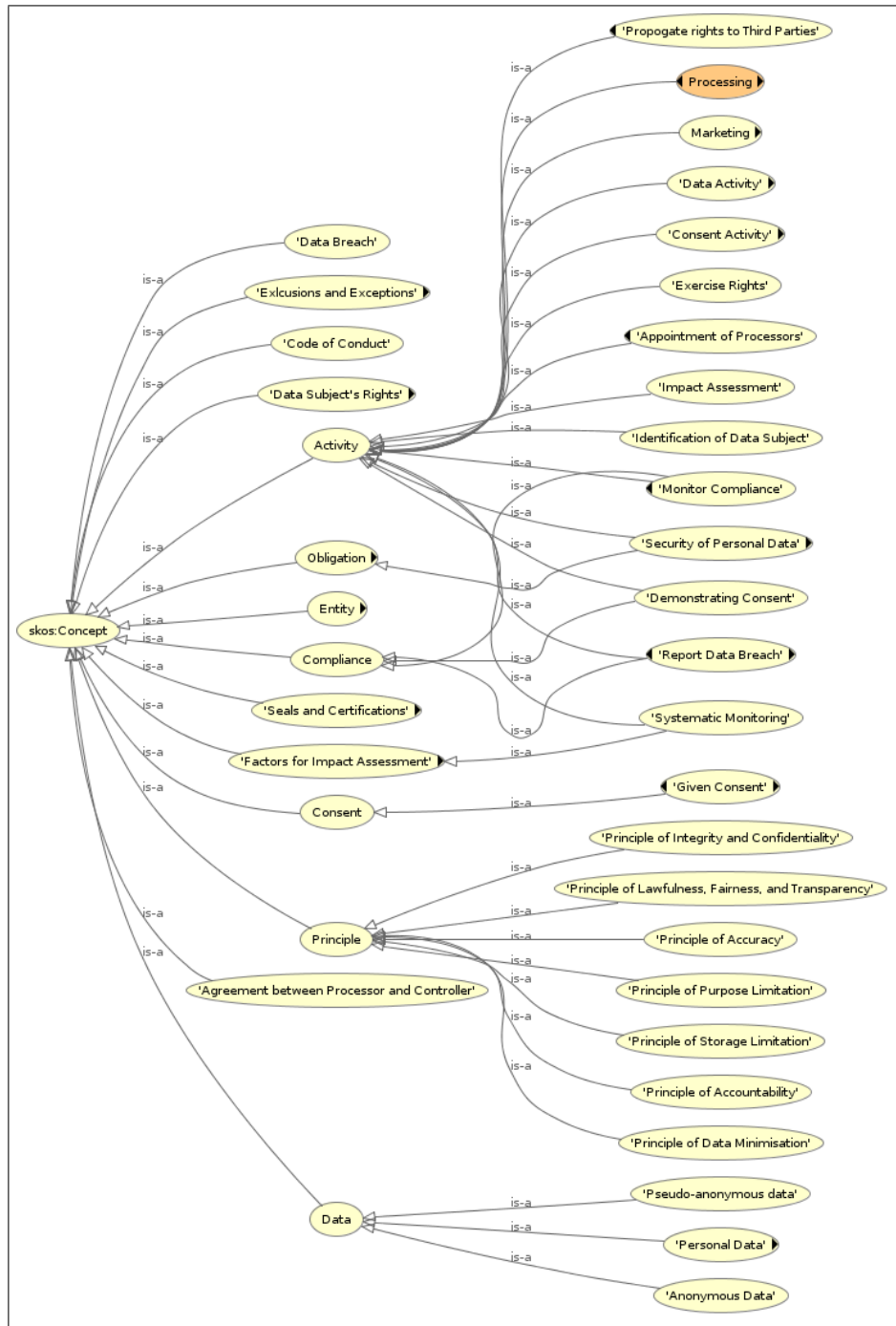
Fig. 3: (a) Terms defined under GDPRtEXT using SKOS

Fig. 3: (b) Terms defined under GDPRtEXT using SKOS