

GConsent - A Consent Ontology based on the GDPR

Harshvardhan J. Pandit, Christophe Debruyne, Declan O’Sullivan, and Dave Lewis

ADAPT Centre, Trinity College Dublin, Dublin, Ireland
{pandith|debruync|declan.osullivan|dave.lewis}@tcd.ie

Abstract. Consent is an important legal basis for the processing of personal data under the General Data Protection Regulation (GDPR), which is the current European data protection law. GDPR provides constraints and obligations on the validity of consent, and provides data subjects with the right to withdraw their consent at any time. Determining and demonstrating compliance to these obligations require information on how the consent was obtained, used, and changed over time. Existing work demonstrates feasibility of semantic web technologies in modelling information and determining compliance for GDPR. Although these address consent, they currently do not model all the information associated with it. In this paper, we address this by first presenting our analysis of information associated with consent under the GDPR. We then present GConsent, an OWL2-DL ontology for representation of consent and its associated information such as provenance. The paper presents the methodology used in the creation and validation of the ontology as well as an example use-case demonstrating its applicability. The ontology and this paper can be accessed online at <https://w3id.org/GConsent>.

Keywords: consent · GDPR · regulatory compliance · OWL2-DL ontology

1 Introduction

The General Data Protection Regulation (GDPR) [19] is the current European data protection law, which affects any service or organisation that uses personal data, and uses large fines to deter non-compliance. Consent is one of the legal basis for processing of personal data under the GDPR (Rec.40, Art.6)¹, and is considered valid only when it is freely given, specific, informed, and unambiguous (Rec.32, Art.2-11); and in the case of minors should be given by their legal guardian (Art.8). GDPR also provides rights regarding changing and withdrawing consent at any time (Art.7-3). To demonstrate compliance with these conditions and obligations of the GDPR, Data Controllers, which are the organisations responsible for deciding how personal data is collected and processed

¹ This is a form of legal notation to denote Recitals (Rec) or Articles (Art) in legal text. These are hyperlinked to where they occur in GDPR using GDPRtEXT [14]

and therefore the ones responsible for obtaining consent when needed, should maintain a demonstrable proof of the given consent [20] by collecting and storing information on how consent was collected, used, and any changes made to it [11] over time.

The information regarding consent and compliance needs to be maintained and shared by multiple parties - data subject, controller, processor, and authorities - which requires its representation to be interoperable between them. Additionally, querying of information is required to comply with requests by data subjects and authorities. Semantic Web technologies are ideal for representing this information because of the flexibility provided for expressing concepts and relationships in an open, interoperable and queryable manner based on standards. Existing work [1,5,8,13,15,16,14] has demonstrated the feasibility of using semantic web technologies for representing and querying metadata for assisting with the GDPR compliance process.

The focus of existing work in terms of consent is mostly on the ‘given’ aspect of consent i.e. consent provided by the data subject. There is a lack of work regarding representing other aspects, or states, of consent such as ‘not given’ or ‘refused’ or ‘withdrawn’ which cannot be modelled in the same manner as ‘given consent’. There is also a lack of modelling representations for events such as delegation or associations with third parties regarding consent which have an effect on its validity regarding compliance.

In this paper, we present our analysis of information associated with consent under the GDPR. We present this through a methodology that creates possible use-cases and scenarios to determine information required for representing consent with a view towards GDPR compliance. We then present the resulting modelling of an ontology in the form of GConsent - an OWL2-DL ontology for representing information associated with consent. The ontology, along with its documentation, is available online at <https://w3id.org/GConsent> under the CC-by-4.0 license.

The rest of the paper is structured as follows: Section 2 presents an overview of the related work regarding representation of consent using semantic web, Section 3 presents the methodology used to create GConsent, Section 4 presents an overview of the GConsent ontology with an example use-case and discusses limitations, with Section 5 concluding the paper by discussing potential future work.

2 Related Work

This section provides an overview of the existing related work in the context of representing information pertaining to consent using semantic web ontologies. Other relevant work can be found through the W3C Community Group for Data Protection Vocabularies and Controls (DPVCG).

A precursor to GConsent was the consent ontology part of the Consent and Data Management Model (CDMM) [5], created before requirements were documented regarding consent and GDPR. GConsent reuses modelling choices such

as consent attributes for medium, location, and delegation, and improves upon the overall ontology by adding additional concepts such as consent status and states, processing, and additional relationships for context, provision of consent, and relationship between instances of consent. GConsent is also linked to the GDPR and is based on guidance and clarifications provided by authorities and legal-domain organisations regarding consent.

The SPECIAL Usage Policy Language (SPL) [8] defines an usage policy as a set consisting of five items - personal data, purpose, processing, storage, and recipients - which represents authorisation provided by the consent. SPL combines several such (basic) policies into a general usage policy, which is used to enforce and verify compliance by ensuring the requirements of executed processes are within the subset of those permitted by the (consent) usage policy. The core attributes describing consent are similar between GConsent and the SPL, which provides some form of compatibility. SPL provides rigid modelling of storage and data recipient while GConsent leaves it open to the adopter. There are also differences in how provenance is modelled, with SPL focusing on maintaining a log of events for the controller, while GConsent is focused on capturing information about all entities and activities as provenance.

Consent Receipt [10] by the Kantara Initiative provides a way to represent the consent granted by the data subject to a controller using JSON. It provides a specification which controllers can implement to provide a receipt of the given consent to the data subject. There is a large semantic overlap between the information modelled by Consent Receipt and GConsent, such as the modelling of data subject, personal data, and purposes. It is currently not compatible with GConsent due to the differences in terminology as the Consent Receipt was created well before the GDPR. For example, Consent Receipt uses the term "PII" (personally identifiable information) whereas GDPR uses "personal data". A key difference is that the Consent Receipt is a record of consent between two parties that is provided to the data subject, whereas GConsent is used to specify the role of various parties and activities in the context of consent. Both feature information that can be useful towards documenting compliance. By aligning the concepts between the two, GConsent can be used to create an updated semantic GDPR-version of the Consent Receipt, which is part of the planned future work.

3 Ontology Creation

3.1 Methodology

The foremost methodology we used in the creation of GConsent was the seminal guide "Ontology Development 101" by Noy and McGuinness [12], which included using Protégé in maintaining the correctness (e.g. unwanted inferences) of the ontology using the HermiT reasoner. The creation of the ontology followed an iterative model. Each iteration of the ontology was tested for suitability and expressiveness by modelling the collected use-cases and scenarios, then evaluating using competency questions. The methodology can be summarised with the following steps:

4 Pandit et al.

1. Gather information about consent from GDPR, articles, academic papers, communications from various supervisory bodies and regulatory authorities
2. Create use-cases and competency questions based on collected information
3. Create ontology to express information about use-cases
4. Evaluate suitability to express information using competency questions

3.2 Information Collection & Analysis

This section describes the information collection process and its analysis used to model the information associated with consent. The primary source of information were the articles and recitals pertaining to consent within the GDPR [19]. Additionally, Article 29 Working Party, which was the official advisory body providing expert opinions regarding data protection, has provided guidelines on consent [17] that assisted in the interpretation of the GDPR. Apart from these, various guidelines and reports published by the Data Protection Offices and legal firms, Handbook on European Data Protection Law, and relevant court laws² were also used to understand and formulate technical requirements regarding consent.

For the scope of our work, we only considered consent as defined within Art.4-11 of the GDPR. Other special cases of consent (Art.9) such scientific research (Rec.33) and children's personal data (Art.8, Rec.38) were not included due to additional requirements and complexity, as well as lack of legal guidance on their compliance requirements. The use of consent as a legal basis (Art.6,Rec.40) includes conditions for consent to be considered valid (Art.7, Rec.42, Rec.43). The burden of proof and requirements for consent is specified to be on the Data Controller (Rec.42), which requires demonstrable proof that the data subject provided the consent and that it was valid as per the obligations specified in the GDPR.

For consent to be informed, it is necessary to provide certain information to the data subject, such as the specific purposes the personal data will be used for. GDPR also provides data subjects with the right to modify or withdraw consent (Art.7-3). In cases where the consent is withdrawn, processing done prior to the withdrawal is considered valid under the valid consent applicable at that time. This information along with other guidelines provided by the collected resources was used to iterate on a model of consent that could represent the required information.

The information regarding consent can be summarised as follow. Consent has associated attributes regarding the data subject the consent is about, their personal data, the purposes and processing operations associated with personal data, and who the consent is provided to. This is similar to the existing model used by SPL for given consent [8].

In addition to these, there are additional attributes such as - entity that provided consent, status, context (location, medium, instant of creation), and expiry

² The recent decision by CNIL (Décision n°MED-2018-042 du 30 octobre 2018) regarding validity of consent was particularly influential.

that are useful in determining whether the specific instance of consent satisfies the obligations of the GDPR. It is also necessary to include the provenance of consent to determine its validity, particularly for qualitative requirements which cannot be machine-evaluated. The provenance aspect shows some overlap with GDPRov [15] which models provenance of consent based on GDPR. This is resolved by clarifying the scope of GConsent to be limited to modelling consent as an entity, and using GDPRov along with PROV-O [9] to define the provenance.

3.3 Use-cases & Scenarios

This section describes the use-cases and scenarios that were used in the creation of GConsent. The use-cases reflect the requirements gathered from the legal documents as well as various real-world scenarios. They were used to identify the information required regarding consent, and how it should be modelled in the form of an ontology. They were also useful to test the expression of consent using GConsent in different contexts. The complete list of use-cases and scenarios can be found in the documentation.

The use-cases are categorised based on the specific information they relate to. There are a total of 15 categories for use-cases based on the provenance of consent, involved persons and organisations, use of delegation, and third-parties. An example use-case for obtaining consent contains scenarios where consent is given via different mediums such as a web-form or a signed document, as well as when it is given implicitly or via delegation. Similarly, use-cases focusing on the agent that provided consent contain scenarios involving a legal representative of the data subject such as parent or guardian for a minor. Use-cases about the provenance identify the agents and activities involved. Similarly, there are use-cases regarding expiry, medium, modification, and revocation of consent.

3.4 Evaluation

The ontology was evaluated regarding its capability to express information about consent using a set of competency questions. The competency questions, listed in Table 1. were based on the collected use-cases and scenarios, and reflect the queries that can arise regarding compliance of consent under the GDPR. The questions were used as SPARQL queries over the information modelled using GConsent.

The validation of GConsent was done by exploring the suitability of using the ontology to define the information required by each competency question. This was an iterative process where the ontology was tested and modified to accommodate the requirements of the competency questions. Changes were made to the ontology where information was found to be missing or incorrectly modelled. The complete list of these questions along with the specific classes and properties involved in answering them can be found in the documentation. The use of competency questions as compliance queries was based on prior work that demonstrated the use of SPARQL in evaluating GDPR compliance [16].

6 Pandit et al.

The questions are grouped into four broad categories based on their context. The first category of questions relates to consent itself, and inquires about things such as personal data or purpose associated with consent. The second category of questions relates to the activity responsible for creation or invalidation of consent. It inquires whether consent was given by delegation, the role played by the person in delegation, and the activity responsible for delegation. The third category of questions inquire about the context of consent, such as location, medium, expiry, or timestamp of instantiation. The fourth and final category of questions inquire about involvement and role of third parties in any purpose or processing.

Table 1. Competency Questions used to evaluate and validate the ontology

ID	Question
Questions about consent	
C1	Who is the consent about?
C2	What type of Personal Data are associated with the Consent?
C3	What type of Purposes are associated with the Consent?
C4	What type of Processing are associated with the Consent?
C5	What is the Status of Consent?
C6	Is the current status valid for processing?
C7	Who is the consent given to?
Questions about how the consent was created/given/changed/invalidated	
P1	Who created/gave/acquired/invalidated the consent?
P2	If consent was created/given/acquired/invalidated through Delegation, who acted as the Delegate?
P3	If consent was created/gave/acquired/invalidated through Delegation, what was the role played by Delegate?
P4	If consent was created/gave/acquired/invalidated through Delegation, how was the delegation executed?
Questions about the context of how consent was created/given/invalidated	
T1	What is the location of associated with consent?
T2	What is the medium associated with consent?
T3	What is the timestamp associated with the consent?
T4	What is the expiry of the consent?
T5	How was the consent acquired/changed/created/invalidated?
T6	What artefacts were shown when consent was acquired/changed/created/invalidated?
Questions related to Third Party associated with the consent	
D1	Is the purpose or processing associated with a third party?
D2	What is the role played by the third party in the purpose or processing?

4 GConsent Ontology

Based on the methodology described in Section 3, we identified requirements in terms of information required to model the use-cases and scenarios identified in Section 4. These were then used to develop GConsent - an OWL2-DL ontology to express information associated with consent for GDPR. We chose OWL2-DL for its expressibility of relationship and constraints while maintaining reasoning capabilities. GConsent aims to model the context, state, and provenance of consent. Its scope is limited to consent as defined in the GDPR, and is meant to assist in the modelling of information associated with compliance but not determining the compliance itself. GConsent does not model consent as a policy or contract, and therefore is not useful for expressing information such as conditions or clauses that affect consent.

GConsent reuses existing vocabularies such as PROV-O [9] and its GDPR-specific extension GDPRov [15] to model provenance and Time Ontology in OWL [3] for temporal values. It has a preferred namespace of `gc` as used in this paper. Terms within GConsent are linked to their respective definitions in the GDPR using GDPRtEXT [14].

GConsent follows best practices and guidelines advocated by the community for self-documenting ontologies [2,6,7,18] and uses a persistent identifier (w3id) for its IRIs. The ontology and its documentation are available online at <https://w3id.org/GConsent> under the CC-by-4.0 license. The online documentation presents a comprehensive overview of the ontology along with describing the methodology used in its creation, including an analysis of the GDPR. The documentation also presents examples of how the ontology can be used using use-cases and scenarios, with one presented in Section 4.2. All figures follow the Graffoo specification [4] and were created using yEd tool.

4.1 Ontology Overview

The core concepts within GConsent, as presented in Fig. 1 are *Consent*, *Data Subject*, *Personal Data*, *Purpose*, *Processing*, and *Status*. These form the essential information that constitute consent as a legal basis under the GDPR. This is similar to other approaches [8] in the state of the art. The status of consent refers to its state or suitability with respect to use as a valid legal basis under the GDPR.

To facilitate its usage, GConsent distinguishes between valid and invalid states for consent, and provides instances to define states such as implicitly or explicitly given, given via delegation, withdrawn, not given, refused, expired, invalidated, and unknown. The property *invalidates* defines the relation between two iterations of consent - such as when a data subject withdraws given consent where only the latest iteration is considered valid.

Context of consent refers to the information associated with how the consent was created, or obtained, or 'given'. GConsent provides classes and properties, as depicted in Fig. 2, to represent location (using *prov:Location*), medium, and instance of creation (using *time:Instant*). Expiry of consent is defined as the

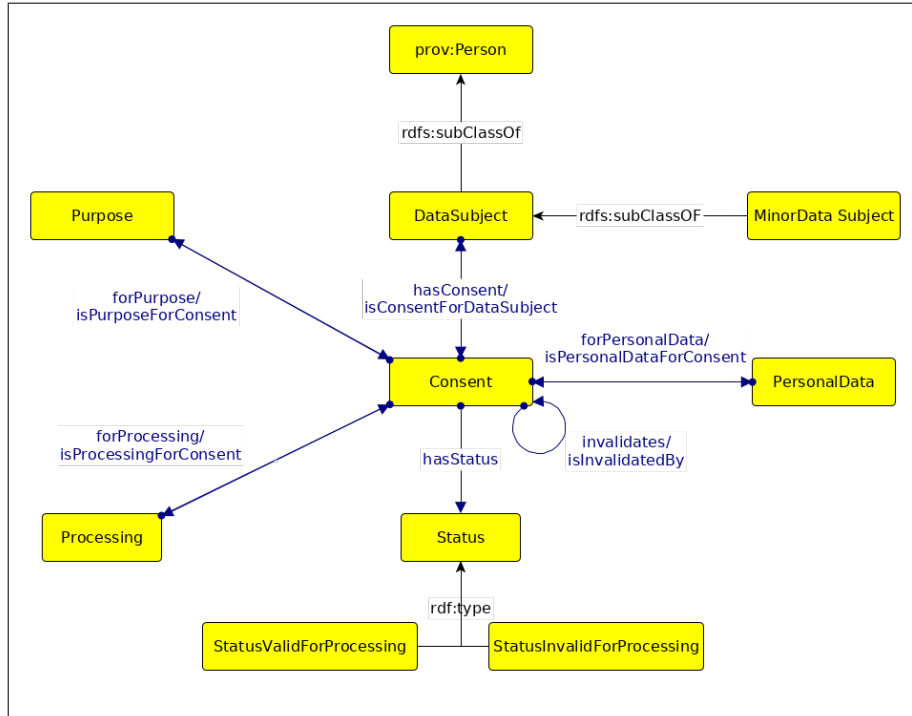


Fig. 1: Overview of the GConsent core ontology

duration or instant after which the consent is no longer considered valid. It is modelled using *time:TemporalEntity* which makes it possible to define it either as a duration (e.g. 6 months) or as an instant in time. To represent the entity that provided consent, GConsent provides the *isProvidedBy* property, whose range is defined as the union of *prov:Person*, *Data Subject*, and *Delegation*, since it is not necessary that the person that provided consent (by delegation) must be a data subject as well. To define other aspects of context, GConsent defines generic properties *hasContext* and its inverse *isContextForConsent* that act as the parent properties for all context relationships.

4.2 Example Use-Case

The example use-case described in Fig. 3 shows implied consent³ in an emergency ward where a nurse provides consent on behalf of the patient. The status of

³ Although the legal basis for obtaining this data under the GDPR could be interpreted as legitimate interest or benefit of data subject, it highlights the recording of information associated with such consent. The example also highlights the potential applicability of GConsent to scenarios other than GDPR such medical consent where additional laws and guidelines apply regarding consent.

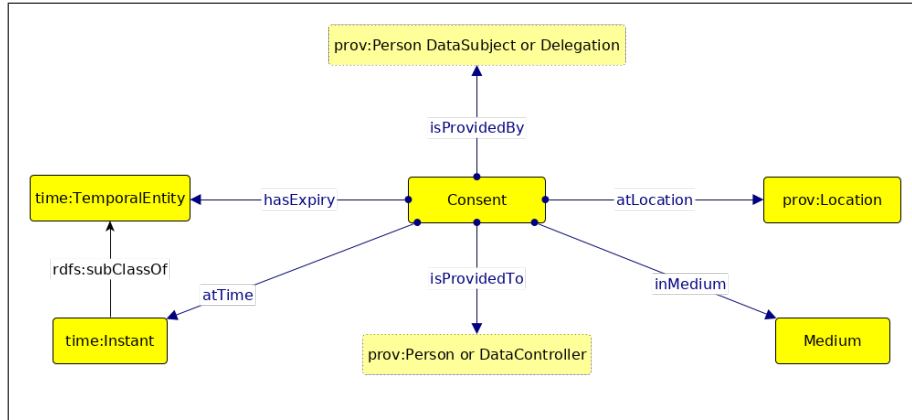


Fig. 2: Concepts representing context of consent

consent in this case is set as implicitly given⁴ even though consent was provided by a delegation where the nurse is the agent that provided consent. The example also shows use of PROV-O and GDPRov vocabularies in capturing provenance aspects of given consent such as the activity that generated consent and entities such as patient records used by it.

When giving consent, sometimes it is required to refer to an abstraction such as a category rather than a specific instance for personal data, processing, or purpose. In the above example, consent is linked using property *gc:forPersonalData* to the broader category of ‘*Health Data*’ rather than some specific instance such as blood group. Such use of punning⁵ allows using a class rather as an instance with a property. As this makes *gc:PersonalData* a meta-class for *ex:HealthData*, further specialisation can be done by defining it as a subclass of an arbitrary class such as *ex:PersonalDataCategory*. Creating examples and guidelines regarding the semantics of such modelling to accurately reflect the use of such abstractions in the real-world⁶ are part of the future work.

4.3 Limitations

GConsent as an ontology has some limitations due to the novelty of consent under the GDPR and the challenges in creating a common ontology for all possible use-cases. In particular, GConsent does not provide a fixed vocabulary for representing temporal and location associated with processing operations such as data sharing or storage. This is due to the perceived ambiguity over whether

⁴ The nurse is the agent that assumes and collects the given consent of the patient, making it an implicit consent given by delegation.

⁵ Punning allows reuse of types. See https://www.w3.org/TR/owl2-new-features/#F12:_Punning.

⁶ Example: privacy policies which mention consent for data categories such as “Account Information” rather than specific instances.

ments gathered from official publications and related resources. This was used to iteratively develop the ontology using a set of use-cases and scenarios which were validated using competency questions. The resulting ontology has applications in modelling information essential in the determination of compliance regarding consent for the GDPR.

GConsent uses PROV-O and its GDPR-specific extension GDPRov to model provenance of consent, and GDPRtEXT to link concepts to the relevant text within the GDPR. Its documentation followed best practices advocated by the community regarding self-documenting ontologies, and contains examples for its use and adoption. The ontology, its documentation, and this paper is available at <https://w3id.org/GConsent> under CC-by-4.0 license.

Compared to the state of the art, GConsent provides additional states for indicating the use of consent other than ‘given consent’. It provides the distinction between valid and invalid states for use as the legal basis for processing of personal data. GConsent also demonstrates the modelling of provenance for activities and agents (such as third parties) and their role in the consent. This is useful to model aspects of provenance such as delegation and agents associated with consent.

Future Work

GConsent provides a generic way to model consent under the GDPR. While the aim of the ontology is to encompass as many use-cases and scenarios as possible, there needs to be a clear and demonstrable application of the work in specific use-cases to drive adoption in the wider community. We plan to develop design patterns that demonstrate the modelling information related to consent and its associated compliance in a variety of contexts. GConsent will play a vital role in such approaches for evaluating compliance based on using consent as a legal basis for processing of data.

One specific example we are working towards takes an existing RDBMS that stores (given) consent information and uses R2RML to produce mappings for generating RDF metadata using GConsent. The resulting data can then be explored and evaluated for compliance using SPARQL queries. The work also aims to address the practice of storing partial information regarding the given consent and combining this information with a common model of the system using GDPRov to generate documentation of consent using GConsent. The approach is expected to demonstrate the feasibility of using a common model versus storing all the information for each instance of consent. This would also facilitate using data validation of information regarding consent.

Acknowledgements

This paper is supported by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

12 Pandit et al.

The authors wish to thank the members of Data Protection Vocabularies and Controls Community Group (DPVCG) for their inputs in the discussion of consent and its related research. The authors also wish to thank Pat McBennett for their help in this work.

References

1. Bartolini, C., Muthuri, R.: Reconciling data protection rights and obligations: An ontology of the forthcoming EU regulation. In: Workshop on Language and Semantic Technology for Legal Domain. p. 8 (2015)
2. Berrueta, D., Phipps, J., Miles, A., Baker, T., Swick, R.: Best practice recipes for publishing RDF vocabularies. Working draft, W3C (2008)
3. Cox, S., Little, C.: Time ontology in OWL. World Wide Web Consortium. Retrieved from <https://www.w3.org/TR/owl-time> (2017)
4. Falco, R., Gangemi, A., Peroni, S., Shotton, D., Vitali, F.: Modelling OWL ontologies with graffoo. In: European Semantic Web Conference. pp. 320–325. Springer (2014). <https://doi.org/10/gfkzgw>
5. 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: Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn) (2017), <http://ceur-ws.org/Vol-1951/#paper-05>
6. Garijo, D.: WIDOCO: a wizard for documenting ontologies. In: International Semantic Web Conference. pp. 94–102. Springer (2017)
7. Gurk, S.M., Abela, C., Debattista, J.: Towards ontology quality assessment. Joint proceedings of the MEPDaW p. 12 (2017)
8. Kirrane, S., Fernández, J.D., Dullaert, W., Milosevic, U., Polleres, A., Bonatti, P., Wenning, R., Drozd, O., Raschke, P.: A scalable consent, transparency and compliance architecture. In: Proceedings of the Posters and Demos Track of the Extended Semantic Web Conference (ESWC 2018) (2018)
9. Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J.: PROV-o: The PROV ontology (2013)
10. Lizar, M., Turner, D.: Consent receipt specification (2017), <https://docs.kantarainitiative.org/cis/consent-receipt-specification-v1-1-0.pdf>
11. Mittal, S., Sharma, P.P.: The role of consent in legitimising the processing of personal data under the current EU data protection framework. Asian Journal of Computer Science And Information Technology 7 pp. 76–78 (2017), <https://papers.ssrn.com/abstract=2975277>
12. Noy, N.F., McGuinness, D.L., others: Ontology development 101: A guide to creating your first ontology. Stanford knowledge systems laboratory technical report KSL-01-05 and ... (2001)
13. Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L.: PrOnto: Privacy ontology for legal reasoning. In: Kő, A., Francesconi, E. (eds.) Electronic Government and the Information Systems Perspective. pp. 139–152. Lecture Notes in Computer Science, Springer International Publishing (2018)
14. Pandit, H.J., Fatema, K., O’Sullivan, D., Lewis, D.: GDPRtEXT - GDPR as a linked data resource. In: The Semantic Web - European Semantic Web Conference. pp. 481–495. Lecture Notes in Computer Science, Springer, Cham (2018). <https://doi.org/10/c3n4>

15. Pandit, H.J., Lewis, D.: Modelling provenance for GDPR compliance using linked open data vocabularies. In: Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn) (2017), <http://ceur-ws.org/Vol-1951/#paper-06>
16. Pandit, H.J., O'Sullivan, D., Lewis, D.: Queryable provenance metadata for GDPR compliance. In: Procedia Computer Science. Proceedings of the 14th International Conference on Semantic Systems 10th – 13th of September 2018 Vienna, Austria, vol. 137, pp. 262–268 (2018). <https://doi.org/10/gfd6r>
17. Party, A.W.: Guidelines on consent under regulation 2016/679 (wp259rev.01) (2018), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051
18. Poveda-Villalón, M., Suárez-Figueroa, M.C., Gómez-Pérez, A.: Validating ontologies with oops! In: International Conference on Knowledge Engineering and Knowledge Management. pp. 267–281. Springer (2012). <https://doi.org/10/gfkzfw>
19. 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) (2016), <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC>
20. Tikkinen-Piri, C., Rohunen, A., Markkula, J.: EU general data protection regulation: Changes and implications for personal data collecting companies. Computer Law & Security Review **34**(1), 134–153 (2018). <https://doi.org/10/gc484m>