Standardisation, Data Interoperability, and GDPR

Harshvarhdan J. Pandit
ADAPT Centre, Trinity College Dublin, Ireland
Christophe Debryune
ADAPT Centre, Trinity College Dublin, Ireland
Declan O'Sullivan
ADAPT Centre, Trinity College Dublin, Ireland
Dave Lewis
ADAPT Centre, Trinity College Dublin, Ireland

ABSTRACT

The General Data Protection Regulation (GDPR) has changed the ecosystem of services involving personal data and information. It emphasises several obligations and rights, amongst which the Right to Data Portability requires providing a copy of the given personal data in a commonly used, structured, and machine-readable format – for interoperability. The GDPR thus explicitly motivates the use and adoption of data interoperability concerning information. This chapter explores the entities and their interactions in the context of the GDPR to provide an information model for the development of interoperable services. The model categorises information and exchanges and explores existing standards and efforts towards use for interoperable interactions. The chapter concludes with an argument for the use and adoption of structured metadata to enable more expressive services through semantic interoperability.

Keywords: Standards, Semantics, W3C, ISO, Vocabulary, Interactions, Business Process, Legal Compliance, Ontology, Data Privacy

INTRODUCTION

Standards emerge when operations have consequences and an agreement is essential for co-operation between stakeholders. In today's world, interoperability is essential for the smooth running of businesses and services that are increasingly dealing with data through the medium of the Internet. With the advent of the Internet as a marketplace with global outreach, the progression of online services has increasingly indulged in personalisation and targeted advertisements. To counter unchecked pervasiveness and instill the responsible use of personal data, privacy laws are enacted and updated to keep pace with ever-evolving technology. The latest of these is the European Council's General Data Protection Regulation ('Regulation (EU) 2016/679...', 2016), which was adopted on 14th April 2016 and entered into force on 25th May 2018. It is the topic of global interest due to the potential of significantly high fines on the order of 20 million euros or 4% of an organisation's global turnover – whichever is higher. Now past its first year, GDPR still continues to be a topic of development and innovation due to its extent of requirements and lack of technological solutions and guidance to address compliance (Good, Rubinstein, & Maslin, 2019).

The GDPR provides the data subject (an individual whose personal data is being processed) with several rights that form an obligation for organisations in order to be compliant. These rights require the provision of information concerning processing in a transparent manner (A12-14) regarding how their personal data is or will be collected, processed, stored, and used along with the specific purposes (A15). The Right to Data Portability (A20) enables the data subject to request a copy of personal data provided to the Data Controller (organisation determining the purposes of processing), or to request it be directly moved, copied, or transferred to another Data Controller. This data is required to be provided in a commonly used, machine-readable, and interoperable format. Thus, the GDPR explicitly mentions and

uses interoperability as a means to ensure a common understanding of data between different Data Controllers, through which it provides the data subject with the freedom to reuse their personal data.

Along with regulating how personal data is used and shared through various processes, the GDPR also provides guidelines, requirements, and obligations on how information is shared or communicated between various entities. For example, when a Data Controller shares data with a Data Processor (organisation performing processing for a Data Controller), the Data Processor is required to carry out its processing limited to the explicit instructions provided by the Controller. These instructions are required to be maintained by the Processor for verifying compliance and ensuring accountability, as well as to clarify the legal responsibilities of each party. Within this arrangement, the Data Processor cannot determine the purpose of the processing, but the Data Processor can share the data with another Data Processor (a Sub-Data Processor) to carry out the processing on its behalf. In such a case, the Data Processor will share the instructions with the Sub-Data Processor, who will, upon completion, notify the Data Processor. The Data Processor will, in turn, notify the Data Controller –thereby establishing a chain where information flows between entities and establishes points of interaction.

While there is no legal requirement for maintaining and using data in a structured and interoperable form, doing so has several benefits for the post-GDPR ecosystem. For Data Subjects and Data Controllers, (semantic) interoperability provides consistency in terms of the understandability of personal data across organisations. For Data Controllers and Data Processors, interoperability enables seamless operations through common mechanisms that also act towards maintaining and demonstrating legal compliance. For Regulatory and Supervisory Authorities, interoperability provides a uniform entry point when conducting investigations into processing operations, and specifically in the case where information flows involve multiple organisations.

In this chapter, we explore these issues of standardisation and data interoperability shaped by the requirements of GDPR and its compliance.

ENTITIES AND INFORMATION IN GDPR

Entities defined by the GDPR

Entities in the context of GDPR are defined and categorised through their roles and responsibilities towards the information required to fulfill the requirements and obligations of compliance. The categorisation of entities also enables identification of relationships through provision and exchange of information between them. Through this, a model emerges representing the commonality and interoperability of information, which is useful to identify and discuss the suitability and applicability of standards for representation, as well as avenues for future work in the standardisation domain.

The model, the entities, and their interactions are visualised in Figure 1. At a broad and abstract level, entities can be categorised into Data Subject (DS), Data Controller (DC), Data Processor (DP), and Supervisory Authority (SA). A Data Subject is an identifiable natural person whose personal data is being processed, and are the user or recipient of a system or service. A Data Controller is an entity that determines the purposes and means for processing of personal data under their control. A Data Processor is an entity that processes personal data on behalf of a Data Controller based on explicitly provided instructions. A sub-processor is a processor acting under another processor. A Sub-Data Processor is bound by the same rules as a Data Processor in terms of limiting processing as per provided instructions. The Supervisory Authority or Data Protection Authority or Regulatory Authority is a governmental institution responsible for monitoring the application of data protection laws.

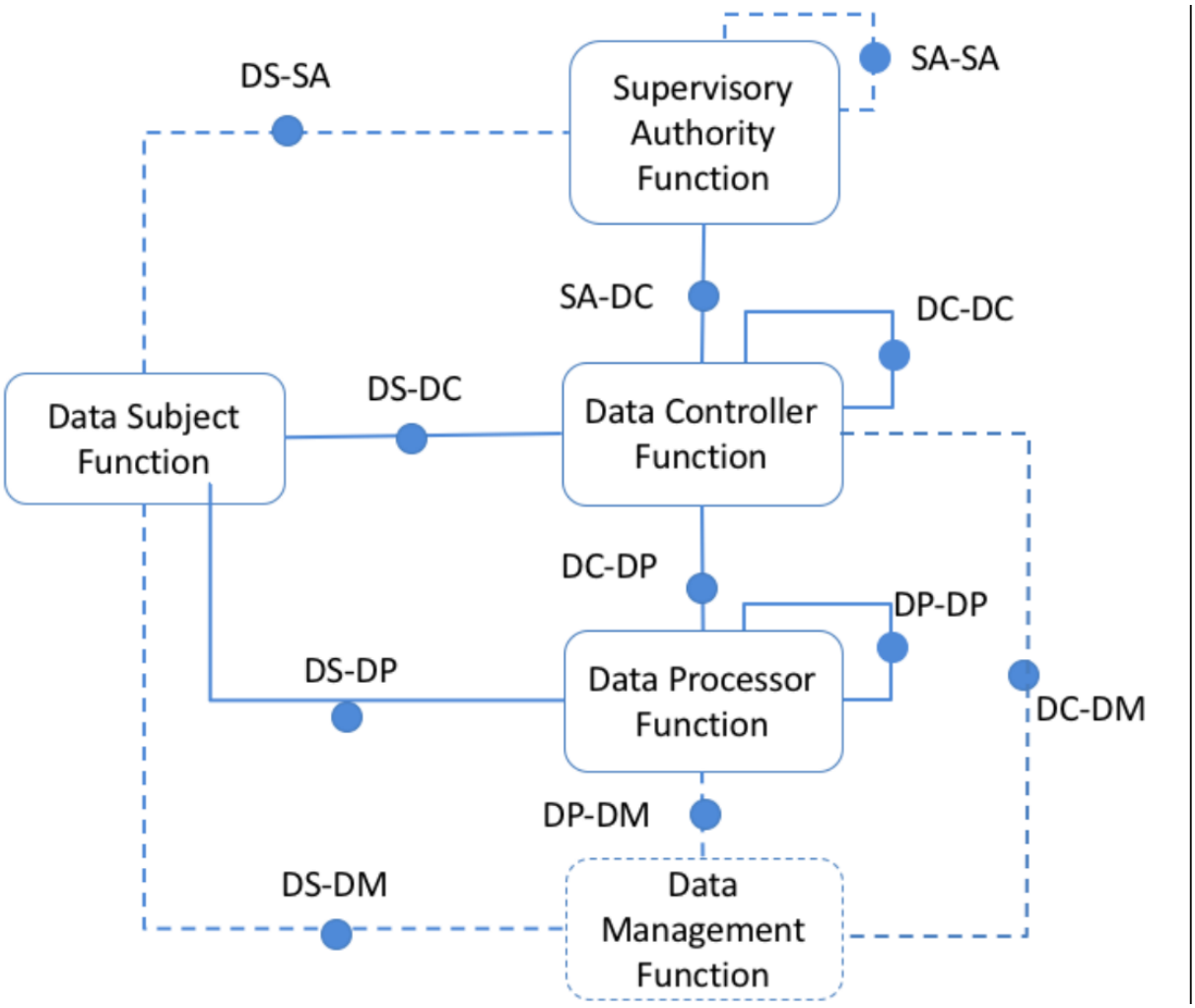

Figure 1: Model of entities and their interactions based on the GDPR (Pandit, Debruyne, et al., 2018)

In addition to these, Data Management (DM) is a virtual entity responsible for the handling and management of information on behalf of a Data Controller or Processor. In this context, "Virtual" refers to the DM not being a separate entity in the legal sense of the term, but being distinct from a DC or DP in terms of functionality and control by virtue of abstraction or automation. An example of a Data Management entity is the use of automated software for interaction with users in an online service, where the Data Subject interacts only with the automated DM for exercising of rights. The DM can be an external interface provided by a third party or a Data Processor contracted by the Controller to act on its behalf. The DM is of interest as the same set of services can be deployed by different Controllers or Processors, thereby providing commonality in terms of functionality and information flows. GDPR can also be interpreted to define more entities such as an Agent or a Representative acting on behalf of another entity such as the Data Subject or Data Controller, Data Protection Officers, organisations that issue certifications, and courts and other authorities involved in the compliance process.

Interactions between Entities

An *interaction* is defined as the exchange of information between two entities irrespective of their type. An interaction between two entities, even of the same type, can be considered as an interoperability point if it involves the communication of some information or structured data between them. Understanding the requirements of this interaction in terms of associated information and context of exchange provides the

basis for exploring opportunities towards standardisation of information practices. In the case of GDPR compliance, the law itself motivates adopting standard practices in terms of interactions between entities – such as in the case of Right to Data portability, which is an interaction between a Data Controller and Data Processor.

Considering all possible combinations of interactions between entities provides a total of six points of interaction without considering the direction of interaction. Also, by including interactions between the same entity types, we have 9 points of interactions, excluding those between Data Subjects as it has no legal basis. For simplicity, we exclude the Data Management functionality as it is a virtual entity and has no mention or role in the GDPR. Similarly, we also exclude the size (large, medium, small, or individual) and nature (commercial, governmental, or not-for-profit) of the entity under the assumption that it has no bearing on the requirements of interoperability within the point of interaction. For specific domains and business sectors, such as health or finance, additional information is exchanged based on specific requirements, which requires a more in-depth review of the domain and its applicable laws. In the case of governmental institutions and organisations that are in a position where information communication needs to be made available for dissemination to the public, the interaction requires such data to be in an 'open' and 'consistent' format, where *open* is defined as being transparent and interoperable towards other entities, and *consistent* is defined as not having temporal changes. Where the interaction is concerned with provision of commercial services, the exchange of information is more concerned with consistency, structure, and correctness.

The interactions between a Data Subject and a Data Controller (other than those governed by the Right to Data Portability), or between a Data Controller and a Data Processor, only require that the provider provides the consumer with the required information in a format that can be understood and used. This provided data is not inherently intended to be made available to anyone else (such as a third-party in this case) and therefore has no requirements in terms of standards as long as the involved entities agree upon the method for sharing of data. Contrast this with the case where a public body such as the Supervisory Authority is involved. Communication from Data Controllers or Data Processors with a Supervisory Authority would have to take into consideration the sensitivity of the private information being shared and therefore would require secure forms of communications that also require security in information itself, such as through encryption or establishment of secure channels. Any warning or ruling by the Supervisory Authority, whether public or private, is also similarly governed by requirements regarding its sensitivity. While currently the SA publishes details of cases (where publicly available) along with decisions through a website, the importance of this information provides incentive to represent it in a more structured format in the future. This representation can adopt existing metadata-rich formats that are used to publish legal documents such as court proceedings and legislations.

Interactions between Data Subjects and Data Controllers

The interaction between Data Subjects and Data Controllers is one of the critical points of interaction addressed by the GDPR. The interoperability between these entities involves the Data Subject providing personal data to Data Controller, which will be in whatever form the Data Controller accepts (by design). However, the Data Subject also provides consent to the Data Controller (which from a legal point of view is specified as the Data Controller collecting consent from Data Subject), which needs to follow specific guidelines stipulated by the GDPR regarding compliance which affects the way consent is collected and stored. Though this does not restrict how the Data Controller obtains consent from the Data Subject, the onus is on the Data Controller to ensure the obtained consent satisfies obligations stipulated by the GDPR for demonstrating the validity of such consent. Therefore, it would be prudent for the Data Controller to obtain or convert consent into a form that makes this process of GDPR compliance more manageable. Storing information about obtained consent bings in requirements towards how this information is structured regarding its representation, storage, and querying and how it can assist in the demonstration of the required compliance.

The interaction of a Data Controller towards Data Subjects also includes the provision of certain information as mandated under the GDPR such as that provided under the Right to Access. Data

Controllers also have to provide this information regarding exercising of rights such as the Right to Data Portability through which a Data Subject can request the Data Controller to provide a copy of their personal data. GDPR also defines the conditions regarding the provision of this data such as its structure or format. Additionally, GDPR also provides Data Subjects the right to have their personal data transferred from one Data Controller to another upon request. The exercising of this right requires both controllers to have some form of interoperability mechanism for mutually understanding the concerned data. This extends to the entity generating it as well as accepting or consuming this data. Such requirements shape the information flow and, therefore, the interoperability of information, and have a role to play in the functioning of the entity and also towards legal compliance. For practical reasons, it is impossible for all entities to have an interoperability agreement or arrangement with each other. Therefore, the provision of such information must be made through open standards and formats that are also commonly used. GDPR provides the same argument for data provided under the Right to Data Portability.

Interactions between Data Controllers and Data Processors

For interactions between Data Controllers and Data Processors, or Data Controllers and Data Controllers, or Data Processors and Data Processors, these already have some ongoing and existing information exchanges outside of GDPR that involve interoperability as part of an organisation's operational practices. Common examples include business arrangements or outsourcing of operations for cost and profit reasons. While such activities are considered a common industry practice, GDPR explicitly mentions the categories of information shared in the operation of such services between these entities. An example of this is the explicit list of instructions provided by the Data Controller to a Data Processor for processing activities over the personal data it provides. The legal acknowledgment of such information-sharing makes its documentation important from the point of compliance. This provides an opportunity for exploring whether a structured and commonly used format can provide advantages to existing practices regarding the sharing of such information.

An approach suggesting an entirely new or different interoperability model would be difficult to uptake due to the diversity and variance of existing infrastructures as well as the cost of changing and adopting them. Therefore, the cost of adopting new practices provides inertia towards keeping existing methods of operation. It is possible to construct a practical interoperability model based on the existing practices with a view towards extending them in an achievable and consistent manner for entities involved. However, this is difficult to achieve in reality due to the earlier mentioned inertia and the cost of change. Since legal compliance is a necessity and GDPR requires operational changes for its obligations, this can be exploited in the adoption of the interoperability model. An approach concerning only that information which is necessary for legal compliance can be proposed as a solution that augments existing services rather than replaces them. Under this, interactions and exchanges between entities through new activities as well as changes to existing ones are defined by the requirements provided by GDPR compliance.

Interactions with Supervisory Authorities

Interoperability as part of GDPR compliance is primarily outlined by the interactions of the Supervisory Authorities with the Data Controllers and Data Processors. Compliance information refers to the data required to demonstrate and determine the organisation's compliance, which legally is acceptable to be in any suitable form as long as it contains the required information. For organisations, the process of maintaining, sharing, and demonstrating compliance using this information becomes a challenge as other entities become involved. For example, under the GDPR, the Data Controller also is concerned with the compliance of the Data Processor as they are provided with the right to carry out reasonable audits for ensuring the Data Processor is acting in accordance with its instructions. Legally, the Data Controller is not responsible for the compliance of the Data Processor. However, since it provides an explicit list of instructions for activities over its personal data, there is a specific and defined relationship between the

compliance of two entities. This motivates towards looking at alternate approaches that can help with the compliance aspect of where information and activities are shared across different entities.

One such example is where information is linked to specific activities associated with the processing of information, which is relevant for compliance. A structured approach that provides an efficient and effective way of storing, managing, and querying of this information presents a technologically structured way to use this information in the demonstration of compliance. In addition, when there are multiple entities involved in the compliance process, the sharing of structured contextual information related to compliance can assist both entities in the demonstration of their compliance. Such requirements also shape the information exchanged between entities and are a part of the interoperability model.

Categorising Information in Interactions

Each interaction point has requirements from multiple GDPR articles that affect the information and activities associated with governing the interoperability between entities. These can be summarised into four categories: Requirements, Processes, Data/Information, and Data Formats. A more detailed exploration of this associates clauses of the GDPR with information and categories and also presents them in a comprehensive tabular format (Pandit, Debruyne, et al., 2018).

The category of ‘Requirements’ reflects a requirement for interoperability that entities are expected to follow or fulfill for compliance. GDPR only states but does not stipulate how this requirement should be fulfilled. The category of ‘Processes’ by contrast, concerns an activity or action as presented in the clause of GDPR, and which leads to processes for the usage, sharing, publication, or exchange of information. For example, Article 16 of the GDPR concerns the Right of Rectification that enables Data Subjects to have their data rectified by the Controller upon request. In this case, the clause specifies the process of rectification, and the requirement for its provision – without specifying how the right should be provisioned.

Where the information consists of some structure or categorisation, it is associated with the ‘Data’ category. Where additional information about category or type of data is specified, this is associated with the ‘Data Format’ category. An example of this is Article 12, which concerns the provision of information about rights in a concise, transparent, intelligible, and easily accessible format. In this case, the clause refers to a process for provision of specific information with criteria for it being valid and compliant.

These requirements do not have a direct bearing on the processing of personal data, but they are useful towards discussions involving requirements gathering, including communication between entities, where standards can be compared or evaluated based on compliance with a requirement or the implementation of a process. For example, Article 30 requires controllers to maintain logs or records of processing activities. While this can refer to abstract information associated with processing activities, it can also be used to formulate records of activities into structured information useful towards demonstration and verification of compliance.

The information associated with information flows can be categorised based on its context and intended usage into five categories - Provenance, Agreements, Consent, Certification, and Compliance. These categories reflect information associated with compliance and organisational processes rather than the personal data that is being processed. The information categories broadly shape and classify the interaction points between entities and refer to the information exchanged. The classification provides a way to refer to the specific type or category of information along with its context without explicitly dealing with specific use-cases or examples of its usage. This abstraction is beneficial towards exploring broad standards towards its representations, such as those for representing provenance or agreements.

The dependence between these categories and their association with interactions between entities suggests an argument for creating create more efficient representations that can enable automation. This can be achieved by integrating the different types of information into a single cohesive model that

operatives at a higher and more abstract level by representing the state of interactions within a system and highlights points of interoperability internally within an organisation.

This presents the possibility of utilising forms of interoperability between the various information categories such that they are capable of referencing each other as required. Such a cohesive set of information forms the basis of the interoperability model, which allows information to be structured systematically for the purposes of storage, querying, and sharing with others. An example can be seen in the case of acquisition of consent, where the consent is represented as an agreement that references the specific processes that will use the data using provenance information while the given consent itself is also recorded as an event using the same or similar provenance mechanisms. This explicit linking of inherently related information allows better representation of information and leads to semantic systems that are capable of intelligent operations. In this case, at a later date, it is possible to identify the given consent for a specific user from provenance logs and to view the process it was obtained against. This itself can further be used to determine if an updated consent is required under the terms of the GDPR upon introducing a change in the process such as an addition of a feature.

Provenance

The provenance information category refers to information about entities and activities involved in producing some data or artefact, which can be used to form assessments about its quality, reliability or trustworthiness. This information is related to the compliance for activities that involve some data that needs to be linked or resolved to the activities that create, use, share, or store it. An example of this is that of consent together with the activities associated with it that obtain, update, or invalidate the consent. For demonstrating compliance, it is essential to show that these activities follow the obligations required for compliance, which requires the presence and maintenance of logs that record the functioning of these activities. These logs can be modelled as a form of provenance in which case they form the life cycle of consent tracking its creation (obtaining), use within different activities, how it is stored, and finally its deletion (invalidation) (Pandit & Lewis, 2017). Compliance then becomes a matter of verifying such provenance logs to see whether the activities recorded the correct and compliant behaviour (Bonatti, Kirrane, Polleres, & Wenning, 2017).

Another example is for checking whether one’s consent was validly given, which requires that the consent should be freely given, be explicit towards specified processes, and must be unambiguous. Since detecting these conditions for validity of consent is not possible without manual oversight, the artefacts and processes involved in the obtaining of provenance can be useful in capturing the state of things as present when obtaining the consent from the Data Subject. Depending on the manner of representing provenance, the life cycle of consent can then be traced with sufficient granularity and abstraction to link it with activities that depend on it, thereby making it possible to also determine whether the consent was used as intended by the terms of the GDPR (Pandit, Debruyne, O’Sullivan, & Lewis, 2019).

As provenance information potentially encompasses all artefacts and processes requiring compliance, it can be argued that having interoperability with relation to sharing and evaluating provenance information would greatly benefit the compliance operations for both the organisation as well as the authorities. Additionally, as compliance itself involves several activities and the creation of artefacts such as compliance reports, this information can also be defined using a common (i.e., shared) provenance model for reuse and dissemination. Such forms of interoperability can be used in any interactions where provenance information needs to be shared or evaluated, such as is also the case with controllers and processors where there is a need to define activities that need to take place or to maintain a joint or collaborative record of activities undertaken that involve both entities. This is especially useful when information needs to be shared that involves life cycles of artefacts such as consent, and personal data need to be tracked or charted across activities. Provenance defined in such manner has led to approaches in the existing corpora of work to create a privacy impact assessment template (Reuben et al., 2016) and a compliance assessment framework (Kirrane et al., 2018).

We mainly identify the use of life cycles for representing the processes and artefacts, whether internal or external to the organisation, as forms of documentation. This provenance information forms

the basis of other information categories as it involves documenting the use of consent and personal data, formation of data-sharing agreements, and recording compliance audits and provision of produced reports. This information is also required to be shared with other entities such as where processors are required to outline their processes to the controllers, and authorities may request to review processes for compliance. The use of provenance also allows recording the occurrence of events such as archival and deletion of consent and personal data which can be vital in the demonstration of compliance.

Data Sharing Agreements

The next category of information we consider is that involving agreements between entities such as that between a Data Controller and a Data Processor, or a Data Controller and another Data Controller, or a Data Processor and another Data Processor. The agreements between these entities have to be in a specific form based on the consideration that they can change depending on factors such as a change in consent or rights being exercised over the personal data provided under the agreement. Therefore, exploring the use of smart agreements (Steyskal & Kirrane, 2015) that can work in an automated manner to a certain extent would benefit systems where a large part of the system can operate on a similar level of automation to ensure compliance. For example, if a Data Controller receives an instruction from a data subject to update their consent for certain activities which are handled by a Data Processor, the Data Controller must update or enforce (depending on the legal term in use to describe the use-case) their agreement to get the Data Processor to also reflect this change in consent over the personal data and activities that they have/had received from the Data Controller. Without some form of automation, such requests would need to be sent and received manually or require manual action, significantly increasing the work and time required to handle them. With automation involved in the process, the Data Controller's system (such as a Data Management interface) can automatically take care of the request by updating the agreement in place for handling the particular consent and personal data with the Data Processor, and can also await a receipt or acknowledgement from the Data Processor for the successful completion of the request. Such agreements that can be iterated, stored, and queried using systems are of benefit to the involved entities as well as other entities that might wish to introspect the agreements such as Certification Bodies and Regulatory Authorities. An example of this is data-sharing agreements that can be explicitly designed to be interoperable based on requirements of the GDPR (Hadziselimovic, Fatema, Pandit, & Lewis, 2017).

Consent

Consent in the context of the GDPR refers to assent or agreement by the data subject about their personal data for the proposed processing activities associated with one or more entities. Given consent refers specifically to the form of consent given by the data subject in relation to their personal data and the proposed usage by activities. Consent can be considered to be an agreement between the Data Subject and the Data Controller (or another entity), and can, therefore, benefit from the same approach as described for implementing data-sharing agreements. This can provide consistency in the application of technology as well as encourage adoption of uniform standards and interoperability in dealing with similar use-cases.

GDPR specifies specific requirements that guide the acquisition and demonstration of consent for it to be evaluated as valid (Mittal & Sharma, 2017). These include the stipulation that consent must be freely given, must be informed, specific, and voluntary. Of these, only the specificity of consent can be gauged from a given consent in a form such as an agreement. Given consent contains the terms which have been accepted by the user, which can be used to gauge the specificity of the agreement, and therefore decide on whether the consent itself was specific or broad under the GDPR. For other stipulations related to valid consent, it is essential to refer to the process and artefacts used to acquire the consent to understand the conditions under which the consent agreement was provided to the data subject and how it was accepted or given or agreed.

For example, in cases where the consent is acquired through a web-form, the entire web-page may need to be preserved to demonstrate that the consent acquisition process complied with the conditions under the GDPR. Therefore, while the given consent may be represented in any form, it also

has to be linked to the processes responsible for acquiring the consent. Additionally, any revision of consent data such as when updating or revoking consent also needs to be stored in a way that can be linked to the processes involved in the change as well as linked to the original consent. This is important as a matter of compliance as GDPR enforcement may require demonstration that a change in consent was carried out correctly, which is only possible through introspection of what the original and changed versions of the consent are. This also introduces the dependency-like relation between data processes and consent where consent should be inherently linked to the processes that depend on it. For example, if the process of using personal data to send emails is dependent on the consent obtained from the user at the time of registration, then it is vital to show that the two are linked together, i.e. the emails are only sent based on the given consent. Such a system must also be able to demonstrate that updated consent has immediate effect on the processes that depend on consent.

These requirements show the inherent dependency of consent and personal data along with the processes involved, which presents a strong argument for representing them together using the same method of provenance. Such a method of capturing the various stages of consent and personal data as life cycles involving processes and artefacts would enable documentation representing the model of the system as a whole. The individual records or logs of activities can then be instantiated based on the model to capture user or event-specific information.

Compliance

Overseeing compliance is an ongoing and continuous process and is specified within the GDPR as an activity to be undertaken by an organisation at certain times. While the interpretation of the law by entities in terms of compliance may vary from use-case to use-case, it is clear that a responsible entity should ensure that all its activities are compliant at all stages of operation. This can be achieved by having proper practices and processes regarding evaluation of compliance from the design stage at the earliest. Such processes ensure that a new service or change in an existing service is compliant before they begin the operation. Several people might be involved in the design and operation of the system, but the responsibility of ensuring the compliance falls on the management or the/a Data Protection Officer (DPO) if appointed (Article 29 Data Protection Working Party, 2016a). In any case, such checks of compliance are integral to audits, done by the organisation itself or by a third-party hired by the organisation, for ensuring the activities meet the required compliance towards legal obligations, and are overseen by the Data Protection Officer (Korff & Georges, 2019). A record of such activities and its outcome is, therefore, an essential outcome of such audits or compliance processes and forms part of the compliance information maintained by the organisation. Such information would prove to be helpful for supervisory authorities who might wish to inspect the activities of an organisation and determine responsibility in cases where multiple entities are involved.

The information associated with compliance-related activities can be represented as provenance information though the processes and artefacts involved in this case are different from those related to the consent and personal data life cycles. To a certain extent, depending on the structuring of compliance activities, it is possible to consider the compliance-related activities as part of a compliance life cycle where the outputs of activities such as reports can be mapped along a timeline using provenance methods similar to those previously outlined. There might be additional requirements for ensuring the security and integrity of such records, though this probably would not have any bearing on the depiction of the information itself. Instead, any concerns related to the data being tampered or accessed without proper authorisation can be mitigated through proper storage and handling of this information. This also allows the provenance representation required for compliance life cycles to be consistent in its purported use-case with those related to provenance of consent and personal data life cycles.

Certifications

GDPR has provisions for seals and certifications that can help organisations with a measure of compliance as well as good practices. These have a maximum validity of three years and have certain conditions or criteria for the creation and issuing of seals and certifications pertaining to GDPR compliance. The seal or certification does not reduce or impact the responsibility of the controller or

processor for compliance with the GDPR but acts as a method of displaying or providing information regarding compliance. The exact nature of such seals and certifications and their role concerning compliance-demonstration to the authorities is still under consideration (European Data Protection Board (EPDB), 2019).

An existing example of such a mechanism is European Privacy Seal ('EuroPriSe', 2019), which carries out an audit of an organisation before providing a seal that is accompanied by a public report published on its website describing the process. The document describes the processes and their compliance with respect to GDPR obligations. While the document itself may be sufficient to demonstrate certain facts regarding the organisation's processes, the fact that it is not published in a format that can be reused by the organisation restricts its usage. The organisation who was the subject of the report has only the option to refer to the report through a legal form of citation.

There are several areas of interest where the information included in the report can be structured for representation in a manner that makes it easy to store, access, query, and, most importantly, share with other entities. For example, if a particular process is responsible for sharing personal data between a controller and a processor, where the processor's processes for handling the said data have been audited through a report, then this information may prove to be sufficient for an agreement between the two entities. However, any such audit and its accompanying report having a validity of a maximum three years require the controller and processor to investigate their respective agreements at the end of this report. Agreements hence need to consider this process as a requirement that hinders the automatic resolution of agreements between the two parties. One way to mitigate this is to keep this requirement out of the automation, in which case the agreements would continue to operate even when the report validity has lapsed. Another case is where processes change, and the processor must renew its certification. If it can demonstrate the changes in its processes, the reports can be linked to the version or iteration of process it evaluated, thereby also providing a way for agreements to view and use this information. Even without use in automated agreements, structuring such information provides motivation for its use within organisation for compliance-related tasks by cross-linking or cross-referencing the information in documentation that can be continuously updated.

EXISTING STANDARDS

When identifying new areas of information representation and standardisation, it is important to acknowledge the importance of reuse and commence by identifying (established) prior work, such as existing standards, and their relevance to the interactions and interoperability discussed previously. At the same time, taking an overview of work carried out within industry, as well as organisations and bodies involved in creating and overseeing standards, and academia allows an outlook into efforts towards standardisation effort. This section provides a summary of existing standards and efforts as applicable for GDPR with a specific focus on their being open for fostering better community participation and adoption.

International Organisation for Standardisation (ISO)

ISO is an international independent non-governmental body composed of representatives from its members' national standards organizations. As such, it represents a global standard-setting body and is widely utilised by the industry and community. While ISO standards are not free (a fee is required to access a standard), their use represents a global agreement, and such is lucrative for large organisations which operate in multiple jurisdictions.

The ISO/IEC 27000 (Disterer, 2013) series concerns Information Security Management Systems and is mostly relevant for the documentation of technical and organisational measures referred to by Article 32 of the GDPR. ISO/IEC 27001 (Lopes, Guarda, & Oliveira, 2019) is a standard for information security management systems and defines information security risks with appropriate measures and controls. It outlines specific requirements and controls to ensure appropriate controls are in place to manage risks to the processing operations, which in the context of GDPR includes personal data.

ISO/IEC 27018 (de Hert, Papakonstantinou, & Kamara, 2016) is a standard concerning 'Personally Identifiable Information (PII) on Public Clouds', which makes it applicable to any processing utilising the cloud. It builds on the abstract control mechanisms defined in ISO/IEC 27002 to specify security issues related to personally identifiable information stored in the cloud. It specifies privacy principles such as consent and choice, purpose legitimacy and specification, rights to access and delete data, information disclosure, and transparency. As such, it addresses several of the requirements for GDPR compliance (Tzolov, 2018), and is intended to be a valuable tool in the compliance certification process. The latest iteration, ISO/IEC 27018:2019, clarifies that it specifies additional controls and guidance for processing which is to be certified with ISO/IEC 27001, rather than a standard on its own. ISO/IEC 27018 corresponds to the ongoing efforts by regulatory bodies to establish uniform collection of standards for processing undergoing in the cloud.

While ISO standards enable agreement over the presence of information and conformance to standard business practices, how this information should be communicated to entities in an interoperable format is not addressed. For example, if a Controller wishes to engage a Processor to carry out some processing and requests information on the technical measures it utilises, the Processor can specify being certified against ISO/IEC 27002 (and 27018) to provide assurance against a standardised set of practices. This can be through a document or a spreadsheet or a list of measures undertaken – that outlines the necessary information in conformance with the ISO standard and is provided by the Processor to the Controller as documentation of its technical measures. While this is sufficient to fulfill legal obligations and business practices, the information can be better represented in an interoperable format to integrate into systems for compliance.

World Wide Web Consortium (W3C)

The World Wide Web Consortium, abbreviated as W3C, is the standards body responsible for information exchange on the Web, which itself is based on the standards and protocols of the Internet. Due to the ever-increasing usage of the Web as a medium for providing services and information, it is important to consider standards that can be readily integrated into mediums such as web pages and web services that form the backbone of interoperability for many organisations, both commercial as well as public institutions. An example of this is email, which is ubiquitous with the Web and illustrates how standards can foster better interoperability. W3C terms its standards as 'Recommendation' to signify its agreement over time by community and members, which reflects its recommended usage rather than adoption as a requirement.

For representing information, W3C has several standards regarding data formats such as XML, CSV, and JSON. These formats provide specifications for the encoding of information into interoperable data streams. The Resource Description Framework ('RDF 1.1 Primer', 2014), or RDF, is a family of specifications that were initially defined as a metadata model but has since been used to model information as web resources. RDF supports several data serialisation formats, including XML and JSON (through JSON-LD), making its usage and adoption easier for information interoperability. RDF allows expression of facts as triples consisting of the subject-predicate-object pattern. This allows the expression of knowledge as a directed graph using a collection of RDF statements, which enables sharing data representations in a consistent manner.

The Web Ontology Language ('OWL 2', 2012), or OWL, is a family of languages for knowledge representations and modelling ontologies using formal semantics built upon RDF. The use of OWL to build schemas (or ontologies) allows the expression and inference of knowledge as well as the use of semantic reasoning. This has attracted interest in the academic as well as industrial community, and there are several public ontologies, with notable examples found in the library and bio-science domains. For querying information declared using RDF, there are mechanisms such as SPARQL ('SPARQL 1.1 Query Language', n.d.) and XQuery ('XQuery', 2017) that operate on standardised forms of data (RDF and XML respectively). Approaches for validating the structure of information defined using RDF include the Shapes Constraint Language (Knublauch & Kontokostas, 2017), which is a W3C Recommendation. To

take advantage of the interoperability offered by commonly used formats such as CSV and JSON, there is ongoing work in creating a standard combining these approaches with RDF. Notable examples for this include CSV on the Web (Tennison, 2016) and JSON-LD ('JSON-LD', 2014). As their names imply, the former uses CSV and the latter JSON. The reuse and combination of standards provide interoperability as well as commonality towards the underlying technology utilised to create, store, and query information represented by these standards. This demonstrates the advantage of combining existing standards towards additional functionality and semantics while ensuring their backward compatibility for technical adoption.

In terms of GDPR, and the information categorised discussed earlier, W3C standards enable the representation of information based on standards that enable machine-readable metadata using RDF, querying using SPARQL and validation using SHACL. In addition, OWL can be used to formulate logic into the metadata to express the interdependencies and relationships inherent in the data. This is especially relevant with the recent interest and trend towards utilising knowledge-graphs where capturing semantics and relationships in the data is of essence. In such a system, GDPR compliance is based on utilising the system to identify the essential information and associate its adherence and validation towards specific clauses for compliance. This has made possible by interpreting the text and concepts of the GDPR itself as machine-readable metadata using RDF in order to link or associate information with its specific clauses (European Union, Publications Office, & ELI Task Force, 2015; Pandit, Fatema, O'Sullivan, & Lewis, 2018)

The Provenance Data Model (Lebo et al., 2013), or PROV, is a W3C Recommendation that provides definitions for interchange of provenance information, which consists of entities and relations between them such as "generated by", "derived from", and "attributed to". PROV was designed to be generic and domain-independent and needs to be extended to address the requirements to represent workflow templates and executions. There are existing approaches in academia that utilise PROV for representing provenance information related to GDPR. Examples include representing the state of a system as a template (Pandit & Lewis, 2017) and representing and maintaining processing logs (Kirrane et al., 2018).

The Open Digital Rights Language (Iannella & Villata, 2018), abbreviated as ODRL, is a W3C Recommendation for policy expression language that provides a flexible and interoperable information model, vocabulary, and encoding mechanisms for representing statements about the usage of content and services. The ODRL Information Model describes the underlying concepts, entities, and relationships that form the foundational basis for the semantics of the ODRL policies. Policies are used to represent permitted and prohibited actions over a certain asset, as well as the obligations required to be met by stakeholders. Policies may furthermore be limited by constraints (e.g., temporal or spatial constraints), and duties (e.g. payments) may be imposed on permissions. ODRL can be utilised for representing agreements, which can include both data sharing agreements as required for Data Controllers and Data Processors, as well as for interpreting GDPR requirements as a policy for compliance checking (Vos, Kirrane, Padget, & Satoh, 2019).

ISA²

The Interoperability solutions for public administrations, businesses, and citizens, or ISA², is a programme that develops and provides digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services. The programme was adopted in November 2015 by the European Parliament and the Council of European Union. ISA² is the follow-up programme to ISA, and aims to ensure interoperability activities are well-coordinated at EU level through a structured plan consisting of a revision to the European Interoperability Framework (EIF) and the European Interoperability Strategy (EIS), along with the development of the European Interoperability Reference Architecture (EIRA) and European Interoperability Cartography (EIC) solutions.

The effort has produced a set of 'Core Vocabularies', maintained by the Semantic Interoperability Community (ALEKSANDROVA, 2016), or SEMIC. Those vocabularies provide a simplified, reusable

and extensible data model for capturing fundamental characteristics of an entity in a context-neutral fashion. Existing core vocabularies include ways to define attributes for people, public organisations, registered organisations, locations, public services, the criterion and evidence required to be fulfilled by private entities to perform public services, and a public event vocabulary. SEMIC has also developed the DCAT Application Profile (DCAT-AP), based on the DCAT specification, for describing public sector datasets in Europe to enable the exchange of descriptions of datasets among data portals. GeoDCAT-AP is an extension of DCAT-AP for describing geospatial datasets, dataset series, and services, while StatDCAT-AP aims to deliver specifications and tools that enhance interoperability between descriptions of statistical data sets within the statistical domain and between statistical data and open data portals. The Asset Description Metadata Schema (ADMS) is a vocabulary to describe and document reusable interoperability solutions, such as data models and specifications, reference datasets, and open-source software. The objective of ADMS is to facilitate the discoverability of reusable interoperability solutions to reduce the development costs of cross-border and cross-sector e-Government systems.

Emerging efforts

Along with efforts towards establishing standards and requirements, the analysis of existing work is also important to identify the potential for reuse and essential drawbacks for adoption. To this end, there have been several critical studies that provide an overview of legal ontologies to date (Leone, Di Caro, & Villata, 2019; Rodrigues, Freitas, Barreiros, Azevedo, & de Almeida Filho, 2019) that present the state of legal ontologies and their usage in the community. At the same time, there are efforts to automate the association of legal requirements with applicable standards – specifically those regarding GDPR and ISO (Bartolini, Giurgiu, Lenzini, & Robaldo, 2017). This provides an important step in the automation of legal compliance by enabling machine-readable and queryable information regarding applicable standards for a specific legal clause. Furthermore, existing work also addresses the requirements of metadata (Wenning & Kirrane, 2018) and standardisation of legal notation associated with compliance (Governatori, Hashmi, Lam, Villata, & Palmirani, 2016).

Data Privacy Vocabulary

The Data Privacy Vocabulary (Pandit & Polleres, 2019) is a work-in-progress effort by the W3C Data Privacy Vocabularies and Controls community group to provide a standardised vocabulary to represent instances of legally compliant personal data handling. It provides a modular vocabulary consisting of concepts for defining personal data categories, purposes of processing, categories of processing, technical and organisational measures, legal bases, recipients, and consent. The vocabulary is defined using RDF and OWL for encapsulating logic and relationships between concepts, which also enables extending it in a compatible manner to define domain-specific use-cases. For example, the vocabulary can be extended for the finance domain by defining the required additional concepts using the W3C standardised mechanisms. Such extensions will remain compatible with the original concepts in the vocabulary while providing domain-specific extensions in the form of a concept hierarchy or ontology. The vocabulary fills an important gap in terms of providing unambiguous definitions that enable interoperability of semantics within the privacy domain.

Consent Receipt

Consent Receipt (Lizar & Turner, 2017) is a standard developed by the Kantara Initiative for representing the consent given by an individual concerning the processing of their personal data. The standard defines the creation of receipts based on equating the giving of consent to a transaction, similar to how a receipt is generated at the end of purchase and payment. The specification requires the receipt to be in human-readable and machine-readable formats for expressing information using predefined categories for personal data collection, purposes, use, and disclosure. In its current state the consent receipt does not address the requirements specified by the GDPR. However, the receipt itself is based on the ISO/IEC

29100 privacy framework and is being discussed for further development in the context of ISO/IEC 29184 regarding online privacy notices and consent.

Data Transfer Project

The Data Transfer Project is an on-going effort to create an open-source platform to facilitate the Right to Data Portability between online services across the Web. Contributors include technology giants such as Apple, Facebook, Google, Microsoft, and Twitter. The code for the platform is currently hosted on Github. The technical approach concerns extracting different information through the available APIs of a service and translating the data to the target platform through the use of intermediate codes or services. The project represents the first step towards an industry-led effort to address the Right to Data Portability and interoperability of information on the Web. At the same time, even though the project was started on July 2018, over a year ago, it has no visible deliverables to date in August 2019.

The Future: An argument for Semantic Interoperability

SEMIC (and EIF) define Semantic Interoperability as the preservation of meaning in the exchange of electronic information (ALEKSANDROVA, 2016). In the context of information exchange, the sender and receiver should understand and interpret information in the same way. Semantic interoperability is achieved through establishment of shared agreements on the meaning and context of information exchanged. These agreements are usually formalized in an artefact called an ontology, vocabulary, or schema. Systems that have semantic interoperability can exchange information in a more flexible manner due to the nature of interpretation being based on a shared agreement for the provision of context. Such context can be represented as metadata describing the system and providing information regarding both the content and the context.

Concerning the Right to Data Portability, GDPR only stipulates that the provided data be intended towards enabling interoperability. The Article 29 Working Party in its guidelines (Article 29 Data Protection Working Party, 2016b) observes that where there are no commonly used formats used within a particular domain or context, Data Controllers should provide personal data in commonly used formats such as CSV, JSON, and XML, along with useful metadata at the best possible level of granularity. This metadata should be used to accurately describe the meaning of the exchanged information to make the function and reuse of data possible. The guidelines further call for cooperation between industry stakeholders to adopt a common set of interoperable standards and formats to deliver the requirements of the Right to Data Portability. Therefore, there needs to be an initiative to go beyond the requirements of providing the data in interoperable formats such as CSV and XML and to work towards the establishment and adoption of semantic metadata.

One possible solution is to utilise existing data formats and extending them to support additional contextual metadata. Examples of this are the CSV on the Web, which augments the CSV data format, and JSON-LD, which encodes RDF in JSON format. Adopting such data formats is easier for existing systems that already support their native formats (CSV and JSON respectively) and can provide the necessary mechanisms for representation of data semantics.

The creation of appropriate metadata to describe information should follow the general guidelines from established methods such as the Semiotic Information Theory (Stamper, 1996), which considers the information content of signs and expressions. In this case, the information content represented by the data would replace signs and expressions in the theory. The structuring of information according to this theory can be represented through Stamper’s Semiotic Ladder (Stamper, 1996) , which is a framework provided by semiotics to discuss and prescribe practical and theoretical methods for the design and use of information systems. This requires agreement between various stakeholders on the creation and adoption of schemas, ontologies, and vocabularies for their respective domains.

The adoption of such semantic metadata would enable better interoperability between systems in terms of requesting data from different providers under an open and shared semantic base. An example of this is requesting a user's profile information from different providers, where a profile contains personal information such as name and email as well as information such as address and references to other social media accounts. This information can be of relevance to generic services such as contact books as well as

for specialised services such as other social media services. By using a common vocabulary to define these pieces of information, a single query can retrieve the information from multiple services, as well as provide it in a manner such that it can be identified by generic as well as specialised services.

A prevalent example of semantic interoperability can be observed on the Web through schema.org ('Schema.org', n.d.), which is a collaborative community effort towards creating and maintaining schemas for use on web pages. Its primary use is to act as a shared vocabulary for metadata on websites that will assist search engines understand the content on the website. A similar effort needs to be undertaken to define interoperable metadata for content being provided as part of the Right to Data Portability, and by extension, other aspects of the GDPR.

CONCLUSION

This chapter presented an exploration of data interoperability based on entities and obligations driven by the General Data Protection Regulation (GDPR). The discussion of other interactions between entities and the categorisation of information flows at such interactions presents sufficient motivation for further work towards identifying commonality and working towards standardisation of information and services based on the requirements of legal compliance with the GDPR and other legislations.

The chapter also provided an overview of existing standards and efforts towards standardisation of information and their relevance with the ecosystem brought about by the GDPR. The promise of automating compliance and its related services and systems provides an argument to also drive efforts for the other information categories identified within the chapter. At the same time, incorporating semantics into information enriches the existing information exchange and enables the creation and utilisation of services with greater flexibility and functionality. For example, a Data Controller can be utilised semantic representations of their system to create interoperable information for documentation of compliance, drafting of privacy policies, and agreements with processors – based on the commonality of information involved and the necessity to exchange this information with other entities. There are existing efforts that are working towards such semantic interoperability and are driven by the various stakeholders including academia and industry. While GDPR itself is a highly contextual domain for services and approaches, it also presents a promising avenue for further standardisation efforts driven by economic and legal incentives.

Acknowledgments

This research is supported by the ADAPT Centre for Digital Media Technology, which is funded by Science Foundation Ireland (SFI) through SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant 13/RC/2106.

REFERENCES

ALEKSANDROVA, Z. (2016, November 25). Core Vocabularies [Text]. Retrieved 23 July 2019, from ISA²—European Commission website: https://ec.europa.eu/isa2/solutions/core-vocabularies_en

Article 29 Data Protection Working Party. (2016a). *Guidelines on Data Protection Officers ('DPOs')* (No. 16/EN, WP-243).

Article 29 Data Protection Working Party. (2016b). *Guidelines on the right to data portability* (No. 16/EN, WP242).

Bartolini, C., Giurgiu, A., Lenzini, G., & Robaldo, L. (2017). Towards Legal Compliance by Correlating Standards and Laws with a Semi-automated Methodology. In T. Bosse & B. Bredeweg (Eds.), *BNAIC 2016: Artificial Intelligence* (Vol. 765, pp. 47–62). https://doi.org/10.1007/978-3-319-67468-1_4

Bonatti, P., Kirrane, S., Polleres, A., & Wenning, R. (2017). Transparent Personal Data Processing: The Road Ahead. *Computer Safety, Reliability, and Security*, 337–349. https://doi.org/10/gfxvtb

de Hert, P., Papakonstantinou, V., & Kamara, I. (2016). The cloud computing standard ISO/IEC 27018 through the lens of the EU legislation on data protection. *Computer Law & Security Review*, *32*(1), 16–30. https://doi.org/10/f8bv3h

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, *34*(2), 193–203. https://doi.org/10/gdtmx7

Disterer, G. (2013). ISO/IEC 27000, 27001 and 27002 for Information Security Management. *Journal of Information Security*, *04*(02), 92–100. https://doi.org/10/gfvhsf

European Data Protection Board (EPDB). (2019). *Guidelines 4/2018 on the accreditation of certification bodies under Article 43 of the General Data Protection Regulation (2016/679)*.

European Union, Publications Office, & ELI Task Force. (2015). *ELI: A technical implementation guide.* Luxembourg: Publications Office.

EuroPriSe. (2019). Retrieved 11 August 2019, from European Privacy Seal (EuroPriSe) website: https://www.european-privacy-seal.eu/EPS-en/Home

Good, N., Rubinstein, I., & Maslin, J. (2019). *'When the Dust Doesn't Settle' – GDPR Compliance One Year In*. Retrieved from https://www.ssrn.com/abstract=3378874

Governatori, G., Hashmi, M., Lam, H.-P., Villata, S., & Palmirani, M. (2016). Semantic Business Process Regulatory Compliance Checking Using LegalRuleML. In E. Blomqvist, P. Ciancarini, F. Poggi, & F. Vitali (Eds.), *Knowledge Engineering and Knowledge Management* (pp. 746–761). Springer International Publishing.

Hadziselimovic, E., Fatema, K., Pandit, H. J., & Lewis, D. (2017). Linked Data Contracts to Support Data Protection and Data Ethics in the Sharing of Scientific Data. *Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci)*, 55–62. Retrieved from http://ceur-ws.org/Vol-1931/paper-08.pdf

Iannella, R., & Villata, S. (2018, February 15). ODRL Information Model 2.2. Retrieved 19 September 2018, from ODRL Information Model 2.2 website: https://www.w3.org/TR/odrl-model/

ISO/IEC 2382-1:1993. (1993). Retrieved from http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/00/72/7229.html

JSON-LD. (2014, January 16). Retrieved 11 August 2019, from JSON-LD 1.0 A JSON-based Serialization for Linked Data website: https://www.w3.org/TR/json-ld/

Kirrane, S., Fernández, J. D., Dullaert, W., Milosevic, U., Polleres, A., Bonatti, P., … Raschke, P. (2018). A Scalable Consent, Transparency and Compliance Architecture. *Proceedings of the Posters and Demos Track of the Extended Semantic Web Conference (ESWC 2018)*. https://doi.org/10/gfxvsf

Knublauch, H., & Kontokostas, D. (2017, July). Shapes Constraint Language (SHACL). Retrieved 19 September 2018, from Shapes Constraint Language (SHACL) website: https://www.w3.org/TR/shacl/

Korff, D., & Georges, M. (2019, July 30). *The Data Protection Officer Handbook*. Retrieved from https://ssrn.com/abstract=3428957

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., … Zhao, J. (2013). PROV-O: The PROV Ontology.

Leone, V., Di Caro, L., & Villata, S. (2019). Taking stock of legal ontologies: A feature-based comparative analysis. *Artificial Intelligence and Law*. https://doi.org/10/gf3z84

Lizar, M., & Turner, D. (2017). *Consent Receipt Specification v1.1.0* (p. 29). Retrieved from Kantara Initiative website: https://docs.kantarainitiative.org/cis/consent-receipt-specification-v1-1-0.pdf

Lopes, I. M., Guarda, T., & Oliveira, P. (2019). How ISO 27001 can help achieve GDPR compliance. *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. IEEE.

Mittal, S., & Sharma, P. P. (2017). 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*, 76–78.

OWL 2. (2012, December 11). Retrieved 11 August 2019, from OWL 2 Web Ontology Language Document Overview (Second Edition) website: https://www.w3.org/TR/owl2-overview/

Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2018). An Exploration of Data Interoperability for GDPR. *International Journal of Standardization Research (IJSR)*, *16*(1), 1–21. https://doi.org/10/gfsn52

Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2019). GConsent—A Consent Ontology Based on the GDPR. In P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A. J. G. Gray, V. Lopez, … K. Hammar (Eds.), *The Semantic Web* (pp. 270–282). Retrieved from https://w3id.org/GConsent

Pandit, H. J., Fatema, K., O'Sullivan, D., & Lewis, D. (2018). GDPRtEXT - GDPR as a Linked Data Resource. *The Semantic Web - European Semantic Web Conference*, 481–495. https://doi.org/10/c3n4

Pandit, H. J., & Lewis, D. (2017). Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies. *Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn)*. Retrieved from http://ceur-ws.org/Vol-1951/PrivOn2017_paper_6.pdf

Pandit, H. J., & Polleres, A. (2019, July 26). DPV. Retrieved 11 August 2019, from Data Privacy Vocabulary v0.1 website: https://www.w3.org/ns/dpv

RDF 1.1 Primer. (2014, June 24). Retrieved 11 August 2019, from RDF 1.1 Primer website: https://www.w3.org/TR/rdf11-primer/

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). *Official Journal of the European Union*, *L119*, 1–88.

Reuben, J., Martucci, L. A., Fischer-Hübner, S., Packer, H. S., Hedbom, H., & Moreau, L. (2016). Privacy Impact Assessment Template for Provenance. *Availability, Reliability and Security (ARES), 2016 11th International Conference On*, 653–660. https://doi.org/10/gfxvsw

Rodrigues, C. M. de O., Freitas, F. L. G. de, Barreiros, E. F. S., Azevedo, R. R. de, & de Almeida Filho, A. T. (2019). Legal ontologies over time: A systematic mapping study. *Expert Systems with Applications*, *130*, 12–30. https://doi.org/10/gf223z

Schema.org. (n.d.). Retrieved 23 July 2019, from https://schema.org/

SPARQL 1.1 Query Language. (n.d.). Retrieved 30 April 2019, from SPARQL 1.1 Query Language website: https://www.w3.org/TR/sparql11-query/

Stamper, R. (1996). Organisational semiotics. In *Information systems: An emerging discipline?* Mc Graw-Hill.

Steyskal, S., & Kirrane, S. (2015). If you can't enforce it, contract it: Enforceability in Policy-Driven (Linked) Data Markets. *SEMANTiCS (Posters & Demos)*, 63–66. Retrieved from https://pdfs.semanticscholar.org/f2c3/cac9b4af913f32dbd5034ed9aa1751a8a337.pdf

Tennison, J. (2016, February 25). CSV on the Web. Retrieved 11 August 2019, from CSV on the Web: A Primer website: https://www.w3.org/TR/tabular-data-primer/

Tzolov, T. (2018). One Model For Implementation GDPR Based On ISO Standards. *2018 International Conference on Information Technologies (InfoTech)*, 1–3. https://doi.org/10/gf3cx7

Vos, M. D., Kirrane, S., Padget, J., & Satoh, K. (2019). ODRL policy modelling and compliance checking. *3rd International Joint Conference on Rules and Reasoning (RuleML+RR 2019)*, 16. Bolzano, Italy.

Wenning, R., & Kirrane, S. (2018). Compliance Using Metadata. In T. Hoppe, B. Humm, & A. Reibold (Eds.), *Semantic Applications: Methodology, Technology, Corporate Use* (pp. 31–45). https://doi.org/10.1007/978-3-662-55433-3_3

Wong, J., & Henderson, T. (2018). How Portable is Portable?: Exercising the GDPR's Right to Data Portability. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 911–920. https://doi.org/10/gfsqrk

XQuery. (2017, March). Retrieved 11 August 2019, from XQuery 3.1: An XML Query Language website: https://www.w3.org/TR/xquery-31/