International Journal of Standardization Research
Volume 16 • Issue 1 • January-June 2018

An Exploration of Data Interoperability for GDPR

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

ABSTRACT

The General Data Protection Regulation (GDPR) specifies obligations that shape the way information is collected, shared, provided, or communicated, and provides rights for receiving a copy of their personal data in an interoperable format. The sharing of information between entities affected by GDPR provides a strong motivation towards the adoption of an interoperable model for the exchange of information and demonstration of compliance. This article explores such an interoperability model through entities identified by the GDPR and their information flows along with relevant obligations. The model categorises information exchanged between entities and presents a discussion on its representation using existing standards. An investigation of data provided under the Right to Data Portability for exploring interoperability in a real-world use-case. The findings demonstrate how the use of common data formats hamper its usability due to a lack of context. The article discusses the adoption of contextual metadata using a semantic model of interoperability to remedy these identified shortcomings.

KEYWORDS

Data Format, Data Standards, GDPR, Interoperability, Semantics

INTRODUCTION

Businesses are increasingly using personal data to provide services, especially online, in various forms such as personalisation of provided services and targeted advertisements. Such services need to adhere to data protection laws governing the collection and subsequent usage and sharing of personal data. Previously, the Data Protection Directive, or DPD (DPD, 1995), in the European Union regulated the processing of personal data. This has been superseded by the General Data Protection Regulation (GDPR, 2016), abbreviated as GDPR, which is the new European data protection legislation that entered into force on 25th May 2018. Non-compliance towards its obligations carries a fine of up to €20 million or 4% of a company's global annual turnover of the previous financial year, whichever is higher, based on the nature of offense (Article 83). This makes GDPR an important legislation in terms of changes to the organisational measures required for compliance. In particular, GDPR

DOI: 10.4018/IJSR.2018010101

focuses on the use of consent and personal data as the basis of operations and provides the data subject with several rights. These new changes have spurred innovation within the community that targets compliance with the various obligations of the GDPR.

Along with providing constraints for how personal data is used and shared through various processes, the GDPR also provides statements about the way information is shared or communicated between various entities. GDPR provides seven key principles (Article 5) that act to guide the processing of personal data. These are - Lawfulness, fairness and transparency, Purpose limitation, Data minimisation, Accuracy, Storage limitation, and Integrity and confidentiality, and Accountability. While these principles are similar to those within the DPD, GDPR encompasses these principles in a larger role in its adherence towards compliance. These principles set out how each data controller should process the personal data of clients or data subjects and forms the guideline for duties and obligations for compliance by entities. For example, a Data Processor under the GDPR is an entity that can only act on the data under the instructions it receives from a Data Controller or another Data Processor (making it the sub-Processor). Therefore, a Data Processor cannot decide the purpose of the data it receives and must adhere to the instructions it receives from the Data Controller or Data Processor that provides the data. Assuming this entity is a Data Controller, the agreement with the Data Processor is expected to state these responsibilities in an explicit manner such that the Data Processor as well as the Data Controller can verify or audit the accountability of this agreement for obligations provided by the GDPR.

The GDPR provides several rights to the data subjects whose adherence is mandatory for organisations. The Right to Inform (Article 12-14) and Right to Access (Article 12, 15) provide the Data Subject the right to be informed regarding how their personal data is or will be collected, processed, stored, and used along with the specific purposes. The Right to Data Portability (Article 12, A20) enables the Data Subject to receive a copy of their personal data which they have provided to the Data Controller. It also allows the Data Subject to request this data to be directly moved, copied, or transferred to another Data Controller. The provided must be in a commonly used, machine readable, and interoperable format. The exercising of these rights involves an explicit interaction between the Data Controller and the Data Subject or another Data Controller where the information exchanged is the personal data under consideration. Additionally, GDPR explicitly mentions interoperability as one of the mandatory properties of this data, making its adoption a necessary part towards its compliance.

While there is no requirement for legally structuring shared data in a particular way, doing so has benefits for all entities involved. For Data Subjects, this provides consistency in terms of understandability and interoperability of their personal data. For Data Controllers and Data Processors, this enables seamless operations through interoperable mechanisms that also act as demonstrable compliance towards required obligations. For Supervisory Authorities, the interoperability of data provides a uniform interface when conducting investigations, being particularly helpful when tracing the flow of information across multiple entities.

This paper investigates interoperability in the context of the GDPR. It presents an overview of the GDPR in terms of entities involved and presents a systematic representation of their interactions. Through this, it presents an analysis of the entities categorised according to their role as defined by the GDPR, the nature of information flows between these entities, and the requirements for interoperability in their interactions through these information flows. Through this analysis, the paper investigates the creation of an interoperability model based on interactions between entities and shaped by relevant obligations enforced by the GDPR which act as requirements for the model. The purpose of this model is to highlight how information exchanged by entities is affected by the GDPR and to discuss its representation using various existing standards and standard-creating bodies. The paper also presents a discussion on how the model is useful for operations involving information exchange and towards ensuring its compliance with the obligations of the GDPR.

The main contribution of this paper is the above-mentioned model for interoperability based on the GDPR. Through the model, the paper identifies the categories of information shared by entities

and discusses their various requirements in terms of interoperability as well as regulatory compliance. Through this discussion, the paper presents its arguments towards the standardisation of representation for these information categories. An investigation of existing standards bodies and available standards for representing the identified information categories is also presented. To demonstrate applicability of the model and its identified information flows, the paper presents an application of the model that investigates interoperability within real-world use-cases based on data provided under the Right to Data Portability. The findings of this investigation reveal that while the data is interoperable, the provided data lacks practical usefulness in terms of semantics. The paper provides a discussion on how this can be resolved using contextual metadata in a semantic model of interoperability based on the semiotic information theory. This discussion also involves the standardisation of such metadata between entities to ensure their interoperability.

This paper is an extension of our previous work (Pandit, O'Sullivan, & Lewis, 2018), where we expand our information model to involve management interfaces as entities along with the additional information flows associated with it and provide a more in-depth analysis of existing standards towards the representation of information. We also present a real-world use-case which analyses the specific data formats used for compliance with the Right to Data Portability and presents our work towards evaluating its usefulness in the context of interoperability. Based on this, we present our arguments towards adopting a more semantic-based approach to define contextual metadata for better interoperability.

ENTITIES AND INFORMATION

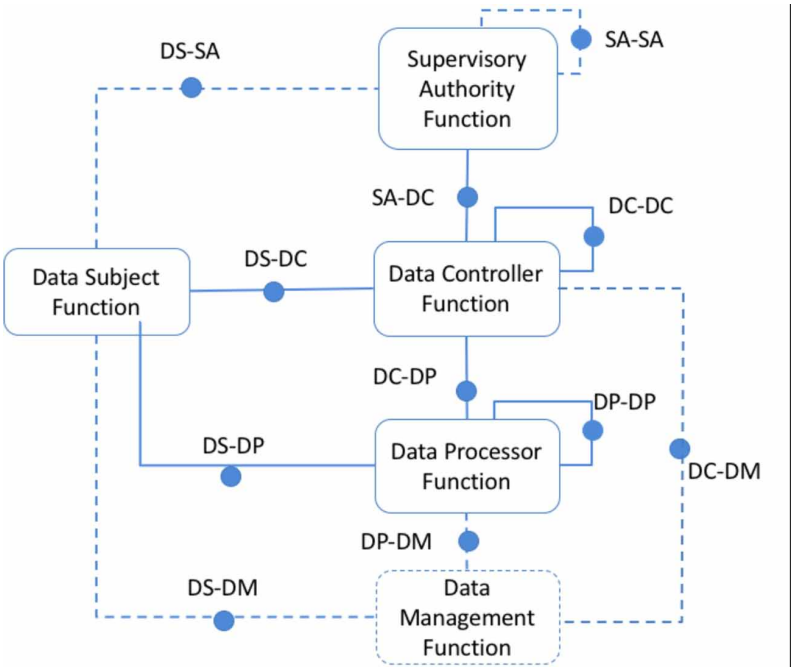
To understand the entities and their relevant obligations under the GDPR, we analysed the text of the GDPR along with various documents provided by supervisory authorities such as the Data Protection Commissioner's Office - Ireland (DPC Ireland, 2018) and Information Commissioner's Office - United Kingdom (ICO, 2018), documents provided by the Article 29 Working Part (abbreviated as WP29) for outlining the nature of obligations under the GDPR, and various information articles and documentation provided by commercial organisations regarding compliance and dissemination of information regarding GDPR. Through this, we first identified the entities and their responsibilities as a matter of compliance under the GDPR, and the information required to fulfil their obligations towards compliance. Based on this, we identified the relationship of entities in terms of exchange of information amongst them. The identified information was then categorised based on the nature of information and relation to compliance. This provided a way to model the commonality and interoperability of data using these categories, as well as to discuss the various standards for their representation. The outcome of this work was a theoretical framework for how the information can be exchanged in an interoperable fashion and provides a background for analysis of use-cases.

An overview of the data interoperability model for GDPR can be seen in Figure 1 which depicts the different entities along with the possible interoperability points between them along with examples of information and processes associated with each such point. Any interaction between two entities, even of the same type, can be considered as an interoperability point if it involves communication of some information or structured data between them towards normal operational practices or for GDPR compliance. Understanding the requirements of this communication between the entities such as what is the associated information, why it is being shared or exchanged, and what are the requirements that shape this information provide the basis for exploring opportunities towards standardisation of information practices. In the case of GDPR compliance, the law itself provides a motivation towards adopting standard practices in terms of interactions between entities.

Categorisation of Entities

For entities and the interoperability between them, we use the following abbreviations or denominations. A Data Subject (DS) is an individual or entity in the context of personal data. They are the user

Figure 1. Data interoperability model for GDPR



or recipient of a system or a service and provide the consent for activities. Data Controller(s) or 'Controller' (DC) is an entity that determines the purposes and means of the processing of personal data. They can act jointly, in which case they are called Joint Controllers. A Data Processor (DP) is an entity that processes personal data on behalf of the controller. The relationship between controllers and processors is many-to-many, i.e. either can be associated with multiple entities of the other type. A sub-processor is a processor acting under another processor. They are bound by the same rules of agreement as the processor they are under with its controller. The Supervisory Authority (SA) or Data Protection Authority (DPA) is a public institution responsible for monitoring the application of data protection laws.

Data Management (DM) is a virtual entity responsible for the handling and management of information on behalf of the Data Controller. Virtual in this case refers to the DM not being a separate entity in the legal sense of the term but having a distinction with the functions of its controlling entity (Data Controller) 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 only interacts with the DM for the operation of the service as well as exercising of rights. The DM was added to the extended version of our work based on the use of automated systems to process and provide data in the real-world by organisations such as Google and Facebook. Apart from these entities, GDPR can also be interpreted to have other entities not considered within the scope of this work. These are an Agent or a Representative acting on behalf of another entity such as the Data Subject or Data Controller, a Data Protection Officer, organisations that issue certifications and seal as provided by the GDPR, as well as additional regulatory bodies and authorities that might be involved in the compliance process.

Interoperability Between Entities

The model in Figure 1 consists of entities and the information flows between them showing several points of interactions. Each point of interaction between two entities is considered to be a point for

interoperability between the two entities. Taking the entities under consideration as Data Subject (DS), Data Controller (DC), Data Processor (DP), and Supervisory Authority (SA), we have a set of 6 possible points for interoperability without considering the direction of interaction. Additionally, controllers, processors, and supervisory authorities can interact with other controllers, processors, and supervisory authorities respectively. This brings the total count of possible points to 9. If a Data Management interface (DM) is used by Data Controllers to interact with Data Subjects, then this adds two more points of interaction bringing the total to 11 points. It is to be noted here that the functionality of DM is not specified by the GDPR in the form of suggestion or requirement but is a practical consideration that could be used by Data Controllers to automate parts of their operations for practical reasons.

The entities depicted in the model are based on an analysis of the text of the GDPR along with other related documents published by various organisations associated with data protection and regulatory compliance. Since only the type of entity is required for understanding and modelling the interaction, their size (large, medium, small, or individual) or nature (commercial, governmental, or not-for-profit) is assumed to have no bearing on the requirements of the interoperability point. Additional information may need to be exchanged based on specific requirements based on the type of the entity, such as additional responsibilities required by larger organisations as compared to individuals, though this requires a deeper review of the law and clarification through legal experts. We therefore do not consider such additional requirements to be within the scope of this paper. For entities such as governmental institutions and organisations that are in a position where information communication needs to be made available for dissemination to the public, we consider this as motivation to explore the requirements of sharing such data in an 'open' and 'consistent' manner, where open is defined as being transparent and interoperable towards other entities, and consistent is defined as not having temporal changes. Where entities are commercial entities, interoperability is more concerned with consistency, structure, and correctness of information being exchanged.

Consider the interactions between a Data Subject and a Data Controller, or between a Data Controller and a Data Processor, where the interoperability between them only requires that the provider should provide the consumer with the required information in a format that can be accepted and operated on. This provided data is not inherently intended to be made available to anyone else (such as another entity which is a third-party in this case), and therefore has no bound requirements in terms of standards at this point of interaction 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 private information being shared, and therefore would require the use of secure forms of communications which may also require security in the structuring of data itself, such as through encryption or establishment of secure channels. Any warning or ruling by the Supervisory Authority that can be considered public information, as in made available to the public, would also need to be published in an appropriate manner in regard to its sensitivity. A modern method of doing this is to publish details of use-cases along with their rulings or decisions on the official website. Such information in the future might be collated in a registry or dataset using appropriate formats and structuring.

The interaction between Data Subjects and Data Controllers is one of the important 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 certain 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 compliance easier. This brings 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.

For interactions between Data Controllers and Data Processors, or Data Controllers and Data Controllers, or Data Processors and Data Processors, these interactions already have some ongoing and existing information exchanges 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 this is the explicit list of instructions shared by the Data Controller to a Data Processor for processing activities over the personal data it provides. The legal acknowledgement of such information sharing makes its documentation important from the point of compliance. This provides an opportunity for exploring whether a structured and common 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 them. Therefore, the cost of adopting new practices provides an 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 towards existing ones are defined by the requirements provided by GDPR compliance.

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 the explicit list of instructions for activities over its personal data, there is a certain relationship between the compliance of the 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 certain activities associated with the processing of information which is relevant for compliance. A structured approach that provides an efficient and effective way for the storage, management, 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 respective compliance. Such requirements also shape the information exchanged between entities and are a part of the interoperability model. We explore the exchange of such information in greater detail through the information flows between various entities in the following section.

Information flows

Each interaction point has requirements from multiple GDPR articles that affect the information and activities associated with that point. This is presented in Table 1 with the relevant articles in GDPR and their relation towards governing the interoperability between entities. An extended version of the table is available online (Pandit, Debruyne, O'Sullivan, & Lewis. 2018) and presents a more granular reference to GDPR articles along with comments describing the relevance to interoperability.

Table 1 contains four types of statements identified in the text of the GDPR that determine or influence the interoperability of information between entities. The first type of statement reflects a requirement for the interoperability and is abbreviated as *REQ*. Entities are expected to follow or fulfil this requirement for compliance. GDPR only states but does not stipulate how a requirement should be fulfilled. Where an activity or action is presented in the statement, these are identified as processes related to usage, sharing, publication, or exchange of information, and are annotated as *PROC* in the table. Where information is categorically mentioned or as information consisting of some form or category, the abbreviation *DATA* is used to identify such statements in the table. Where additional information about category or type of data is specified, this is annotated with *FORMAT*, with the statement either specifying an explicit data format or providing guidelines governing the choice of formats which are acceptable or need to be enforced.

Where these requirements might not have a direct bearing on the processes and the data involved, they are useful towards the discussion involving the abstract concept of the associated data. These form the background of the requirements gathering process for processes, including communication between entities, where the compliance of a requirement or the implementation of a process might guide the available standards for representing the data involved. For example, in Article 30-1, the statement requires controllers to maintain logs or records of processing activities. While this statement refers to the abstract information associated with processing activities, it can also be used to interpret and formulate records of activities into a structured form of information useful towards discussing standardisation of the associated data. In the next section, we identify and explore this abstract notion of information from the selected articles outlined in Table 1 by categorising them based on their content and intended usage.

Categorising Information Flows

The information associated with information flows can be categorised based on its context and intended usage into the following five categories which we explore in this paper: Provenance, Agreements, Consent, Certification, and Compliance. The relationship of information categories and their association with the different entities is presented in Table 2. It describes the role each entity plays for the corresponding information category, as well the interoperability with another entity. For

Table 1. Interaction points between entities in GDPR with type of statement

Article	Interaction Point	Type(s)
5	DS -- DC, DC -- SA	REQ, PROC
7	DC -- SA, DS -- DC	PROC
12	DS -- DC	REQ, PROC, DATA, FORMAT
13	DS -- DC	DATA
14	DS -- DC	DATA
15	DS -- DC	DATA
16	DS -- DC	REQ, PROC
18	DS -- DC	REQ, PROC
19	DS -- DC, DC -- DC, DC -- DP	REQ, PROC, DATA
20	DS -- DC, DC -- DC	REQ, PROC, DATA, FORMAT
25	DC -- SA	PROC
26	DC -- DC	REQ, PROC
27	DC -- SA	REQ, DATA, FORMAT
28	DC -- DP, DP -- DP	REQ, PROC, DATA
30	DC -- SA, DC -- DP, DP -- SA	REQ, PROC, DATA, FORMAT
33	DC -- SA, DC -- DP	REQ, PROC, DATA
34	DS -- DC	REQ, PROC
35	DC -- SA, DS -- DC	REQ, DATA
36	DC -- SA, DP -- SA	REQ, PROC, DATA
42	DC -- SA, DP -- SA	REQ
47	DC -- DP, DP -- SA, DC -- SA	PROC
49	DS -- DC, DC -- SA, DP -- SA	REQ, PROC
57	DS -- SA, SA -- SA	REQ, PROC, DATA
58	DC -- SA, DP -- SA	REQ, PROC, DATA
60	SA -- SA	REQ, PROC
77	DS -- SA	REQ, PROC, DATA

Table 2. Describing the relation between information categories and entities

Category	DS	DC	DP	SA
Provenance	--	Maintain	Maintain	Inspect
Agreements	--	With DC and DP	With DC and DP	Inspect
Consent	Provide	Collect	--	Inspect
Certification	--	Audit	Audit	Provision
Compliance	--	Maintain, Demonstrate, Audit DP Compliance	Maintain, Demonstrate (SA and DC)	Check

example, consent is provided by the Data Subject, is collected by the Data Controller, and is inspected by the Supervisory Authority. We use the information categories to broadly shape and classify the information flows between entities as well as to refer to the information exchanged within them. 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.

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 along 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). Compliance then becomes a matter of introspecting such provenance logs to see whether the activities recorded the correct and compliant behaviour. Another example is for checking whether a 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.

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 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 creating components based on activities (Mense & Blobel, 2017).

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, greatly 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 an 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 the assent or agreement by the data subject in relation to 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 (Ross, 2017). 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 certain requirements which guide the acquisition and demonstration of consent for it to be evaluated as valid (Mittal 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 (Fatema et al., 2017), the entire web-page may need to be preserved to demonstrate that the consent acquisition process was in accordance 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 an 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 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 the 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. Compliance under the GDPR is a continuous process rather than a single operation to be carried at the end of an activity. Instead, it is essential to maintain compliance at all times by ensuring related activities are compliant at all stages of their 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 are compliant before they begin the operation. Several people might be involved in design and operation of the system, but the responsibility of ensuring the compliance falls on the management or on the/a Data Protection Officer (DPO) if appointed. 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. 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 of 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 which 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 criterion 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 with respect of demonstration of compliance to the authorities is still under consideration.

An existing example of such a mechanism is European Privacy Seal (EuroPriSe, 2018) which carries out an audit of an organisation before providing a seal which 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 certain 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 requires the controller and processor to investigate their respective agreements at the end of this report. Agreements therefore needs to consider this process as a requirement which 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 is able to demonstrate the changes in its processes, the reports can possibly 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, the structuring of such information may provide a strong use within the organisation of compliance related information by cross-linking or cross-referencing the information in documentation that can be continuously updated.

Identifying Opportunities for Commonality and Interoperability

The model provides an overview of information exchange in the context of the GDPR. It identifies the relevant entities, their roles and requirements, the categories of information exchanged, and the obligations of the GDPR applicable over these. Through this, we identify information that is common in terms of its requirements for operation and compliance as well as representation in terms of practicality and adoption. Interoperability of information can be discussed using these as attributes and using the model to structure the information flows. The commonality and interoperability of identified information flows and their associated information categories are useful towards discussing how these can be exploited towards the standardisation of information being exchanged.

As seen from the previous descriptions of various information categories, provenance forms an underlying structure where processes or data artefacts can be captured and represented for various use cases of the other information categories. Similarly, the dependence between the different information types also demonstrates the advantages of linking them together to create more efficient systems capable of automation and better documentation. This provides an opportunity to combine the approach towards representing the different types information into a cohesive model that operatives at a higher and more abstract level to represent the entire system's information model. It also highlights the points of interoperability internally within an organisation. While it is still possible to pick and choose which information or category should be represented individually, the overall benefits afforded by a cohesive model are better suited for the functioning of the service and its compliance.

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.

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 structuring of information in a systematic manner 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.

Existing Standards

In this section, we explore existing standards and their relevance with respect to the representation of information and interoperability discussed previously in this paper in the context of the GDPR. Our focus is primarily on standards for information being exchanged over the medium of the internet. This is due to its increasing prevalence for information exchange as well as provision of services. Where applications and services are not provided over the internet, they either depend on it for communication (including messages and updates) or internally use it as the medium for services with other parties (such as analytics). We consider approaches both within industry as well as academia, as well as organisations and bodies involved in creating and overseeing standards. We intentionally emphasise on standards that are open and non-proprietary due to their greater usability and freedom of adoption by the community at large.

World Wide Web Consortium (W3C)

The World Wide Web Consortium (W3C, 2018), 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 provision of services and information, it is important to consider standards that can be readily integrated into mediums such as web pages and web services which form the backbone of interoperability for many organisations, both commercial as well as public institutions.

W3C standards undergo various stages of development starting from "Working Draft (WD)" to "Candidate Recommendation (CR)" which are then moved to the "Proposed Recommendation (PR)" stage before being set as a "W3C Recommendation (REC)". Due to the continuous participation of the stakeholders and the community at each stage of development, the standards at stages PR and REC are considered to have been sufficiently matured to be adopted into usage. We therefore consider only those standards which fall in either stages as being suitable for recommendation in this paper.

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, 2014), or RDF, is a family of specifications that were originally defined as a metadata model but have 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 data modelling in a consistent manner.

The Web Ontology Language (OWL, 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 a large interest in the academic as well as commercial community, and there are several public ontologies, with notable examples found in the library and medical domains. For querying information declared using RDF, there are mechanisms such as SPARQL (SPARQL, 2013) 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 (SHACL, 2017) which is a W3C Recommendation, and the Shape Expressions (ShEx, 2017) language, which is currently being drafted by the W3C community.

To take advantage of the interoperability offered by commonly used formats such as CSV and JSON with the semantics provided by RDF, there is significant work in creating a standard combining these approaches. Notable examples for this include CSV on the Web (CSVW, 2016) which uses CSV and JSON-LD (JSON-LD, 2014) which uses JSON. Reusing (and in this case combining) standards provides interoperability as well as commonality towards the underlying technology utilised to create, store, and query information represented by these standards. Therefore, any additional standards or formats developed for application-specific approaches should be based on existing forms of standards in order to take advantage of existing practices and adoption of technologies. This line of argument is consistent with the recent uptake of open-data publishing requirements by the European Publications Office (Eur-LEX, 2018) using mechanisms based on RDF and open data formats. In the next sections, we discuss W3C standards and approaches for modelling and representing the various information categories discussed in this paper.

For representing provenance, we have the Provenance Data Model (PROV-DM, 2013), or PROV, which is a W3C recommendation since 30th April 2013 and provides definitions for interchange of provenance information. Using PROV, we can define entities and the various relations and operations between them such as generated by, derived from, and attributions. PROV has been successfully utilised in several domains and applications including encapsulation of scientific workflows and provenance repositories. 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 in approaches specific to the representation of provenance information related to GDPR (Pandit & Lewis, 2017).

The Open Digital Rights Language (ODRL, 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 meet by stakeholders. In addition, policies may 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 representing consent as an agreement between the Data Controller and the Data Subject.

CEN / CENELEC / ETSI

The European Committee for Standardization (CEN, 2018), or CEN, is a public standards organization consisting of thirty-four national members that work together to develop European Standards (ENs) in various sectors. CEN is officially recognised as a European standards body by the European Union. The other official European standards bodies are the European Committee for Electrotechnical Standardization (CENELEC, 2018) and the European Telecommunications Standards Institute (ETSI, 2018). Together, these standardisation bodies provide a large framework of interoperable standards that aim to foster free trade and public benefit. Previously, there was a significant overlap and between the CEN and ISO standards, with some of them presenting potentially conflicting resolutions. This was rectified by the Vienna Agreement where CEN and ISO agreed to avoid duplication of standards. This has resulted in CEN adopting a number of ISO standards which have superseded or replaced existing CEN standards. An example of this is CEN ISO/IEEE 11073, which is a standard for medical and health device communications. It enables communication between medical, health care and wellness devices and with external computer systems for automatic and detailed electronic data

capture of client-related and vital signs information, and of device operational data. Following such standards allows easier operations between multiple entities, as well as for supervisory authorities to assess its workings. This is important when considering that data obtained via such devices can be considered to be sensitive personal information under the GDPR, and therefore will have additional obligations regarding its collection, usage, storage and sharing. Using standards for data collection and communication allows compliance to be assessed based on known mechanisms part of implementing the standard.

ISA²

The Interoperability solutions for public administrations, businesses and citizens (ISA2, 2018), 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 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 (SEMIC, 2018), or SEMIC, that 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 so as 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, in order to reduce the development costs of cross-border and/or cross-sector e-Government systems.

Interoperability and Right to Data Portability

The Right to Data Portability provided by the GDPR allows a Data Subject to request a copy of their personal data from the Data Controller within the timeframe of one month (or two months for demonstrably complex requests). It also allows Data Subjects to request their personal data to be transferred directly from one Data Controller to Another. The data provided under this right must support re-use, and therefore must be provided in a format that is structured, commonly used, machine readable, and is interoperable. It is to be noted that GDPR does not specify any specific data formats that satisfy these requirements. Therefore, it is up to the Data Controller to ensure that the data format chosen to provide data satisfies these requirements. Recital 68 encourages Data Controllers to develop interoperable formats that enable data portability, but without an obligation for controllers to adopt or maintain processing systems which are technically compatible. Therefore, a Data Controller have no requirement to investigate which data formats are supported by other Data Controllers (or their Data Management systems). Clarifications provided by the WP29 (Article 29 Data Protection Working Party, 2017) state that Data Controllers do not have an obligation to support a particular data format used by another Data Controller to provide data. Direct transfer from one Data Controller to another

is only possible when the receiving Data Controller has technical systems that can accept the specific data format used. WP29 also comments on the data formats used specifying that being structured, commonly used, and machine-readable are specifications for the data format with interoperability being its desired outcome. Further comments by WP29 on Recital 68 state that the aim of data portability is to produce interoperable systems and not compatible systems. Interoperability in this case is defined by ISO/IEC 2382-01 as "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units."

Based on this understanding of the requirements for the data provided under the Right to Data Portability, we undertook a short study of the various data formats used by online social services to provide personal data. We selected these services based on their popularity and used the specific mechanisms provided by the service itself to exercise our right to obtain a copy of our personal data. The study was undertaken only on the services used by the primary author of the paper. The obtained data was analysed for the specific data format used and whether it satisfies the specifications laid out by the GDPR. These data formats were then analysed to evaluate whether they are based on existing standards and support for interoperability as defined by the WP29 guidelines. The purpose of the study was to understand the provision of information and the specific standards used in its representation. The study provides information regarding how data is exchanged in the industry and allows an analysis of the model in terms of existing standards and representations. We present here our report on these findings.

We analysed the personal data obtained from following organisations: Apple, Facebook, Fitbit, Google, Instagram (owned by Facebook), LinkedIn, Snapchat, Twitter, and WhatsApp (owned by Facebook). The requests were made in the first three weeks of June. In all cases, the services offered an online interface to request a copy of the personal data. In most cases, the data was provided on the same day, with the maximum time taken to provide being 5 days from the day of request. All organisations provided the data as an archive using the Zip file format, except Fitbit which provided a link to a Dropbox folder that could be accessed without a Dropbox account and contained the requested data. CSV, HTML, and JSON were the most common formats used to provide (generic) data, with other formats such as VCF, vCARD, iCalendar, MBOX used for specific data such as calendars and emails. Table 3 lists more information about the data formats used by organisations.

To assess whether these data formats satisfy the requirements set forth by the GDPR is a matter requiring legal expertise and an authoritative interpretation of the law. Here, we evaluate them

Table 3. Data formats used for data obtained under the right to data portability

Organisation	Archive Format	Data Format(s)	Response (in days)
Apple	Zip	CSV, PDF, VCF, ICS	3
Facebook	Zip	HTML, Images, JSON	0
Fitbit	Dropbox link	XLS, TSV	5
Google	Zip, Gzip	HTML, iCalendar, vCard, Document formats, JSON, CSV, MBOX…	0
Instagram	Zip	JPEG, JSON	2
LinkedIn	Zip	CSV	0
Snapchat	Zip	HTML, JSON	0
Twitter	Zip	JavaScript	2
WhatsApp	Zip	HTML, JSON	3

from a technical perspective, and structure our argument on the points set forth by WP29 regarding interoperability. The most common data formats, HTML, CSV, and JSON, are all considered to be commonly used, structured, and machine-readable. However, using HTML is not suitable for provision of data due to being largely a data presentation and markup language. Therefore, any extraction of information from HTML would need an introspection of the entire document, which is not suitable for interoperability. While CSV and JSON can be easily read and consumed by automated systems, some manual oversight is still needed to evaluate what the data itself actually is. For example, data in a single row of a CSV file can be interpreted only as raw information, as there is no indication of what the data represents. Although the first row may provide values that act like column headers to describe the data, these are for human-consumption, and cannot be interpreted by machines readily. Similarly, data provided as JSON is structured and machine-readable, but cannot be consumed without first understanding its structure.

Different providers might represent similar data in different structures, requiring human action in the creation of systems that can interpret this information. For example, two different organisations providing data regarding purchases might use different structuring of information in their respective CSV files. One might provide the item cost first and then the item name, while the second might reverse this order to list the item name first and cost after. Even though the data contained within both CSV files might be exactly the same, their interpretation differs due to the difference in structuring due to a lack of context regarding the information. Therefore, although the data can be consumed, its usability depends on its interpretation which may differ between different providers. This potentially satisfies the WP29 guidelines regarding creating interoperability between systems since the data is indeed interoperable with additional actions for interpretation. However, the usability of information can be enhanced through the use of contextual metadata that can assist in the interpretation of given information for greater interoperability between systems.

Semantic Interoperability

SEMIC (and EIF) defines "Semantic Interoperability" as the preservation of meaning in the exchange of electronic information (SEMIC, 2018). In the context of an information exchange, the receiver and the sender of information can understand and interpret it the same way. Semantic interoperability is achieved through the establishment of common agreements on the meaning of different entities exchanged in the context of the information. 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 common agreement for the provision of context. Such context can be represented as metadata describing the system and providing information such as the type of data or what it represents. The use of contextual metadata to augment the personal data provided under the Right to Data Portability will lead to such semantic interoperability between the systems.

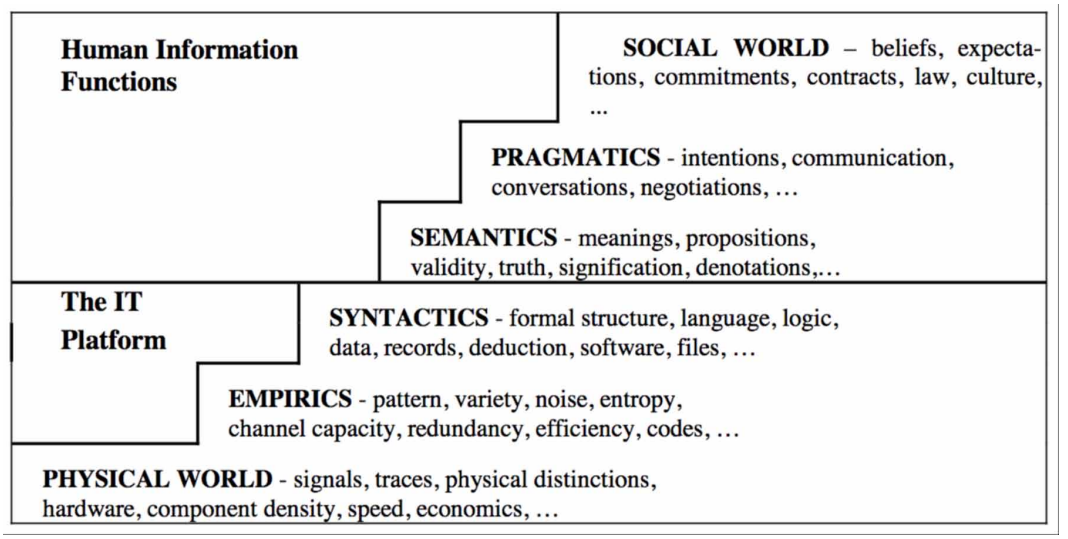
While GDPR only requires that the provided data (and by extension, their representative systems) be interoperable, these requirements can be further expanded upon towards using semantic interoperability. WP29 in its guidelines observes that where there are no common formats used within a particular domain or context, the 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. WP29 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.

One possible solution towards this is using data formats based on existing data formats with support for adding contextual metadata. Examples of this are the CSV on the Web data format which

augments the CSV data format, and JSON-LD data format which allows encoding linked data for mappings between JSON and RDF. 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 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), visualised in Figure 2, 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. Adopting these would enable better interoperability between the systems in terms of requesting the data from different providers. An example of this is requesting a user's profile information from different providers. Profile information in this case contains personal information such as name and email as well as information such as address and references to other social media accounts. By using a common vocabulary to define these pieces of information, a single query can retrieve the information from multiple services.

There is an existing example of such semantic interoperability on the web through the use of schema.org (Schema.org, 2018), which is a collaborative community effort to create and maintain schemas for structured data for use on the internet. Its primary use is to act as a shared vocabulary used to structure metadata on websites to help search engines understand the content being published. A similar effort needs to be undertaken to define interoperable metadata for content being provided as part of the Right to Data Portability. An initiative by industry giants such as Google, Facebook, Twitter, and Microsoft to create such interoperability has led to the data transfer project (Data Transfer Project, 2018). The aim of the platform is the creation of a common framework with open-source code that can connect online service providers to enable a seamless, direct, and user-initiated portability of data. The architecture of the project describes common data models acting as 'adapters' that can plug into various services to ensure universal interoperability. Such data models explicitly declare the specific metadata useful to the exchange of information based on its context. This is an example of semantic metadata enhancing interoperability through the provision of context.

Figure 2. Stamper's (2016) semiotic ladder for design and use of information systems



CONCLUSION

This paper explored an interoperability model based on entities and obligations within the context of the General Data Protection Regulation (GDPR). The interoperability model was created from an analysis of the GDPR where relevant entities were identified along with their roles and responsibilities in the collection and sharing of information. The exchange of information was represented through interaction points between entities and information flows between them. Relevant articles in GDPR were identified for each interaction point and analysed for their effect on the interaction for legal compliance. The information exchanged through these was categorised and discussed using the five categories of provenance, agreements, consent, certification, and compliance. Commonalities between these categories were identified and used as motivation for exploiting them towards interoperability for organisations. The paper also presented a discussion on the representation of information using existing standards along with various bodies and organisations involved in the standardisation process. To explore the real-world application of the interoperability model, the paper presents an investigation and analysis of the data provided under the Right to Data Portability. Its findings demonstrate that although this data used was interoperable due to structured and machine-readable data formats but was not readily usable due to lack of context regarding its meaning and structure. A solution was presented that uses metadata to provide context for the provided data and enables semantic interoperability.

Through this work, we hope to have presented sufficient motivation for further exploration of the interoperability model for GDPR based on the relation of compliance requirements with information flows between different entities. Along with this, the information categories of consent, certifications, and compliance need further work regarding their usage and representation for interoperability. For this, the paper presents strong arguments towards adopting models such as those based on semantic web technologies, which present an open and extensible framework for representation and sharing of information. Their usage also enables in the creation and adoption of semantic interoperability between various entities and their systems.

ACKNOWLEDGMENT

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.

REFERENCES

W3C. (2018). World Wide Web Consortium. Retrieved from https://www.w3.org/

Article 29 Data Protection Working Party. (2017). Guidelines on the right to data portability.

CEN. (2018). Home - European Committee for Standardization. Retrieved from https://www.cen.eu/

CENELEC. (2018). Home - European Committee for Electrotechnical Standardization. Retrieved from https://www.cenelec.eu/

CSVW. (2016). CSV on the Web: A Primer. Retrieved from https://www.w3.org/TR/tabular-data-primer/

Data Transfer Project. (2018). Data Transfer Project. Retrieved from https://datatransferproject.dev/

DPC. (2018). Home - Data Protection Commission - Ireland. Retrieved from https://dataprotection.ie/

DPD. (1995). Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data. *Official Journal of the European Union, 281*, 31–50.

ETSI. (2018). Home - European Telecommunications Standards Institute. Retrieved from https://www.etsi.org/

Eur-LEX. (2018). Home - EU Law and Publications. Retrieved from https://publications.europa.eu/

EuroPriSe. (2018). EuroPriSe - European Privacy Seal for IT Products and IT-Based Services. Retrieved from http://european-privacy-seal.eu/

Fatema, K., Hadziselimovic, E., Pandit, H. J., Debruyne, C., Lewis, D., & O'Sullivan, D. (2017). 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). Retrieved from http://ceur-ws.org/Vol-1951/#paper-05

GDPR. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union, L*, *119*, 1–88.

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. In Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci) (pp. 55–62). Retrieved from http://ceur-ws.org/Vol-1931/#paper-08

ICO. (2018). Home - Information Commissioner's Office - United Kingdom. Retrieved from https://ico.org.uk/

ISA. (2018). ISA^2 - Interoperability solutions for public administrations, businesses and citizens. Retrieved from https://ec.europa.eu/isa2/

JSON-LD. (2014). JSON-LD 1.0 - A JSON-based Serialization for Linked Data. Retrieved from https://www.w3.org/TR/json-ld/

Mense, A., & Blobel, B. (2017). HL7 Standards and Components to Support Implementation of the European General Data Protection Regulation. *European Journal for Biomedical Informatics*, *13*(1), 27–33.

Mittal, S. (2017). The Role of Consent in Legitimising the Processing of Personal Data Under the Current EU Data Protection Framework.

ODRL. (2018). ODRL Information Model 2.2. Retrieved from https://www.w3.org/TR/odrl-model/

OWL. (2012). OWL 2 Web Ontology Language. Retrieved from https://www.w3.org/TR/owl-overview/

Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2018, July 5). data interoperability in GDPR.csv (Version 2). figshare. 10.6084/m9.figshare.6752798.v2

Pandit, H. J., O'Sullivan, D., & Lewis, D. (2018). GDPR Data Interoperability Model. *Presented at the 23rd EURAS Annual Standardisation Conference*, Dublin, Ireland.

PROV-DM. (2013). PROV-DM: The PROV Data Model. Retrieved from https://www.w3.org/TR/prov-dm/

RDF. (2014). Resource Description Framework. Retrieved from https://www.w3.org/RDF/

Reuben, J., Martucci, L. A., Fischer-Hübner, S., Packer, H. S., Hedbom, H., & Moreau, L. (2016). Privacy Impact Assessment Template for Provenance. In *2016 11th International Conference on Availability, Reliability and Security (ARES)* (pp. 653–660). IEEE. doi:10.1109/ARES.2016.95

Ross, H. (2017). Data subject consent: How will the General Data Protection Regulation affect this? *Journal of Data Protection & Privacy*, *1*(2), 146–155.

Schema.org. (2018). Home - Schema.org. Retrieved from https://schema.org/

SEMIC. (2018). Semantic Interoperability Community - Joinup. Retrieved from https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic

SHACL. (2017). Shapes Constraint Language (SHACL). Retrieved from https://www.w3.org/TR/shacl/

ShEx. (2017). Shape Expressions Language 2.0. Retrieved from http://shex.io/shex-semantics/

SPARQL. (2013). SPARQL 1.1 Query Language. Retrieved from https://www.w3.org/TR/sparql11-query/

Stamper, R. (1996). Signs, information, norms and systems. In *Signs of Work: Semiosis and Information Processing in Organisations* (pp. 349–397).

Steyskal, S., & Kirrane, S. (2015). If you can't enforce it, contract it: Enforceability in Policy-Driven (Linked) Data Markets. In SEMANTiCS (Posters & Demos) (pp. 63–66).

XQuery. (2017). XQuery 3.1: An XML Query Language. Retrieved from https://www.w3.org/TR/xquery-31/

Harshvardhan J. Pandit is a PhD Researcher at the ADAPT Centre in Trinity College Dublin. His area of research concerns how technology can assist in the maintenance, evaluation, and demonstration of compliance towards GDPR through the use of open technologies, such as semantic web.

Christophe Debruyne is a Research Fellow at Trinity College Dublin (Ireland) and is affiliated with the ADAPT Centre. He received his PhD in Computer Science from the Vrije Universiteit Brussel (Belgium) in 2013. His main research interests are collaborative ontology engineering, data governance, and data integration. Dr. Debruyne has a keen interest in industry collaboration to demonstrate and validate his research, and to tackle problems that industry face. He has applied his research in various domains including HR, B2B, GLAM, and geospatial data.