GDPR Data Interoperability Model

Harshvardhan J. Pandit, Declan O'Sullivan, Dave Lewis

ADAPT Centre, Trinity College Dublin, Dublin, Ireland

{ harshvardhan.pandit | declan.osullivan | dave.lewis } @adaptcentre.ie

Abstract: *Laws such as the General Data Protection Regulation (GDPR) specify information to be shared, provided, or communicated between entities as part of its operation and compliance. Its compliance additionally requires provision of information between the entities to follow the requirements and procedures outlined in the text. GDPR also provides explicit cases of data interoperability such as that provided by the right of a data subject to move their personal data between different data controllers. In this paper, we explore the interoperability of information between the various entities mentioned within GDPR. We identify various procedures outlined for information flows which also contain explicit requirements such as presence of structured data or specific data formats being used and provide a discussion of existing standards by evaluating the state of the art with respect to the standards provided by the World Wide Web Consortium (W3C) for representing information.*

Keywords: *GDPR, Interoperability, Data Sharing*

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 in European Union regulated the processing of personal data. This has been superseded by the General Data Protection Regulation, abbreviated as GDPR, which is the new European data protection legislation that enters into force on 25th May 2018 (European Union, 2016). It is an important legislation in terms of changes to the organisational measures required for compliance. In particular, GDPR focuses on the use of consent and personal data and provides the data subject with several rights. These new changes have spurred innovation within the community that targets compliance with the various obligations of the GDPR.

Along with providing several explicit constraints and procedures over the use and sharing of personal data, the GDPR also provides statements about the way information is shared or communicated between various entities. Compared to its predecessor, GDPR stipulates larger transparency and accountability of data being shared, as well as the responsibilities of all parties involved. For e.g. a Data Processor under the GDPR is an entity that can only process the data under the provided instructions from another Data Processor or Data Controller. An agreement between a Data Controller and Data Processor is expected to state the specific responsibilities of the Data Processor, as well as the mechanisms under which the Data Controller can verify or audit the accountability of this process as provided by the GDPR.

While there is no requirement for legally structuring this data in a particular way, doing so has benefits for all entities involved. This includes the personal data given by the data subject to an organisation, which can then be shared with other entities, each governed by a different agreement. Information flows also exist between the data subject and the

supervisory authorities in the form of complaints and their resolutions. We explore such information flows in this paper and explore how they can benefit from interoperability based on the requirements of the GDPR. This follows from our previous work on the creation of the GDPRtEXT (Pandit, Fatema, OSullivan, & Lewis, in-press) resource which provides a linked data version of GDPR text along with a thesauri of concepts.

In particular, we focus on identifying the nature of such information flows to create a data model for interoperability. Our aim in undertaking this project, and in presenting this paper is to present this model and explore how interoperability can benefit the various entities both externally as well as internally, and in doing so, to explore existing standards for adoption towards interoperability. In the following paper, we identify entities and information flows associated between them to identify the categories of information in the context of the GDPR. We then present our discussion on how these can be used to achieve interoperability and its advantages towards compliance as well as providing benefits to the operations of an organisation. We finally conclude the paper with an evaluation of existing standards with a focus on those provided by the World Wide Web Consortium (W3C) due to the prevalence of online services, and discuss how they can be combined towards the objectives of this work.

Entities and Information

An overview of the data interoperability model for GDPR can be seen in Fig. 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 specific information or structured data between them. Understanding the requirements of such communication and the associated information is a good step towards exploring the opportunities for standardisation.

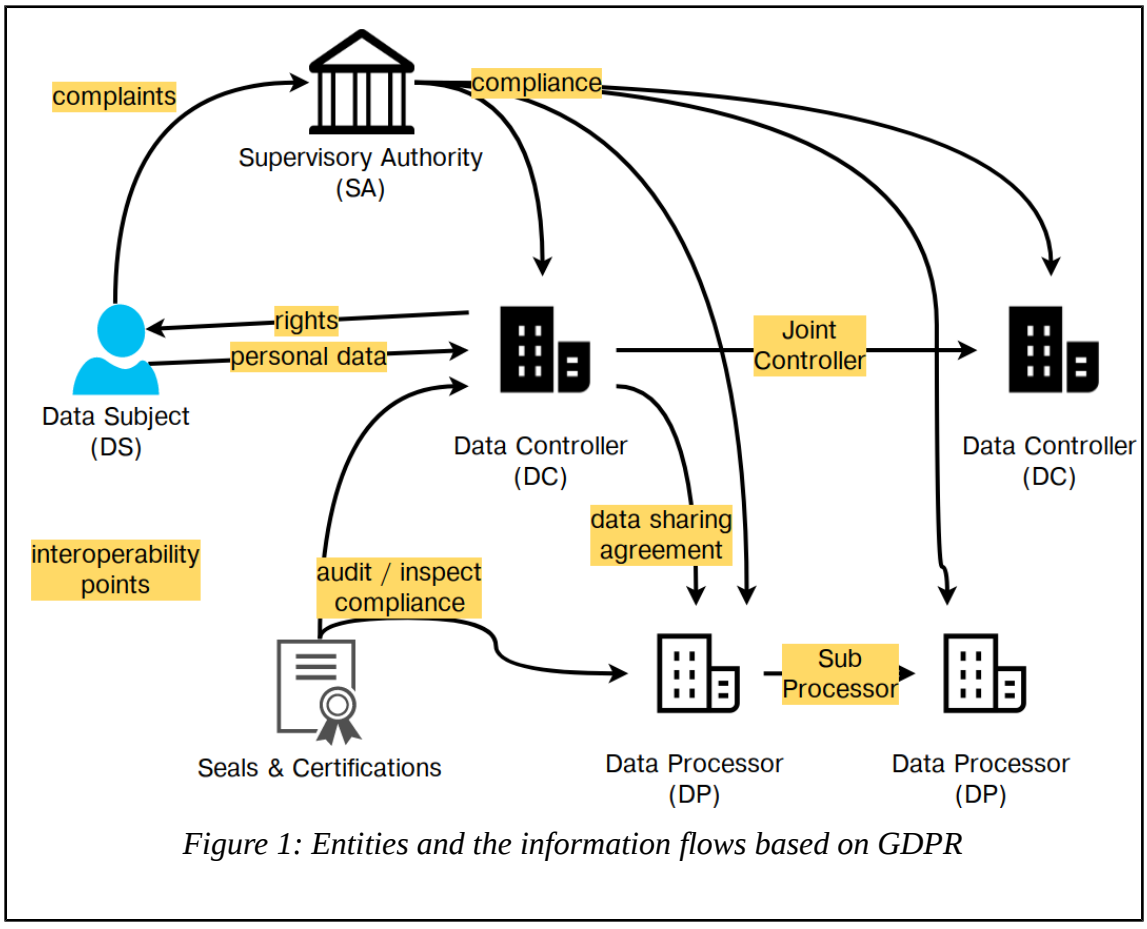

Figure 1: Entities and the information flows based on GDPR

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 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. These are the entities we consider to be within the scope of this paper. In addition, there are other entities mentioned in the GDPR such as - Representative of Controller, Processor, Data Subject, as well as the Regulatory Authority, and Certification Bodies.

Interoperability between Entities

There are several possible cases of interoperability between these entities as reflected by the model in Fig. 1. In this, 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 DS, DC, DP, and SA, we have a set of 6 possible points for interoperability without considering the direction of interaction. Additionally, since controllers and processors can interact with other controllers and processors respectively, and supervisory authorities can interact with other supervisory authorities, this brings the total count of possible points to 9. These are shown in the figure as well as listed in Table. 1 along with examples of information associated between the two entities.

The entities depicted in the model are taken from the text of the GDPR. 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) has no bearing on the interoperability point. Additional information may need to be exchanged based on specific requirements, 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 entities are commercial entities, interoperability is more concerned with consistency, structure, and correctness of information being exchanged.

Consider the interactions between Data Subjects and Data Controllers or Data Controllers and Data Processors, where the interoperability requires only that the provider and consumer or recipient merely provide the required information in an agreed format. This information is not inherently intended to be made available to anyone else (other entity), and therefore has no bound requirements in terms of standards at this point of interaction. Contrast this with the case where the Supervisor Authority, a public institution is involved. Communication with Data Controllers and Data Processors would have to take into consideration the sensitivity of private information and therefore would require the use of

secure forms of communications which may also require security in the information structuring itself, such as through encryption or establishment of secure channels. Any warning or ruling by the SA that can be considered public information, as in made available to the public, would need to be also published in an appropriate manner. The current norm for this is for the SA to publish details of use-cases along with their rulings or decisions on its website. Such information in the future might be collated in a registry or dataset using appropriate formats and structuring.

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

The interaction between Data Subjects and Data Controllers is one of the major points of concern addressed by the GDPR. The interoperability between these entities involves the DS providing personal data to DC, which will be in whatever form the DC offers to accept.

However, DS also provide consent to the DC, which needs to follow certain guidelines stipulated by the GDPR regarding compliance which can affect the way consent is collected and stored. Though this puts no restriction on how the DC obtains this consent from the DS, the onus is on the DC to ensure that the obtained consent satisfied the obligations of the GDPR for demonstrating the validity of such consent. Therefore, it would be prudent for DC to obtain or convert the consent in a form that makes this process of compliance easier.

The interaction from DC to DS includes the provision of information regarding services and rights as mandated by GDPR. DC also has to provide information for rights such as right to data portability which allows DS to receive a copy of their personal data upon request. GDPR has particular clauses regarding the provision of this data, and its structure or format. Additionally, GDPR also provides for DS to get their personal data transferred from one DC to another, which requires both controllers should have some form of interoperability mechanism to send as well as accept the data in a manner that is mutually understandable.

For interactions between DC and DP, or DC and DC, or DP and DP, these points already have some interoperability as part of an organisation's business practice. However, under GDPR, certain additional information may need to be shared in the operation of such services. This provides an opportunity for exploring whether a structured or a common format might provide advantages to existing practices. An approach suggesting an entirely new form of interoperability model would be difficult to uptake due to existing infrastructures. Alternatively, an approach that only considers the information necessary for compliance can be proposed as a solution that augments existing services rather than replaces them.

Interoperability as part of GDPR compliance is primarily outlined by the interactions between the Supervisory Authority and Controllers/Processors. The compliance information reflects the data required to demonstrate and determine the organisation's compliance, which legally can be in any form as long as it contains the correct and sufficient information. For organisations, the process of maintaining, sharing, and demonstrating compliance using this information is a challenge that warrants looking at alternate approaches that can help with these activities. One avenue of possibility is provided in the case where this information is linked to information such as those related to processing activities that are also involved in compliance. Taking advantage of this relation, one can use a structured approach that provides an efficient and effective way for the storage, management, and querying of such related information.

Information flows

Each interaction point can have multiple GDPR clauses that affect the information as well as the processes associated with the information. This is presented in Table. 1 which presents a summary of the relevant articles in GDPR and their relation to interoperability, and is intended to present the GDPR clauses governing the data interoperability between different entities. While the table only shows a summarisation of the information, we have made available the full table online which contains annotations referenced with individual points within Articles and includes additional comments regarding interactions.

The table depicts 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 and this is considered under compliance. GDPR only states but does not stipulate how a requirement is fulfilled. Where an activity or action is presented, 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 and has 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 which specifies an explicit data format mentioned in text or some guidelines governing the choice of formats which need to be enforced.

While requirements and processes might not have a direct bearing on the data and specific formats used to carry out a particular activity, these are an important point of discussion as they present an abstract concept of the associated data. These form the background of the requirements gathering process for any such communication 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 more structured information which would be useful to discuss standardisation of such records. In this section, we identify and explore such types or categories of abstract information from the selected articles outlined in Table. 1.

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	Ask for consent	--	Inspect
Certification	--	Audit	Audit	Provision
Compliance	--	Ongoing, Demonstrate, Check (DP)	Ongoing, Demonstrate (SA and DC)	Check

Categorising Information Flows

For the current paper, we only consider certain information categories and explore them in more detail. These are - provenance, agreements, consent, certification, and compliance. The entities and interactions involving these information categories are listed out in Table. 2 which describes the information associated with each information category and the entities involved.

Provenance

The provenance information category, as defined by provenance, 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 functioned correctly, which requires the presence and maintenance of logs that record the functioning of these activities. These logs can be modelled as provenance in which case they form the lifecycle 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 introspection such provenance logs to see whether the activities recorded the correct and compliant behaviour. A deeper example is that of checking whether a consent was validly given, which requires that the consent be freely given, specific, and unambiguous. Since detecting 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 specific consent. Depending on the manner of representing provenance, the lifecycle 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 format 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 lifecyles 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 DC and DP, or DC and DC, or DP and DP. The form of agreements needs to take into consideration that they can change depending on factors such as change in consent and rights being exercised. Therefore, exploring the use of smart agreements (Steyskal & Kirrane, 2015) that can be interacted with in an automated manner to certain extent would benefit systems where a large part of the system operates on a similar level of automation. For example, if a controller receives an instruction from a data subject to update their consent for certain activities which are handled by a processor, the controller must update or enforce (depending on the legal term in use here) their agreement to get the processor to also reflect this consent over the personal data and activities that they have/had received from the controller. Without some form of automation, such requests would need to be sent and received manually, greatly increasing the work and time required to handle them. With automation involved in the process, the controller’s system can automatically take care of the request by updating the agreement in place for handling the particular consent and personal data with the processor, and can also await for a receipt from the processor for successful completion of the request. Such agreements that can be iterated, stored, and queried using systems are of benefit to the involved parties 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 other entity), and can therefore benefit from the same approach for implementing the data sharing agreements. This provides consistency in technology as well as encourages 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 deemed to be 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 somehow 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 dependant 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 lifecycles 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. While the interpretation of the law may vary, 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 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 Data Protection Officer (DPO) if appointed. In any case, such checks of compliance are integral to audits by the organisation itself or a third-party hired by the organisation for ensuring proper legal requirements are met by their activities. 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.

Such information of compliance related activities can be represented as provenance though the processes and artefacts involved in this case are different from those related to the consent and personal data lifecycles. 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 lifecycle where the outputs of activities such as various 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 lifecycles to be consistent in its purported use-case with those related to provenance of consent and personal data lifecycles.

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 EuroPriSe (European Privacy Seal, n.d.) 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

As seen from the previous descriptions of various information categories, provenance forms a strong underlying structure for all categories where processes or data artefacts can be captured and represented for various use cases. 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, we think 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 lifecycles 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 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.

W3C Standards

In this section, we explore existing standards and their suitability towards interoperability of information specific to the requirements set forth in the previous analysis of the GDPR. We consider all approaches - including research and academia, commercial, bodies involved in maintaining standards - with the caveat of the standard being open and non-proprietary. We argue that the use of such open standards fosters better community participation and adoption. For the scope of this paper, we restrict our exploration of standards to those published by the World Wide Web Consortium (W3C) as it is the main standards body responsible for information exchange on the internet. As the exchange of information is most likely supported and occurs through the medium of the web, these standards can be readily integrated into mediums such as websites 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)”. It is usually considered that standards at stages PR and REC are sufficiently matured to be adopted into usage as they rarely have any significant changes. We therefore consider only those standards which fall in either stages as being suitable for recommendation in this paper.

For representing provenance, we have the Provenance Data Model i.e. PROV-DM (PROV-O: The PROV Ontology, n.d.), 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, abbreviated as ODRL (ODRL Information Model 2.2, n.d.), is a proposed recommendation 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 well as the representation of consent.

For representing the information itself, W3C has several standards with regards to the structuring and organisation of information such as RDF and OWL along with data formats such as XML. For querying information, there are mechanisms such as SPARQL and XQuery/XPath that operate on standardised forms of data (RDF and XML respectively). For validating the structure of information, Shapes Constraint Language (SHACL) can be used as a standardised form for validating RDF. There are approaches based on commonly utilised formats such as CSV on Web for CSV and JSON-LD for JSON that combine the understandability of such formats with the standardised representation of information such as

that in RDF. Reusing such standards provides a form of interoperability in the form of underlying technology utilised to create, store, and query the information. 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 line with the recent uptake of open-data publishing by the European Publications Office using mechanisms based on RDF and open data formats. Additionally, these existing standards can be used to provide data related to the right of data portability accorded by the GDPR that requires organisations to provide a copy of personal data in a format that is structured, commonly used and machine-readable.

CEN / CENELEC / ETSI

The European Committee for Standardization (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) and the European Telecommunications Standards Institute (ETSI). 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.

ISA²

The Interoperability solutions for public administrations, businesses and citizens (ISA²) programme 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), 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 that is currently under development. 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.

Conclusion

Through this paper we have explored the information flows influenced by the General Data Protection Regulation (GDPR) with a goal towards a data model for interoperability. We identified the various entities involved in such information flows and the nature of their relationships with respect to information interoperability. This was done using an analysis of the text of the GDPR where relevant article for each entity and their associated information flow were identified and published in this paper. This was then used to categorise the relevant articles in GDPR based on their effect on each interaction point. We explored five information categories in this paper, which were provenance, data sharing agreements, consent, certification, and compliance along with the dependencies between them. We presented ideas and motivations for exploiting the commonality in these information categories with a view towards interoperability at both internal and external levels for organisations. For representing this information, we evaluated W3C standards based on maturity and recommendation for the identified information categories.

Through this work, we hope to have presented a sufficient motivation for further exploring the interoperability model and the relation of information flows between different entities. We also would like to further work on information categories such as consent and compliance which have not seen much exploration in terms of interoperability and standardised representation. While there is no legal requirement to adopt standards for exchange of data, doing so has several benefits for the involved entities. Primarily, it allows a common language for exchange of data, which reduces the burden for developing targeted solutions for different organisations. A common technological base can help in the creation and adoption of community-wide frameworks that foster interoperability. Additionally, such commonality in data would also help the supervisory authorities as well as data subjects in understanding and acting on given data.

Acknowledgements

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

References

European privacy seal. https://www.european-privacy-seal.eu/ . (Accessed: 2018-03-26)

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 5th workshop on society, privacy and the semantic web - policy and technology (PrivOn 2017). Retrieved from http://ceur-ws.org/Vol-1951/#paper-05

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

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 information model 2.2. https://www.w3.org/TR/odrl-model/ (Accessed: 2018-03-26)

Pandit, H. J., Fatema, K., OSullivan, D., & Lewis, D. (in-press). GDPRtEXT – GDPR as a linked data resource. The Semantic Web - 15th International Conference, ESWC 2018, Crete, Greece, June 3 - June 7, 2018. https://2018.eswc-conferences.org/wp-content/uploads/2018/02/ESWC2018_paper_135.pdf

Pandit, H. J., & Lewis, D. (2017). Modelling provenance for gdpr compliance using linked open data vocabularies. In 5th workshop on society, privacy and the semantic web - policy and technology (privon2017) (privon). http://ceur-ws.org/Vol-1951/#paper-06

PROV-O: The PROV ontology. https://www.w3.org/TR/prov-o/ (Accessed: 2018-03-26)

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, May 4). Official Journal of the European Union, L119, 1–88. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

Reuben, J., Martucci, L. A., Fischer-H¨ubner, S., Packer, H. S., Hedbom, H., & Moreau, L. (2016). Privacy impact assessment template for provenance. In Availability, reliability and security (ares), 2016 11th international conference on (pp. 653–660).

Ross, H. (2017). Data subject consent: How will the general data protection regulation affect this? Journal of Data Protection & Privacy, 1 (2), 146–155.

Steyskal, S., & Kirrane, S. (2015). If you can't enforce it, contract it: Enforce ability in policy-driven (linked) data markets. In Semantics (posters & demos) (pp. 63–66).