Reconciling Data Protection Rights and Obligations: An Ontology of the Forthcoming EU Regulation

Cesare Bartolini
University of Luxembourg
cesare.bartolini@uni.lu

Robert Muthuri
University of Turin
robert.kiriinya@unito.it

Abstract

Knowledge theory has made its way into modern computing, through the use of models and annotations to organize it. The bottom layer of knowledge organizations makes use of ontologies, which are models based on a formal language structure and designed to express the concepts pertaining to a domain and the relationships between them. The use of ontologies is popular also in the legal domain to organize legal documents and as a support to legal reasoning. A legal topic which is currently under the limelight at the European level is data protection. Under the pressure of the last years' technological developments, the data protection legislation has shown its weaknesses, and is currently undergoing a long and complex reform that is finally approaching its completion. The reform will urge businesses dealing with personal data to comply with the new Regulation. The aim of the current paper is to provide a basic ontology for the upcoming data protection legislation, highlighting the duties of the data controller, to ease the transition of systems and services from the existing legislation to the new one.

1 Introduction

The goal of privacy and data protection domains of law is to protect the personal information of the individuals in a given jurisdiction. While businesses have a legitimate interest in appropriating personal data as information assets to achieve their business goals, they should also comply with regulatory requirements particularly on client and employee accuracy and security of their information. However, this information is often subject to abuse. With the advent of social media and the soon to be Internet of Things, people are generating even more content on various platforms. Accordingly, businesses are continually developing methodologies and tools to exploit these valuable assets such as machine learning, big data analytics and natural language processing techniques.

Legislators therefore enact data protection laws to secure proper information handling procedures. The traditional concerns associated with data protection include identity theft, fraud and deception. However, the application of the foregoing technologies to user-generated data to profile potential clients for advertising overshadows the other concerns. However, the enactment of privacy and data protection laws around the world results in a complex patchwork which may compromise the integrity of personal information for many individuals while also jeopardizing many business opportunities.

This is why the European Union (EU) is in the process of upgrading the current data protection law, which is based on the so-called Data Protection Directive (DPD) [1] to a more modern and uniform legislation (Reding, 2010). The reform, which is being developed since 2010, is made up of two main legislative documents: a General Data Protection Regulation (GDPR), and a Directive for the exchange of personal data in criminal investigation. The former document constitutes the basis for the general protection of personal data. While a Directive may be implemented differently in different countries, a Regulation immediately becomes enforceable in all member states in a uniform way.

The reform process is underway, and although the new legislation is in its final stages, it will not be in force before 2018. The text of the GDPR is not yet finalized, and the latest official version released by the Commission dates back to early 2012[2], although some versions containing the amendments of the Parliament and the Council have either been published or leaked to the general public.

One significant concern in the data protection reform is that it will introduce high fines against data controllers who do not comply with the Regulation, and inquisitory powers of the Data Protection Authorities (DPAs). In other words, a DPA

[1] Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

[2] EU Commission Document COM(2012) 11 final of 25 January 2012.

will be entitled to verify if the data controller complies with the data protection rules, and issue fines when it does not. Therefore, businesses dealing with personal data, especially Small and Medium Enterprises (SMEs), will have a compelling need to fulfill all the requirements of the GDPR. On the other hand, most of these requirements are expressed in vague and uncertain terms – one above all, the "appropriate technical and organisational measures" for security (Article 30 of the draft Regulation) – making it difficult for the data controller to know the exact extent of its obligations.

In this paper, we propose a basic ontology of the data protection domain in the context of the GDPR[3]. The purpose of this work is to provide a base structure to identify the scope and extent of the obligations of the data controller, especially in relation to the rights of the data subject. The ontology will be used as a basis for future research in the compliance of the business process of the data controller with data protection rules.

2 Related Work

Ontologies are nowadays widely used in knowledge bases and the Semantic Web (Berners-Lee et al., 2001) as a means of expressing concepts from the legal domain in a formal structure (Benjamins et al., 2005). Legal ontologies are generally created to describe a legal system or norms (Breuker et al., 2005), however they can express a number of different perspectives, from general knowledge to specific domain terminology.

As mentioned in section 1, the data protection reform will put a pressure on data controllers to be compliant with the GDPR (Mikkonen, 2014). However, achieving compliance is not an easy task, given that the legal text is not clear with regards to the actual requirements. The transition of organizational and technical measures adopted by businesses would be eased by the existence of standards to adopt, and auditing companies to verify the adherence to those standards. However, no significant standards currently exist for data protection, much less addressing it in the light of the upcoming reform.

An alternative idea is to use security standards as a substitute of data protection standards. While in law security and data protection refer to two distinct domains, there is no doubt that some overlapping exists between the two. In particular, this is reflected by the fact that some provisions in data protection legislation required that the data processing be performed under appropriate security measures. In computer science, data protection (often called "privacy", a term which creates some confusion with the legal concept of the same name, mainly elaborated by the American doctrine, in particular (Prosser, 1960) and (Bloustein, 1964)) is considered as a subdomain of security: see for example (Pfleeger and Pfleeger, 2006) and (Massacci et al., 2003). An early-stage research (introduced in (Bartolini et al., 2015)) aims at evaluating the overlapping between the GDPR and security standards, such as the ISO 27000 family, and in particular ISO 27001:2013 (Int, 2013), to measure the degree of coverage of the data protection rules a security standard would cover. This would help a data controller who adopts a widespread security standard (which relies upon many years of expertise and consolidated auditing firms and methodologies) better understand what is required on their part to achieve GDPR compliance.

A suitable ontology for the protection of personal data does not seem to exist yet. An attempt to build an ontology was made in the context of the NEURONA project (Casellas et al., 2010). However, there are several problems that make it unsuitable for the purposes of the current research. The ontology itself is not publicly available because it was developed in the course of an industrial project; it is focused on the Spanish national data protection law; and it does not address the point of view of the duties of the data controller.

Another interesting approach is presented in (Rahmouni et al., 2010). However, that work is not focused on the obligations of the data controller, but rather on expressing the legal norms using an ontology to enforce access control policies.

Additionally, no existing work specifically addresses the data protection legislation in the light of the reform.

3 Ontology

An ontology specification represents a given level of consensus in a particular community. For the data protection domain, this includes some basic data protection principles which we should represent in the ontology. Although the legislation differs between the various countries, and even between the different Member States of the EU, the data protection principles are the outcome of many years of evolution of human rights. The data protection principles have been established over the years by the Council of Europe (CoE)[4], then evolved by the EU, both in legislation and in the decisions of the European Data Protection Supervisor (EDPS) and the national Data Protection Authorities.

3.1 Noteworthy Concepts

The data protection principles serve as the foundation for our ontology. It is from these concepts

[3] In the official Commission version of 2012. Although subsequent texts present numerous differences with the 2012 version, the core definitions, principles and rules are unchanged.

[4] Convention 108 of 1981.

that we define the obligations of the data controller while contrasting them to the rights of the data subject. The following is an enumeration of the principles as classified under the European data protection handbook (Eur, 2014):

- data must be processed lawfully, and in particular
  - in compliance with the European Convention on Human Rights by the CoE, and in particular with Article 8 of the Convention ("Right to respect for private and family life");
  - in compliance with the Charter of Fundamental Rights of the European Union, specifically with Article 8 ("Protection of personal data");
- personal data must be processed for specified and lawful purposes, and not for other purposes which exceed the stated ones (principle of purpose limitation);
- personal data must be collected according to some criteria, minimizing the impact on the data subject (data quality principles), and in particular:
  - data must be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed;
  - data must be accurate and, where necessary, kept up to date;
  - data must be deleted after they are no longer necessary for the specified purpose;
  - personal data must be collected only to the extent where anonymous data, or data which do not allow identification of the data subject, are inadequate for the purpose of the processing (principle of data minimization[5]);
- personal data shall be processed on the basis of a fair relationship with the data subject, and in particular:
  - the data subject is entitled to be informed at all times about the processing of personal data and how the data are being used;
  - the processing should be accessible to the data subject, not performed in secret, and documented. The data subject should also be put in the condition to exercise other rights granted by the law (e.g., the right to object, the right to rectification);
- personal data are processed under the liability of the data controller, who must implement all measures, both technical and organizational, to ensure that no harm is caused to the data subject by the processing.

3.2 Methodology

Two premises are in order. First, the work is solely based on the European draft Regulation for a data protection reform. It does not refer to data protection principles that may exist in Member State legislation (and that would integrate the GDPR if compatible), nor the outcomes of the decisions of courts or DPAs. Only the GDPR is taken into account, in a form which is subject to changes in the final text.

Secondly, the purpose of this work is not to define a model of the legal text (although to some extent this is inevitable), but rather to model the requirements and duties that the controller must meet by enacting appropriate measures to be compliant with the legislation.

The first step in creating the ontology was the choice of the approach, i.e., what perspective the ontology is supposed to address (as described by (Breuker et al., 2005)). From this point of view, the structure of the ontology presented here is the result of the combination of two different approaches:

- the skeleton has been derived from (Eur, 2014), with only a few slight modifications to adjust to the upcoming legislation;
- the details are based on the long-term research focus described in section 1, highlighting the obligations of the data controller and (when possible) matching them with the corresponding rights of the data subject. In this sense, the ontology is structured in a way similar to the Hohfeldian model (Hohfeld, 1917).

The ontology for the EU data protection legislation was created using the Protégé 5 software and the OWL/XML language (Antoniou and van Harmelen, 2004). A graphical depiction of the ontology[6] is shown in Figure 1.

3.3 Ontology Development

In developing this ontology, we follow the approach suggested by (Noy and McGuinness, 2001). The first step is to define and limit the scope of the ontology as we have already described in the foregoing section. This could be refined by the development of competency questions. These are questions that help to delimit the representation of a particular area of the domain or particular level

[5]This principle is not present in the Directive and has been developed in the following years in the German and Italian legislation.

[6]Note to reviewers: the actual ontology has not been uploaded to an ontology database, but the current version can be retrieved at `https://drive.google.com/open?id=0B9l91sx_gYBXSjFhSVdaZGNqZjA`.

of detail (Grüninger and Fox, 1995). Not only do the answers given help describe the application of the ontology, they also help clarify its scope and domain. As indicated in section 1, this work anticipates the impact that the new Regulation is likely to have on firms particularly the SMEs once it enters into force. For our data protection ontology, the following are possible competency questions:

- What are the obligations of a data controller?
- What are the functions of a data processor?
- What are the rights of the data subject?
- How do the rights of the data subject relate to the obligations of the data controller and the functions of the processor?
- How can a data subject interact and/or enforce their rights against a data controller?
- What are the possible fines and sanctions issued in response to violations by a data controllers?
- Who supervises a data controller?

The next step is to determine the hierarchy that the ontology will take. The Data Protection Handbook (Eur, 2014) provides a high-level partitioning of the European data protection into general principles, founding rules (constituting most of the duties of the data controller), and rights of the data subject. Since the rules and the data subject's rights are applications of the general principles of data protection, they have been defined as subclasses of some principle.

On the other hand, the purpose was to highlight the duties of the data controller, not only at a general level but in terms of what are the requirements for the processing of personal data to be legitimate.

Following the first perspective, and according to (Eur, 2014), the ontology has been divided into three main areas:

- the data protection principles;
- the rules of data processing;
- the data subject's rights.

The first area contains the principles as defined above, and they represent a high level of abstraction, general concepts that are the result of decades of evolution in data protection legislation and constitute the basis of modern data protection laws. The second and third areas describe the single provisions of the GDPR, which are expressed either as data processing rules (duties of the data controller), or as rights of the data subject. These represent more low-level concepts that must be applied to data processing and that can be directly verified by a DPA.

The ontology is designed in such a way that every data processing rule or data subject's right is an application of one of the data protection principles. Some associations are straightforward, while others are derived from (Eur, 2014).

The obligations of the data controller contain some general provisions which apply to any form of personal data processing. For example:

- the processing must be based on one of the legal grounds of Article 6, such as the data subject's consent, a legal obligation of the controller, and so on;
- the processing must be performed in compliance with the Regulation and additional normative sources such as codes of conduct;
- prior to the processing, the controller must perform an risk impact assessment.

The grounds for processing (Article 6 of the Regulation) have been structured in detail, also by clarifying the differences in consent depending on whether it is:

- an ordinary consent concerning non-sensitive data pertaining to the consenting data subject;
- a consent to the processing of sensitive data pertaining to the consenting data subject;
- a consent to the processing of a data subject who is legally represented by the consenting person.

Additional provisions (Articles 40–45) concern the transfer of personal data to third countries, which is a processing activity in itself, as per the definition contained in Article 4(3). Therefore, additional rules, applicable only to the transfer of personal data, have been included in the ontology.

In addition to complying with the aforementioned obligations, the data controller must ensure that the processing allows the data subject to exercise the rights to which he or she is entitled by the law, such as the right to access the data, to request their erasure (a new right introduced by the GDPR but not present in the DPD), or to request the rectification of incorrect or outdated data.

Also, for the purposes of our research it is important to relate the rights of the data subject with the corresponding duties of the data controller. Some of the outcomes of such an approach are the following:

- the data subject has a right to be informed about the processing. This is achieved by

means of transparent information and communication that the data controller must provide (Article 11 of the regulation), as a requirement for the processing to be lawful. Providing transparent information is also a means of establishing trust between the data controller and the data subject, which in turn is a facet of the fairness principle;

- for the data subject to enact a right to access, the data controller has a corresponding duty to provide some means to request access to the data. To exercise the right, the data subject must perform a single access, which is defined as a subset of the right to access, and is bound by a relationship with the data for which access is requested;

- the right to object is similarly structured: the data subject can object to the processing of personal data, and the objection (a subset of the right to object) is related to a specific processing. The relationship is a functional property, called *isObjected*, defined in the domain of Processing. This property is also used to define the lawfulness of the processing, because personal data cannot be lawfully processed if the data subject has exercised the right to object.

The granularity of the ontology is still quite coarse. Following the last example, there is currently no means of expressing whether the grounds on which the objection of the data subject is based (Article 19 of the draft Regulation, in connection with Article 6(1)), or the "compelling legitimate grounds" on which the processing can be carried out in spite of the objection. Such a degree of detail can be significant in a judicial perspective, but not in the scope of the current research. The complexity of the data protection legislation is such that a complete coverage would be excessive for the current scope.

Some of the concepts expressed in the ontology remain vague because they are expressed as such in the law, and are not fit for direct usage. Samples of such concepts are the "appropriate safeguards [...] in a legally binding instrument" for the lawfulness of a data transfer (Article 42), or the "appropriate technical and organisational measures to ensure [...] security" (Article 30). These concepts are fluid and must be coordinated with knowledge from other fields than data protection, such as general contract law or the state of the art and best practices in data security. Standards for computer security can partly fill these gaps, so understanding the relationship between them and the GDPR would be key to a fast transition to the new legislation once it enters into force.

4 Application Examples

(Mommers, 2001) identifies a number of ways legal ontologies may be applied, ranging from information systems to knowledge-based systems. We extend those applications to actualise our competency questions. We envision the data protection ontology as being suitable for the following uses.

4.1 Information retrieval

The encoding of the meaning of concepts and the relations among them empowers users of information retrieval systems. Data controllers and their processors will be able to determine what their duties against the rights of the data subject are. We facilitate this by linking data processing principles to the processing rules of personal data belonging to a data subject. We also encode the hierarchical relations that for instance demonstrate the extra properties mandatory for sensitive data over and above those of personal data. This will help a data processor comprehend such information about concepts relevant to his duties.

4.2 Transition from Directive to Regulation

The EU is currently made up of 28 Member States, each with its own domestication of the DPD, whereas the GDPR will introduce a homogeneous legislation. Making the meaning of legal terms in the new GDPR explicit could help compare the impact of the new legislation on the existing national regimes. For instance, in the United Kingdom (UK), international data flow is currently treated as a data protection principle (UK Data Protection Act 1998, Schedule I, Art. 1(8)), while in the draft regulation it is not.

4.3 Translation of Legal Documents

Related to the transition is the fact that even though the Regulation will have a uniform application in all the Member States, it may need to be implemented in different languages according to the official language of each State. For that nation's transition to the new regime a data protection ontology might be crucial, as it may function as a uniform base upon which the domestication process is grounded.

4.4 Automated classification and summarizing

Alongside information retrieval, automated classification is meant to facilitate finding documents. Ontologies, combined with statistical techniques and natural language processing techniques, can support classification as well. In the context of Big Data, such techniques would help a company mine all documents that, for instance, contain sensitive data, involve transfers to foreign recipients, have

proper consent from their respective data subjects, or for which data subjects have previously objected to the processing. This also applies to automatically building summaries of such documents for managerial review or compliance purposes when reporting to a DPA.

4.5 Question answering

Automatic questions answering requires thorough representation of knowledge in order to let a system "understand" both the question and the source of knowledge on which automatic answering is based. For instance, with the help of a data protection ontology, a data controller could implement a knowledge system that could help data subjects make queries related to their rights, the personal data being processed, the purposes of the processing, the accuracy of the data, or the retention period. A DPA could also integrate the ontology into a knowledge system, to provide services such as:

- for data subjects, to determine their rights and remedies in case of breaches and violations;
- for data controllers, to understand their obligations;
- for data processors, to understand their functions.

4.6 Decision support and decision making

Legal (procedural) regulations often contain decision structures that allow to take certain decisions or qualifications. Although such structures can be modelled in relatively simple decision trees, such decision trees still require user intervention to make a choice in each step. An ontology can be used to encode not only the decision steps, but also the content of the decision rules.

For now, the data protection ontology only models the contents of the decisions that a data controller would need to make. The obvious next stages will involve modelling the possible decision steps that would assist a data controller in handling personal data. This may involve reconciling such steps with the data and process models of the data processors acting on behalf of the controller. Advantages of using an ontology in such a case will include the consistency of the models and the reusability of the underlying ontology for other modelling activities. This would also come in handy for the internal audits by the Data Protection Officer and demonstrating compliance to a DPA.

4.7 Agent technology

Agents (Jennings and Wooldridge, 1998) are assumed to allow intelligent autonomous communication between different computer systems. For such communication, the modelling of rules governing that community is necessary. Of particular interest here is the intersection between data protection and computer security domains. Security in data protection, an expression of the accountability principle according to (Eur, 2014), requires the secure processing of personal data (Art. 30–32 of the GDPR). However, the relation of this requirement to security standards from the security domain (as discussed in section 1 above) still has to be thought on. As in the case of decision support and decision making, such modelling can be supported by an underlying ontology.

5 Conclusions

In this work, the authors have structured an ontology for personal data protection, with two main objectives: to emphasize the duties that the data controller must fulfill by enacting appropriate solutions, and to address the upcoming data protection reform.

The ontology presented here is a preliminary step in a work which is at its early stages. It is a very simple ontology, almost naive, leaving out many aspects of the legal text. However, it does not aim at expressing the GDPR, as a whole or in part, from a normative point of view. The provisions that contain duties for the data controller, and matching rights for the data subject, have been selectively identified and built into the ontology. It will act as a starting point which was necessary to pursue the long-term goal of verifying compliance with the GDPR.

This ontology is by definition a work in progress, because it will have to be adapted to the changes in the legal text when a final version of the GDPR is released. However, the unfinished status of the new legislation does not mean that the current work is too preliminary, because the final text will not drift significantly from the current ones, at least with respect to the core principles, data subject's rights and data processing rules. Additionally, the work done here is also subject to change as new requirements emerge from the long-term goal. The ontology will be subject to corrections, and also to refinements in case the current structure turns out to be too coarse for the long-term objectives.

This work will proceed in several research directions. On one side, it will be necessary to develop, possibly identifying existing solutions in literature, a methodology to formally express the compliance with the GDPR, by stating that the data controller fulfills, or does not fulfill, certain requirements. Another mandatory research direction will be to develop a similarly-structured ontology for security standards, so that the two can be compared. An appropriate comparison methodology will also be required.

6 Acknowledgments

This paper is supported by the Joint International Doctoral (Ph.D.) Degree in Law, Science and Technology (Last-JD) coordinated by CIRSFID, University of Bologna, and by the Luxembourg Privacy Cluster (LPC) project at the University of Luxembourg.

References

[Antoniou and van Harmelen2004] Grigoris Antoniou and Frank van Harmelen. 2004. Web ontology language: OWL. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems, chapter 4, pages 67–92. Springer Berlin Heidelberg, second edition.

[Bartolini et al.2015] Cesare Bartolini, Gabriela Gheorghe, Andra Giurgiu, Mehrdad Sabetzadeh, and Nicolas Sannier. 2015. Assessing IT security standards against the upcoming GDPR for cloud systems. In *Proceedings of the Grande Region Security and Reliability Day (GRSRD) 2015*, pages 40–42, 3.

[Benjamins et al.2005] V. Richard Benjamins, Pompeu Casanovas, Joost Breuker, and Aldo Gangemi, editors. 2005. *Law and the Semantic Web*, volume 3369 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.

[Berners-Lee et al.2001] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific American*, 284(5):28–37, 5.

[Bloustein1964] Edward J. Bloustein. 1964. Privacy as an aspect of human dignity: An answer to dean prosser. *N.Y.U. Law Review*, 39(6):962–1007, 12.

[Breuker et al.2005] Joost Breuker, André Valente, and Radboud Winkels. 2005. Use and reuse of legal ontologies in knowledge engineering and information management. In Benjamins et al. (Benjamins et al., 2005), chapter 2, pages 36–64.

[Casellas et al.2010] Núria Casellas, Juan-Emilio Nieto, Albert Meroño Antoni Roig, Sergi Torralba, Mario Reyes, and Pompeu Casanovas. 2010. The neurona ontology: A data protection compliance ontology. In *Proceedings of the Intelligent Privacy Management Symposium*, pages 34–38, 3.

[Eur2014] European Union Agency for Fundamental Rights, 2014. *Handbook on European data protection law*, 4.

[Grüninger and Fox1995] Michael Grüninger and Mark S. Fox. 1995. Methodology for the design and evaluation of ontologies. *Proceedings of the 1995 International Joint Conference on AI, Workshop on Basic Ontological Issues in Knowledge Sharing*, 8.

[Hohfeld1917] Wesley Newcomb Hohfeld. 1917. Fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal*, 26(8):710–770, 6.

[Int2013] International Organization for Standardization, 2013. *ISO/IEC 27001 – Information technology – Security techniques – Information security management systems – Requirements*, second edition, 10.

[Jennings and Wooldridge1998] Nicholas R. Jennings and Michael J. Wooldridge, editors. 1998. *Agent Technology: Foundations, Applications, and Markets*. Springer Berlin Heidelberg.

[Massacci et al.2003] Fabio Massacci, Marco Prest, and Nicola Zannone. 2003. Using a security requirements engineering methodology in practice: The compliance with the italian data protection legislation. Technical Report DIT-04-103, University of Trento, 11.

[Mikkonen2014] Tomi Mikkonen. 2014. Perceptions of controllers on eu data protection reform: A finnish perspective. *Computer Law & Security Review*, 30(2):190–195, 4.

[Mommers2001] Laurens Mommers. 2001. A knowledge-based ontology of the legal domain. In *Second International Workshop on Legal Ontologies, JURIX*.

[Noy and McGuinness2001] Natalya F. Noy and Deborah L. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory, 3.

[Pfleeger and Pfleeger2006] Charles P. Pfleeger and Shari Lawrence Pfleeger. 2006. *Security in Computing*. Prentice Hall, Upper Saddle River, NJ, USA, fourth edition, 10.

[Prosser1960] William Lloyd Prosser. 1960. Privacy. *California Law Review*, 48(3):383–423, 8.

[Rahmouni et al.2010] Hanene Boussi Rahmouni, Tony Solomonides, Marco Casassa Mont, and Simon Shiu. 2010. Privacy compliance and enforcement on european healthgrids: an approach through ontology. *Philosophical Transactions of the Royal Society A*, 368(1926):4057–4072, 9.

[Reding2010] Viviane Reding. 2010. The upcoming data protection reform for the european union. *International Data Privacy Law*, 11.

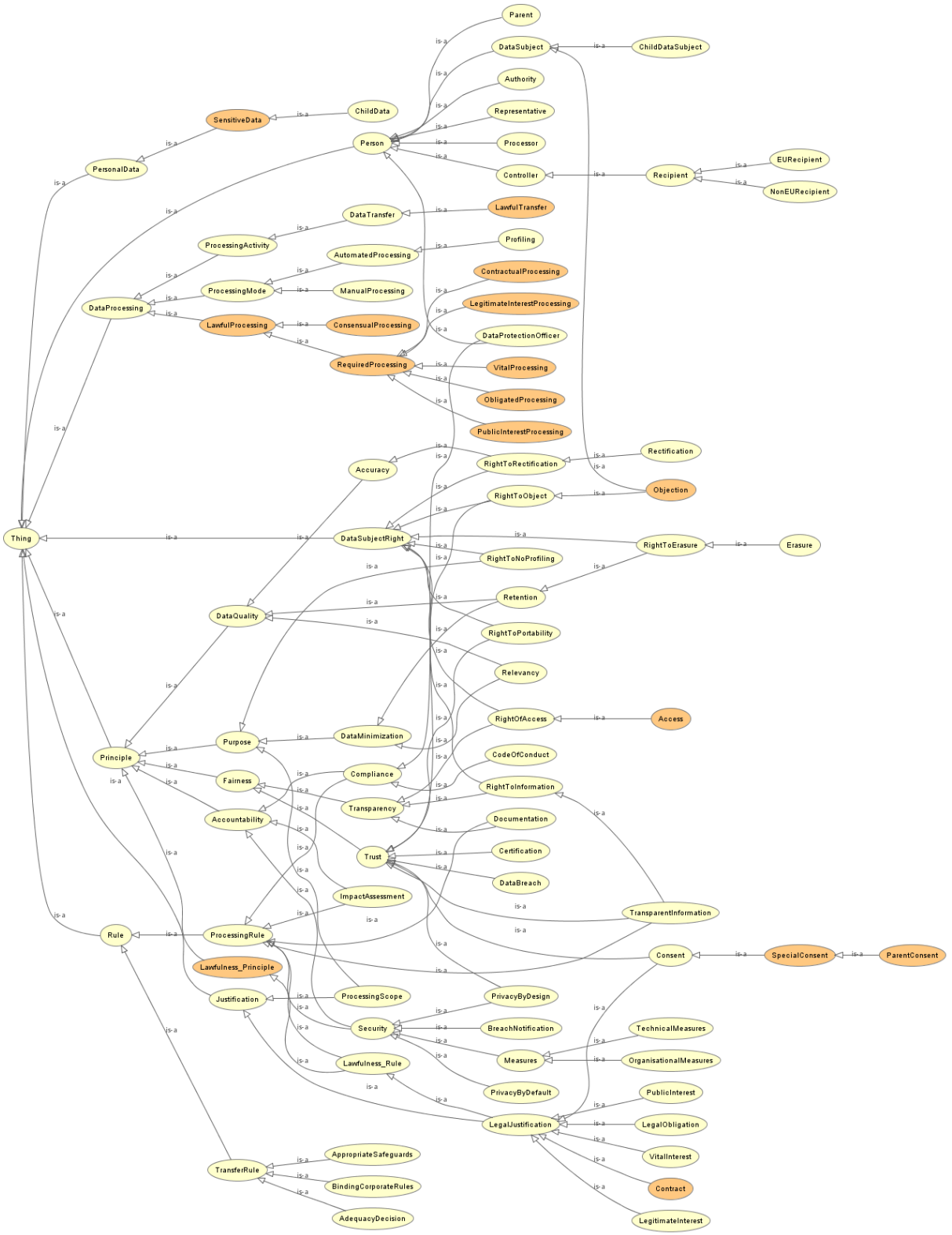

Figure 1: Schema of the data protection ontology.