Privacy-by-Design in Intelligent Infrastructures

Knockaert, Manon; Laurent, Maryline; Malina , Lukas; Matulevicius, Raimundas; Petrocchi, Marinella; Seeba, Mari; Tang, Qiang; Tasidou, Aimilia; Tom, Jake

Published in:
Deep diving into data protection

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Privacy-by-Design in Intelligent Infrastructures

Manon Knockaert[1], Maryline Laurent[2], Lukas Malina[3], Raimundas Matulevičius[4], Marinella Petrocchi[5], Mari Seeba[6], Qiang Tang[7], Aimilia Tasidou[8], Jake Tom[9]

Introduction

For the European Union, Intelligent Infrastructure Management involves Smart Infrastructures that comprise several operators from different domains of activity, such as energy, public transport, or public safety[10]. A smart infrastructure means "(…) an interconnected sensing network that provides real-time digital information about the state of the system[11]"[12]. Consequently, "intelligent structures have controls that are operated automatically, with additional sensor-enhanced capability to

[1] This work has been done with the financial support from the European Union's Horizon 2020 research and innovation program under Grant Agreement n° 830892 (SPARTA) and the Ministry of the Interior of the Czech Republic under Grant VJ01030002.The publication only reflects the opinions of its authors and neither the European Commission nor the Ministry of the Interior of the Czech Republic can be held responsible for the use which could be made of it.

University of Namur, Faculty of Law, CRIDS/NaDi. The author would like to thank Jean Herveg and Michael Lognoul for their precious collaboration.

[2] TelecomSudParis.

[3] Brno University of Technology.

[4] University of Tartu, Institute of Computer Science.

[5] IIT-CNR, Pisa.

[6] University of Tartu, Institute of Computer Science.

[7] Luxembourg Institute of Science and Technology (LIST).

[8] TelecomSudParis.

[9] University of Tartu, Institute of Computer Science.

[10] https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/smart-infrastructure.

[11] R. Morimoto, "Estimating the benefits of effectively and proactively maintaining infrastructure with the innovative smart infrastructure sensor system", *Socio-economic Planning Sciences* 44(4), 2010, pp. 247-257.

[12] R. Ogie, P. Perez and V. Dignum, "Smart infrastructure: an emerging frontier for multidisciplinary research", *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction*. 1-9. 10.1680/jsmic.16.00002, 2017, p. 9.

adjust operations to suit user needs in real time; while smart structures are simply intelligent structures that provide a broader range of automated services that can scale gracefully to better adapt to both user and environmental conditions[13]. In other words, the main distinction is that, unlike intelligent structures that are reactive in exercising their control functions, smart structures are more adaptive, able to better handle issues of fragmentation and interoperability in the use of information and robust enough to dynamically adjust its built form to accommodate changes in use as well as environmental conditions[14]"[15].

As advocated by the European Union Agency for Cybersecurity ("ENISA"), "smart infrastructures rely on remote management of resources, and deploy and operate "cyber-physical systems" that are made up of data-controlled equipment which interacts with the physical world. They collaborate and exchange data under several schemes, depending on their level of maturity"[16]. Intelligent Infrastructures in which IoT technologies are encompassed receive particular attention from the European institutions. As stated by the Article 29 Working Party (now replaced by the European Data Protection Board): "The Internet of Things (IoT) is on the threshold of integration into the lives of European citizens (…). Already today, connected devices successfully meet the needs of EU citizens on the large-scale markets of quantified self and domotics. The IoT thus holds significant prospects of growth for a great number of innovating and creative EU companies, whether big or small, which operate on these markets"[17]. This technology is an infrastructure in which billions of sensors embedded in a lot of devices (such as home robots, smartphones, etc.) are designed to collect and process a lot of (personal and non-personal) data and, as they are associated with unique identifiers, interact with other devices or systems using networking capabilities[18].

In its agenda for Europe, the President of the European Commission, Ursula von der Leyen insisted on the opportunities offered by the IoT:

[13] M.B. Hoy, "Smart buildings: an introduction to the library of the future", *Medical Reference Services Quarterly* 35(3), 2016, pp. 326-331.

[14] A.H. Buckman, M. Mayfield and S.B.M Beck, "What is a smart building?", *Smart and Sustainable Built Environment* 3(2): 2014, pp. 92-109.

[15] R. Ogie, P. Perez and V. Dignum, "Smart infrastructure: an emerging frontier for multidisciplinary research", *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction*. 1-9. 10.1680/jsmic.16.00002, 2017, p. 11.

[16] https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/smart-infrastructure

[17] Article 29 Working Party, Opinion 8/2014 on the Recent Developments on the Internet of Things. WP223, 16 September 2014, p. 3.

[18] Article 29 Working Party, Opinion 8/2014 on the Recent Developments on the Internet of Things. WP223, 16 September 2014, p. 4.

"The Internet of Things is connecting the world in new ways. After knowledge and people, it is now physical devices and sensors that are linking up with each other" [19].

In addition to the issue of securing intelligent infrastructure systems, Internet of Things poses several challenges for the protection and preservation of personal data. For example, data processing may be invisible to data subjects. Indeed, they may not be really aware of the data being used, the various processing and the potential consequences. Furthermore, the IOT is also characterized by a multitude of actors and stakeholders in the development process that has an impact on the increasing number of processing of personal data and the exchange of information. In addition, another important challenge is the fact that we are facing a miniaturization of these connected objects. They are smaller and smaller and easily transportable, which lead to geo-tracking and profiling activities.

While technology entails risks for data protection, it can also become a key tool in the preservation of personal data and compliance with the legal framework surrounding the processing of personal data. In this respect, the General Data Protection Regulation (GDPR) places technology at the heart of data management. The architectural design of the IoT system and the different algorithmic operations must integrate in themselves the guarantees of data protection, at all stages of the processing of the personal data (from the collection to the deletion or anonymization after a specified retention period) [20]. To this end, privacy-enhancing encryption technologies and privacy-enhancing computations enable encrypted data utilization, ensuring that data remain protected from information leakages in case of data breach. Privacy-enhancing digital signatures and authentication methods usually offer anonymity/pseudonymity to users and ensure certain security features (e.g. non-repudiation) to system providers in various ICT use cases. Furthermore, the privacy-preserving communication techniques can also be used as an additional privacy approach in the communication layer in current networks to prevent unauthorised data analysis. The appropriate combination of such Privacy-Enhancing Techniques (PETs) can then

[19] Ursula von der Leyen, "A Union that strives for more-My agenda for Europe". Available at: https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf, p. 13.

[20] C. DE TERWANGNE K. ROSIER and B. LOSDYCK, « Lignes de force du nouveau Règlement relatif à la protection des données à caractère presonnel », *Journal de droit européen*, 2016, pp. 32-33.

address the various sets of the privacy-by-design and GDPR principles and requirements.

The objective of this contribution is to give an overview of various technologies that could be used to manage personal data when using IoT in Intelligent Infrastructure environment. This paper uses the following structure. Section 2 presents the principles applicable to the management and processing of personal data in Intelligent Infrastructures ("IIs"). Section 3 focuses on the management of Intelligent Infrastructures. Section 4 presents several Privacy Enhancing Technologies to reinforce the protection of personal data. Then, two examples of concrete scenarios are explained, one concerns the management of personal data in parking reservation situation; the other concerns the management and enforcement of privacy policies.

1. Key Data Protection elements in Intelligent Infrastructures

This first section focuses on a particular component of the intelligent infrastructures, which is the use of connected objects. Indeed, IoT technologies may include the processing of personal data and therefore have an impact on the privacy of European citizens. In this paper, we do not detail all the principles governing the processing of personal data, but only those for which IoT technologies pose the most challenges.

According to the General Data Protection Regulation (GDPR), "personal data means any information relating to an identified or identifiable natural person. An identifiable natural person is any individual who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person"[21].

1.1. The privacy by design requirement[22] reinforces the personal data protection principles. This imperative requires the data controller[23] to ensure that the IOT system put in place is compliant with the fundamental

[21] Article 4.1 1) of the GDPR.

[22] Article 25.1 of the GDPR.

[23] According to Article 4.1, 7) of the GDPR, data controller means "he natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data".

principles of personal data protection. Article 25.1 of the GDPR provides that the data controller has to implement appropriate technical and organizational measures, both at the time of the determination of the means for processing and at the time of the processing itself[24]. In other words, it is about thinking the process differently. By an *a priori* integration of legal norms with technical developments, the objective pursued by the European legislator is to reverse a situation where the development of technology precedes the legal constraints[25].

1.2. Purpose limitation. The first principle is the purpose limitation. All personal data may only be processed for specified, explicit, and legitimate purposes[26]. This is the principle of purpose limitation that will allow the data controller to determine, with regards to the purposes pursued by the data processing, the personal data he/she can collect and process, what he/she can and cannot do with the personal data and the duration of the personal data retention. Furthermore, the purposes pursued by the data processing must be explicit and cannot be vague or imprecise. For example, purposes such as "Promote safety and security" or "Provide, improve and develop services" are rejected[27].

In addition, the GDPR prohibits any further processing that is incompatible with the original purpose[28]. To determine the legality of a further processing, the GDPR establishes a list of factors that should be considered (e.g. the existence of a link between the original purpose and the

[24] However, it should be noted that manufacturers of products that are not data controllers are not subject to the privacy by design rule. Nevertheless, the GDPR encourages product manufacturers, service providers and application producers to consider the data protection regulation when developing and designing their products or services (Recital 78 of the GDPR). This provision is only included in a recital and not in a binding rule. However, this recital is important to ensure that all those involved in the design and development of an IoT system consider data protection requirements.

[25] E. Degrave and B. Vanderose, 'Privacy by design et E-gouvernement: un modèle inédit en Belgique", Pyramides, 2014, p. 74; Bygrave, Lee A., Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements (June 20, 2017). Oslo Law Review, Volume 4, No. 2, 2017, p. 106. Available at SSRN: https://ssrn.com/abstract=3035164.

[26] Article 5.1, b) of the GDPR.

[27] C. de Terwangne, « Les principes relatifs au traitement des données à caractère personnel et à sa licéité », *in Le règlement général sur la protection des données (RGPD/GDPR) – Analyse approfondie*, C. de Terwangne et K. Rosier (coord.), Brussels, Larcier, p. 95.

[28] Article 6.4 of the GDPR. Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered incompatible with the initial purposes.

new one, the nature of the data and the existence of safeguards such as encryption and pseudonymization)[29].

The purpose limitation is also crucial for a data controller active in IoT applications and services. On this matter, the Article 29 Working Party states that: "The increase of the amount of data generated by the IoT in combination with modern techniques related to data analysis and cross-matching may lend this data to secondary uses, whether related or not to the purpose assigned to the original processing. Third parties requesting access to data collected by other parties may thus want to make use of this data for totally different purposes"[30]. IoT stakeholders must therefore be vigilant concerning the compatibility test for raw data, extracted data or, displayed data[31].

For example, in the case of a connected vehicle, the data concerning the accelerometer and the gyroscope of a smartphone could be used to detect individuals' driving habits[32]. A test should be made to verify, on a case by case basis, if such a further purpose should be deemed legal or not.

1.3. Minimization. In addition, data controllers have to collect and process only the personal data that is adequate, relevant and necessary to carry out the purpose(s) pursued by the data processing[33]. The legislation provides that the collection of personal data cannot be excessive in relation to the purposes pursued by the data processing and with consideration for the loss of privacy. As mentioned by the European Data Protection Supervisor (EDPS), the principle of necessity implies the need for a combined factual assessment of the effectiveness of the measure for

[29] Article 6.4 of the GDPR. Note that there is controversy about the applicability of these criteria for particular categories of personal data (sensitive data). See the Opinion of the EDPB 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (art. 70.1.b)). CRIDS advices to be careful on this matter because, in principle, these personal data cannot be processed. On this aspect, see J.-M. Van Gyseghem, « Les catégories particulières de données à caractère personnel », *in Le règlement général sur la protection des données (RGPD/GDPR) – Analyse approfondie*, C. de Terwangne et K. Rosier (coord.), Brussels, Larcier.

[30] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 7.

[31] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 8.

[32] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 7.

[33] Article 5.1, c) of the GDPR.

the purpose and whether it is less intrusive compared to other ways to achieve the same goal.[34]

It is possible to select – initially – the data strictly necessary for the accomplishment of the purpose announced and – in a second step – to add personal data in order to ensure and maintain the integrity of the developed system[35]. This possibility consists in balancing data minimization and security. While it is true that security requirements are, for the first time, formally raised as a legal principle within the GDPR[36], it must be emphasized that the law only requires a level of security that is proportional to the risks through the use of appropriate measures. Consequently, the collection of additional personal data is limited to what is strictly necessary to achieve the objectives of the system. It entails a case-by-case analysis depending on the operating of the tool, the nature, and the volume of data as well as the risks to rights and freedoms for the data subject. If the data controller can reasonably consider the collection of additional personal data to verify the identity of the person in order to guarantee a safe processing, article 11 of the GDPR demonstrates the overall philosophy of the data protection law. This provision describes the situation where the identification of the data subject is not required. If the purpose for which personal data is processed does not or no longer requires the identification of a data subject, the data controller is not obliged to maintain, acquire or process additional information in order to be able to identify the data subject[37].

The minimization principle is intrinsically linked to the privacy by default requirement. Article 25.2 states that the data controller has to implement appropriate technical and organizational measures to ensure that, by default, only personal data which are necessary for each specific purpose of the processing are processed[38]. It implies that the principles of data protection have to be considered during the elaboration and conception of any processing system.

---

[34] European Data Protection Supervisor (EDPS), Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit, 11 April 2017, p. 5.

[35] Gürses, Seda and Troncoso, Carmela and Diaz, Claudia. Engineering Privacy by Design Reloaded. Available at http://carmelatroncoso.com/papers/Gurses-APC15.pdf, pp. 10-13.

[36] C. DE TERWANGNE, K. ROSIER and B. LOSDYCK, « Lignes de force du nouveau Règlement relatif à la protection des données à caractère personnel », *J.D.E.*, 2016, p. 21.

[37] Veale, Michael and Binns, Reuben and Ausloos, Jef, When Data Protection by Design and Data Subject Rights Clash (February 20, 2018). International Data Privacy Law (2018) doi:10.1093/idpl/ipy002, p. 12. Available at SSRN: https://ssrn.com/abstract=3081069 or http://dx.doi.org/10.2139/ssrn.3081069.

[38] Note that the privacy by default requirement also applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility.

The data controller must first consider the use of anonymized data. If the use of anonymous data is inappropriate for the purpose for which it is intended, the controller must then consider the possibility of processing pseudonymised data[39]. The GDPR encourages the use of this method because it is a technique to implement the privacy by design and by default requirements[40].

The objective pursued by the legislator is to ensure that, by default, the product, application, or service is preconfigured in a manner that respects the key rules of data protection. In this respect, the product, service, or application must be preconfigured to meet the requirements of Article 5 of the GDPR, mainly the minimization of data used and the access rights, data retention, and data integrity and confidentiality.

As stated by ENISA, "Data protection by default implements the rule to limit the data processing to what is necessary for its purpose, namely the data protection principles of data minimization and storage limitation on the basis of the principle of purpose limitation. Although its main focus is on necessity, Article 25(2) is also linked to other data protection principles, such the principle of transparency, as well as the principle of integrity and confidentiality and the overall security of the processing. (…) data protection by default does not force the deactivation of any lawful processing, but it requires the limitation of the processing to the minimum depending of each specific purpose"[41].

1.4. Anonymization v. pseudonymization. According to the former Belgian Commission for the Protection of Privacy (now replaced by the Data Protection Authority), anonymization must be understood as the "processing that takes identifiable personal data as a starting point and results in data that are no longer identifiable"[42].

Before attesting that he or she works with personal data that have been rendered anonymous, the data controller must verify the risk of

[39] According to Article 4.5 of the GDPR 'pseudonymisation' means the "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

[40] Article 25 and Recital 58 of the GDPR.

[41] ENISA, Recommendations on shaping technology according to GDPR provisions, p. 12. Available at: https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions-part-2.

[42] *Autorité de Protection des Données*, https://www.autoriteprotectiondonnees.be/faq-page/10026#t10026n20918.

re-identification of data subjects. It should be noted, however, that the Article 29 Working Party (now replaced by the European Data Protection Board) is cautious about the possibilities of real anonymization in the light of big data technologies and the large amounts of information available to third parties. The increased capacity for cross-checking information means that the controller runs the risk of considering the data anonymous when it might not be the case[43].

According to Directive (EU) 2019/1024 on open data and the re-use of public sector information, anonymization means the process of changing documents into anonymous documents which do not relate to an identified or identifiable natural person, or the process of rendering personal data anonymous in such a manner that the data subject is not or no longer identifiable[44].

On the contrary, pseudonymization means "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person"[45].

Pseudonymized data remain as personal data and are therefore subject to the GDPR.

1.5. Storage limitation. Personal data cannot be kept and processed for a longer period than necessary to accomplish the purpose for which it has been collected[46]. For each type of data collected and in consideration of the relevant purposes, it is necessary to determine if and for how long the personal data needs to be stored or whether it must be deleted or anonymized.

The data controller must carry out this operation of deletion or anonymization of the data spontaneously, and not at the request of the data subject[47]. Recital 39 of the GDPR suggests that deadlines should be set from the outset by the data controller for the erasure of the personal data or for a periodic check, to ensure that the storage does not exceed what is necessary. By applying the principles of privacy by design and by

[43] See Article 29 Working Party, WP207, 5 June 2013.

[44] Article 2.7 Directive 2019/1024 on open data and the re-use of public sector information.

[45] Article 4.5 of the GDPR.

[46] See Article 5.1 (e) GDPR.

[47] C. DE TERWANGNE, « Les principes relatifs au traitement des données à caractère personnel et à sa licéité », *in Le règlement général sur la protection des données (RGPD/GDPR) – Analyse approfondie*, C. DE TERWANGNE et K. ROSIER (coord.), Brussels, Larcier, p. 113.

default, a technical mechanism can be established whereby conservation of the data automatically ends as soon as the time required to achieve the stated purpose has passed[48].

Particularly in the context of IoT, the period of conservation could be different according to the various stakeholders (e.g. creator of the algorithm, creator of the sensors, vehicle manufacturers). In this regard, the Article 29 Working Party indicated that "This necessity test must be carried out by each and every stakeholder in the provision of a specific service on the IoT, as the purposes of their respective processing can in fact be different. For instance, personal data communicated by a user when he subscribes to a specific service on the IoT should be deleted as soon as the user puts an end to his subscription. Similarly, information deleted by the user in his account should not be retained. When a user does not use the service or application for a defined period of time, the user profile should be set as inactive. After another period of time the data should be deleted"[49].

1.6. Integrity and confidentiality. Both the data controller and data processor are responsible for the security of the processing, and therefore of the system used to process data, in particular for preventing unauthorized access to personal data and to the equipment used for processing such data. The security of the system should also prevent any illegal/unauthorized use of personal data or the equipment[50].

The Article 29 Working Party highlights the balance between security and efficiency in IoT services. Indeed, it states that "The IoT raises several security challenges, namely as security and resource constraints force device manufacturers to balance battery efficiency and device security. In particular, it is not yet clear how device manufacturers will balance the implementation of confidentiality, integrity and availability measures at all levels of the processing sequence with the need to optimise the use of computational resources – and energy – by objects and sensors.

IoT devices and platforms are also expected to exchange data and store them on service providers' infrastructures. Therefore, the security of the IoT should not be envisioned by considering only the security of

[48] C. DE TERWANGNE, « Les principes relatifs au traitement des données à caractère personnel et à sa licéité », *in Le règlement général sur la protection des données (RGPD/GDPR) – Analyse approfondie*, C. DE TERWANGNE et K. ROSIER (coord.), Brussels, Larcier, p. 114.

[49] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 17.

[50] Article 5.1, f) of the GDPR.

the devices but also the communication links, storage infrastructure and other inputs of this ecosystem. In the same way, the presence of different levels of processing whose technical design and implementation are provided by different stakeholders does not ensure the adequate coordination amongst all of them and may result in the presence of weak points that can be used to exploit vulnerabilities. For example, most of the sensors currently present on the market are not capable of establishing an encrypted link for communications since the computing requirements will have an impact on a device limited by low-powered batteries. With regard to the end-to-end security, the result of the integration of physical and logical components provided by a set of different stakeholders only guarantees the level of security provided by the weakest component". [51]

1.7. Transparency. First, personal data must be processed lawfully, fairly, and in a transparent[52] manner. Transparency implies that certain information is provided spontaneously by the controller to the persons concerned by the processing of their personal data[53] (data subjects). Furthermore, information shall be provided in a concise, transparent, intelligible, and easily accessible way[54].

The principle of transparency is crucial in the IoT context. Indeed, the Article 29 Working Party highlights the lack of control for the users and information asymmetry. It states that "As a result of the need to provide pervasive services in an unobtrusive manner, users might in practice find themselves under third-party monitoring. This may result in situations where the user can lose all control on the dissemination of his/her data, depending on whether or not the collection and processing of this data will be made in a transparent manner or not" [55].It adds that "unlike other types of content, IoTpushed data may not be adequately reviewable by the data subject prior to publication, which undeniably generates a risk of lack of control and excessive self-exposure for the user. Also, communication between objects can be triggered automatically as well as by default, without the individual being aware

[51] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 9.

[52] See article 12 of the GDPR

[53] C. DE TERWANGNE, « Les principes relatifs au traitement des données à caractère personnel et à sa licéité », *in Le règlement général sur la protection des données (RGPD/GDPR) – Analyse approfondie*, C. DE TERWANGNE et K. ROSIER (coord.), Brussels, Larcier, p. 91.

[54] See article 12 of the GDPR.

[55] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 6.

of it. In the absence of the possibility to effectively control how objects interact or to be able to define virtual boundaries by defining active or non-active zones for specific things, it will become extraordinarily difficult to control the generated flow of data. It will be even more difficult to control its subsequent use, and thereby prevent potential function creep".[56]

The figure below summarizes the main principles that frame the processing of personal data.

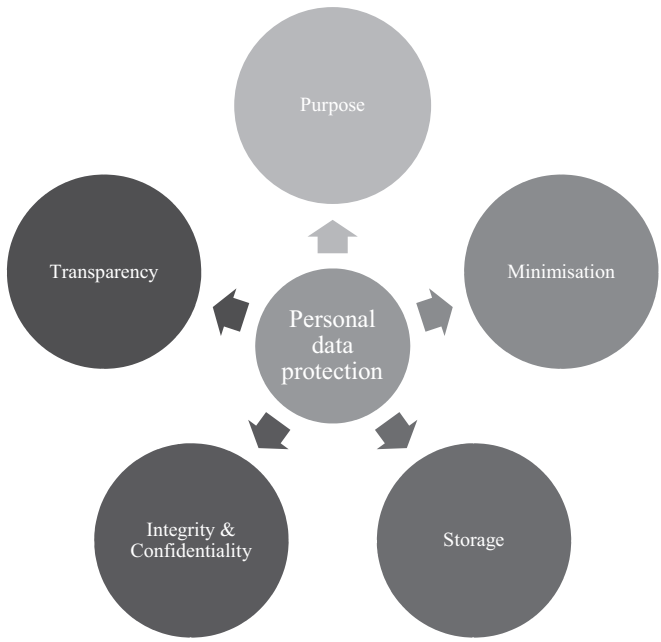

Figure 1 – GDPR principles

[56] Art. 29 Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, 16.09.2014, WP 223, p. 6.

1.8 Graphical GDPR Presentation. Figure 2 presents the extract of the GDPR model[57]. The model is expressed using the Unified Modeling Language (UML) class diagrams. The diagram can be viewed as being composed of three fragments:

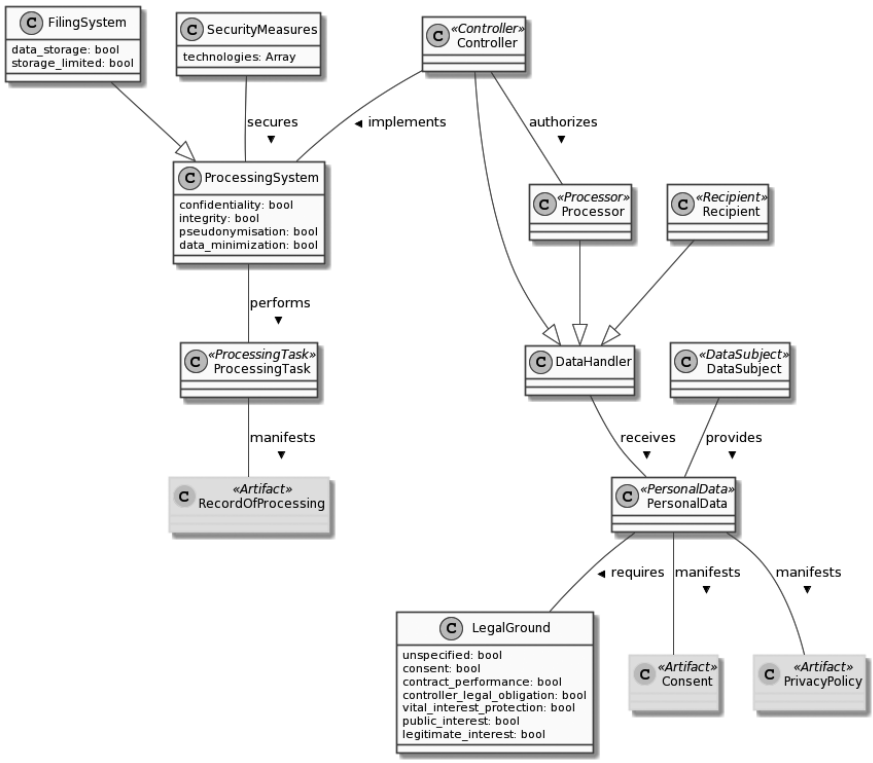

Figure 2 – Extract of the GDPR model (adapted from[58,59])

[57] J. Tom, E. Sing and R. Matulevičius.: "Conceptual Representation of the GDPR: Model and Application Directions. In: Perspectives in Business Informatics Re- search". *BIR* 2018. vol. 330. Springer, 2018); K. Kala, *Refinement of the General Data Protection Regulation (GDPR) Model: Administrative Fines Perspective*. Master's thesis, University of Tartu, 2019.

[58] J. Tom, E. Sing and R. Matulevičius.: "Conceptual Representation of the GDPR: Model and Application Directions. In: Perspectives in Business Informatics Re- search". *BIR* 2018. vol. 330. Springer, 2018.

[59] K. Kala, *Refinement of the General Data Protection Regulation (GDPR) Model: Administrative Fines Perspective*. Master's thesis, University of Tartu, 2019.

GDPR roles: When evaluating a business process, it is important to determine the roles that are represented within the context of article 4 of the GDPR. As such, they can be classified as actors that provide personal data represented by the class DataSubject and those organizational entities who interact with the data in any capacity represented by the generalization DataHandler. The data handlers are specialized into Controllers, Processors, Recipients and ThirdParties (not included into adapted fig 2.).

Legal ground and consent of processing: The second part of the diagram covers articles related to consent and legal ground of processing of personal data represented by the classes PersonalData, LegalGround, SpecialLegalGround (not included into adapted fig 2.). In regard of legal ground, PersonalData is manifested by the presence of a Consent with specific characteristics mentioned in article 7 (Conditions for consent). Personal data is identified as either general or special, depending on categories mentioned in article 9(1). The classes LegalGround and SpecialLegalGround correspond to articles 6 (Lawfulness of processing) and 9 (Processing of special categories of personal data) respectively. It is important to understand the impact that identifying the category of data has on the requirement of consent. If the category of data is identified as general, consent is only needed if the LegalGround class is evaluated to general. In any other case, the requirement of consent can be overridden according to article 6. Similarly, if the category of data is evaluated to anything other than general, the SpecialLegalGround class becomes the deciding factor on whether consent is required. If the data category is evaluated to a special category and SpecialLegalGround is evaluated to general, then consent is required. However, if it is evaluated to anything else, consent will not have to be collected according to article 9.

Processing/filing system, processing task and technical measures: The final part of the GDPR model captures the technical aspects of the processing/filing system used to carry out processing tasks represented by the classes ProcessingSystem, FilingSystem and ProcessingTask. Processing systems must fulfil specific security criteria as described in article 32 (Security of processing) such as confidentiality, integrity, pseudonymization, etc. The distinction between a processing and a filing system is based on whether data is stored using a structured record system such as a relational database. Should the controller be evaluated as using a filing system, it becomes necessary to identify whether there is a clear limit on the duration of data storage captured by the attribute storage limited (article 5(e)). The class TechnicalMeasures corresponds to the articles in the GDPR that mention the necessity for controllers and processors to implement appropriate technical measures to ensure compliance to the

regulation (article 24(1) and article 28(1)). Its singular attribute technologies are marked as an array to indicate that usually, it is a combination of technologies that are to be utilized to fulfil this requirement. At the tail end, we can see that the processing system carries out specific processing tasks represented by the class ProcessingTask. It is necessary to record all processing tasks by maintaining a record of processing activities as put forth by article 30 (Records of processing activities). This is manifested by the artifact RecordOfProcessing.

2. Intelligent Infrastructures Environment

2.1. Key components

As a first insight into the Intelligent Infrastructures environment, this section is dedicated to their functioning and management. The objective is to present how this technology works, what can push a company to embark on such a service and how to ensure an efficient management of II considering the regulatory framework.

Decision making processes of Intelligent Infrastructure. Decision making processes proceed from basic data collection and end up, after traversing a relatively long path, with better decision making. These processes follow a path from basic "collection of data" taken from multiple sources (SCADA systems, customer billing, GPS, Ticketing / counting, social media, sensors, lasar surveys, satellite imagery, BIM/ GIS, Manufacturer's data, CCTV, Scanned images, control systems) and onto the process of data management of a wide range of categories such as various assets, data cleaning, customers, data structures, costs, data storage and various activities. The "sense making"-process consists of modelling, Big Data analysis, Analytics, Data mining, all of which may lead to "improved insight" (or improved intelligence). The final stage involves the actual and crucial decision-making step, which is supported by cutting-edge tools like optimization algorithms, rule-based automation, decision support tools, and machine learning, all of which lead to "improved decisions".[60]

Management of failures and undefined situations. Legacy systems are failure prone and costly to maintain. A key goal for II management

[60] K. Bowers, V. Buscher, R. Dentten, M. Edwards, J. England, M. Enzer, A. K. Parlikad and J. Schooling, "Smart Infastructure: Getting more from strategic assets". Available at: https://www-smartinfrastructure.eng.cam.ac.uk/system/files/documents/the-smart-infrastructure-paper.pdf.

is to decrease costs. Dependability of a system is the ability to avoid service failures that are more frequent and more severe than what is acceptable. Dependability often comes hand in hand with security since cyber-attacks as well as service and device vulnerabilities are frequent causes of service failures, and thus the most common attributes that it encompasses are availability (readiness for correct service provision); reliability (continuity of a service which functions without failures); integrity (absence of improper system alteration, service accuracy/consistency); maintainability (ability of a service to undergo modifications and repairs); and confidentiality (absence of unauthorized disclosure of information). High service availability is also key as it requires exceptional operational performance and maintenance which can ensure long mean time between failures and very short mean time to repair; and safety addresses the absence of serious or catastrophic consequences for the user environment[61].

Reasonable Utilization of Resources. One of the key challenges that organizations face in the management of Intelligent Infrastructures (IIs) is converting from a traditional inter-connect networking topology to a cross-connect topology, which could potentially risk disruption and increased costs. However, as Harel arues[62], the availability of next-generation II products has in great part eliminated this concern by working equally well in cross-connect, inter-connect and "mixed" environments (eg. with copper, fiber or "mixed" cabling). Thus, through effective utilization of resources, organizations can expect to cut operational costs by 20-30 % or more, decrease downtime, optimize power and space utilization, accelerate service deployment and enhance security. In the end, Harel concludes that the utilization of a suitable II next-generation product as a best practice platform, should bring manageability, security, and controllability to the enterprise.

Efficient Intelligent Infrastructure management (Efficiency Model). Next-generation IIs bring efficiency and automation to a broad range of previously-manual tasks. The II concept means the continuous monitoring of a "self-aware"-network, and together with power and environmental apparatuses, as well as a central data repository, it can determine

[61] S. Poledna, "Course: Dependable Computer System". Available at: https://ti.tuwien.ac.at/cps/teaching/courses/dependable-systems/slides/2-DCS-basic-concepts-and-taxonomy.pdf.

[62] T. Harel, "Intelligent Infrastructure Management: The Best-Practice Platform for Data Centers", https://www.datacenterknowledge.com/archives/2014/01/07/intelligent-infrastructure-management-best-practice-platform-data-centers (accessed May 6, 2020).

network status in real time, and intelligent processes can streamline and error-proof operations efficiently[63].

Remote sensors and alerting management. Consider that an IoT platform includes sensor-enabled assets that are connected to a wide range of IoT devices. Sensors allow for the monitoring of the condition of assets and either produce data or actuate responses to data[64].

As Shen describes in detail, the essential building blocks of an intelligent infrastructure framework would consist of an II asset management system that receives data from, say, an IoT platform, and applies business rules to that data and automatically initiates workflow processes. Such an IoT platform would receive location intelligence and other data from sensor-enabled assets, which would standardize, aggregate, and analyse data. It would then, if necessary, alert management and initiate a workflow to trigger the next step in a business process[65].

Manage and maintain goals and value. The principal goal in the context of efficient II management is delivering uninterrupted value as organizations upgrade and evolve their intelligent infrastructures.

Consider the value of the *Efficiency Model*. On the one hand, the value of correct and efficient provisioning lies in the significant reduction in human error, reduction of downtime, increases in productivity, and optimal utilization of resources. Areas affected would involve: Fault Management, IT Asset Management, Security, and Environment and Power Management.

On the other hand, the value of the IoT platform lies in the organization's ability to successfully leverage infrastructure asset management systems and all the data coming from connected devices. In this context, value is generated by the insights arising from utilization of the tools of data analytics, for example, that take into account the methods of data-driven design and leverage the power of the Internet, which is the gateway of value that translates protocols, connects devices, filters and processes data, providing security and performing other critical functions

[63] T. Harel, "Intelligent Infrastructure Management: The Best-Practice Platform for Data Centers", https://www.datacenterknowledge.com/archives/2014/01/07/intelligent-infrastructure-management-best-practice-platform-data-centers (accessed May 6, 2020).

[64] M. Weber and Z. Ivana Podnar. "A Regulatory View on Smart City Services." *Sensors (Basel, Switzerland)* vol. 19,2 415. 21 Jan. 2019, doi:10.3390/s19020415, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6358906/ (accessed May 6, 2020).

[65] Yi Shen, "Create Synergies and Inspire Collaborations Around the Development of Intelligent Infrastructure for Human-Centered Communities", *Journal of the association for information science and technology*, 70(6):596–606, 2019.

to receive data from sensor-enabled assets residing in the IoT platform and intelligent infrastructure asset management framework.

Governance Model for Intelligent Infrastructure Ecosystems. Legal and regulatory framework. As Weber and Žarko suggest, regulatory characteristics are related to legal acts and bylaws relevant to the provision of services. They are predefined by the National laws or EU Directives and Regulations applicable to all EU Member States. The most critical characteristics of a service relate to: i) lawful interception of data traffic, including IoT traffic; ii) service dependability, i.e. the ability to avoid frequent and severe service failures; iii) personal data protection; iv) secure systems that prevent cyber-attacks at the device and service level; v) operator switch, i.e. the ability to change an IoT operator within the value chain; vi) regulated services regarding roaming devices that are registered in one network but used in visited networks; and vii) interoperability and open access to data and services, not only in a technical sense but also as a regulatory requirement[66].

2.2. Parking Reservation Generation Scenario

Figure 3 introduces a process model that captures a parking ticket generation scenario[67]. In this scenario, the vehicle owner first logs into the system using his/her personal device by inputting the system credentials. After logging in, the credentials are verified by the Parking Service Provider (PSP) who responds with a login notification in the case of a successful verification. After this, the user requests a parking space for which personal data (location) is required. The PSP checks the availability of parking spaces in that parking lot by sending a request to the Parking Lot Terminal (PLT). The PSP then handles the responsibility of making a parking reservation according to the user's request and sends it to the PLT who generates a parking permit according to the reservation. This permit is then relayed to the user via the PSP.

[66] M. Weber and Z. Ivana Podnar. "A Regulatory View on Smart City Services." *Sensors (Basel, Switzerland)* vol. 19,2 415. 21 Jan. 2019, doi:10.3390/s19020415, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6358906/ (accessed May 6, 2020).

[67] Business process model is created using the bpmn.io tool. The process model is used as the import to the DPO tool to evaluate its compliance.

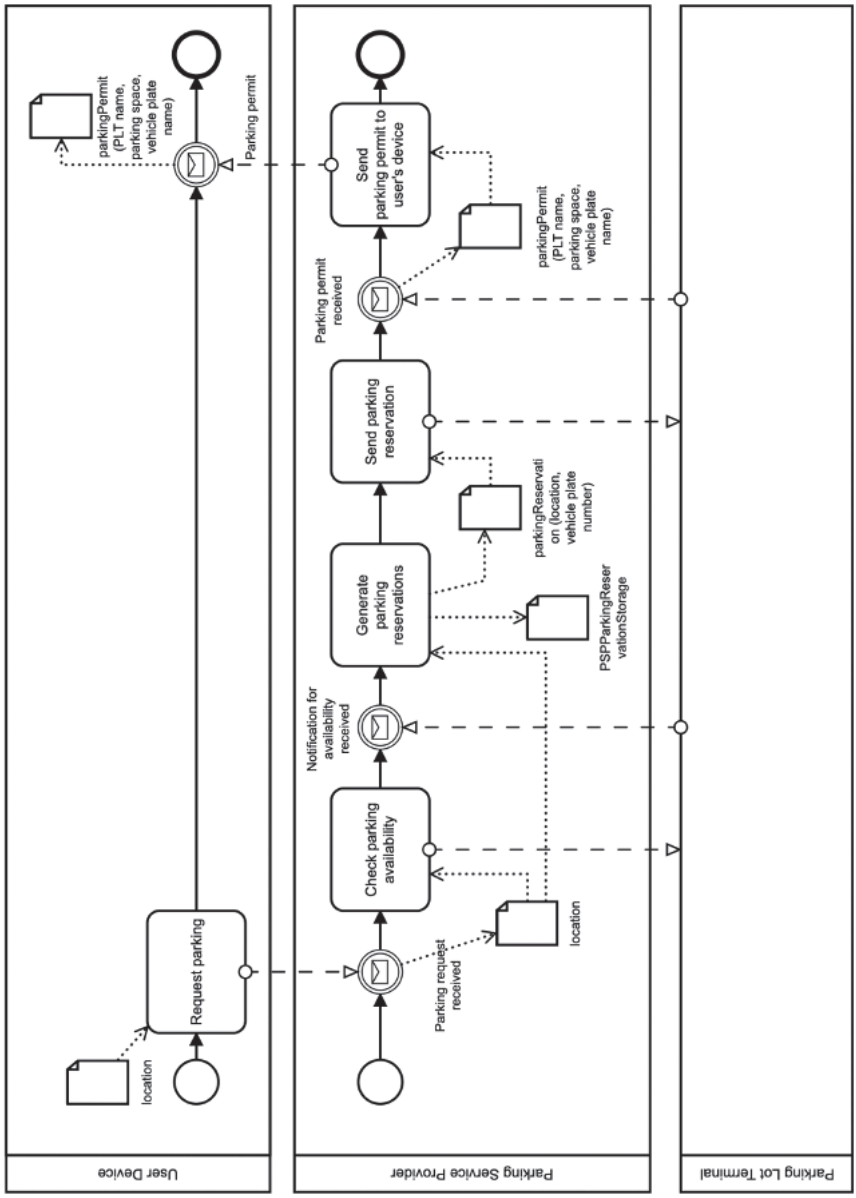

Figure 3 – Parking Ticket Generation Scenario (adapted from[68])

[68] N. A Onyinye, *A Comparison of Privacy Enhancing Technologies in Internet of Vehicles Systems*, Master thesis, University of Tartu, 2020.

3. Personal Data Protection and Intelligent Infrastructures: the added-value of Privacy Enhancing Technologies

Privacy Enhancing Technologies are technologies to support the implementation of the GDPR principles into an intelligent infrastructure system and in IoT tools. In this section, we briefly introduce current Privacy-Enhancing Technologies (PETs) categorized into 6 areas, i.e. digital signatures, authentication, communication systems, encryption technologies, computations, and general anonymization technologies. We map how each PET area and approach complies with GDPR processing principles.

3.1. Overview of current Privacy Enhancing Technologies to support GDPR principles

Privacy-Enhancing Digital Signatures. Privacy-preserving digital signatures mainly comply with GDRP processing principles explained above such as minimization, integrity, and partly with confidentiality. Privacy-enhancing digital signatures allow users to sign messages with integrity, authenticity, and non-repudiation properties such as common digital signatures, and also provide some additional privacy features. For instance, group and ring digital signatures provide a signer anonymity or a signer pseudonymity. Any user (a group member) can then anonymously sign a message on behalf of the group. Group and ring signature schemes are often used in group-based authentication scenarios in order to ensure data integrity and authenticity but also to keep the privacy of signers. The valid signatures should not be linkable to a concrete person who is using the private key. The signatures can be verified by anyone by using one public group key that does not point to a concrete signer. Ring signatures are similar to group signatures but are based on decentralized models without a group manager. Group and ring signatures can be suitable basic cryptographic tools for systems that require the anonymization of users such as e-voting, e-payments, e-coins, and other privacy-preserving use cases. Some group and ring signature schemes are already included in the standard ISO/IEC 20008-2:2013[69].

[69] International Organization for Standardization. ISO/IEC 20008-2: Information technology - security techniques - anonymous digital signatures - part 2: Mechanisms using a group public key. stage 60.60, 2013. More technical details about group signatures and ring signatures can be found in the works of Bellare et al. and Camenisch and Groth's; M. Bellare, D. Micciancio, and B. Warinschi. "Foundations of group signatures: Formal definitions, simplified requirements, and a construction based on general assumptions."

Other privacy-enhancing digital signatures are blind digital signature schemes that enable to hide (blind) the content of the message. Blind signatures are used in use cases where the message owner and signer are different entities. The signer is usually a third party that should not have access to data content. The signatures are then publicly verifiable against the unblinded message as a standard digital signature. Blind signatures are mostly used in payment systems such as PayCash. More about blind signatures can be found in papers[70] [71].

All above mentioned techniques provide strong privacy features and can increase the privacy in specific use cases.

Privacy-Enhancing Authentication. Privacy-enhancing authentication techniques such as Attribute-Based Credentials (ABC) or anonymous credentials and anonymous and pseudonymous entity authentication protocols mainly comply with the data minimization and partly purpose limitation GDRP processing principles. These authentication techniques usually enable users to join services without revealing their real identities and personal data. Some schemes also provide unlinkability and untraceability to prevent profiling users' behaviour in a service. These requirements and principles are also described in the standard ISO/IEC 29191:2012[72]. Nevertheless, anonymous authentication schemes should ensure that malicious users can be revoked from the system. For instance, ABC schemes are based on personal characteristics instead of user identity (i.e. full name, unique identifier, digital certificate X.509). The digital identity is considered to be a set of characteristics (personal attributes) that describe certain person, such as a driving licence, age, a group membership and more. These attributes can be shown selectively, anonymously and without anyone's ability to trace or link the showing transactions. For example, adults who want to access to liquor eshops only must prove that reach age limits by showing and proving these concrete attributes (i.e. older than 21). The attributes are issued usually by some third trusted party or service providers. ABC schemes and anonymous credentials are

International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg, 2003; J. CAMENISCH and J. GROTH. "Group signatures: Better efficiency and new theoretical aspects." *International Conference on Security in Communication Networks*. Springer, Berlin, Heidelberg, 2004.

[70] D. POINTCHEVAL and J. STERN, "Security arguments for digital signatures and blind signatures." *Journal of cryptology* 13.3, 2000, pp. 361-396.

[71] D. SCHRÖDER and D. UNRUH, "Security of blind signatures revisited." *Journal of Cryptology* 30.2, 2017, pp. 470-494.

[72] ISO/IEC 29191:2012 Information technology – Security techniques – Requirements for partially anonymous, partially unlinkable authentication.

usually based on asymmetric cryptographic primitives and comply with the user-centric approach[73].

Privacy-Enhancing Communication Systems. Privacy-enhancing communication protocols and systems are mainly focused on providing integrity and confidentiality when data are transferred via communication networks. The common security protocols such as IP security (IPsec), Transport Layer Security (TLS) or Secure Shell (SSH) offer authenticated encryption of data in client/server or peer-to-peer connections and prevent to eavesdrop the vital and personal data. Nonetheless, these protocols usually require to authenticate and identify data senders and receivers. Mix-networks, proxies and onion routing techniques enable users to create anonymous communication networks that protect against complex traffic analysis. Senders can communicate with destinations without revealing their identity or location. This contributes to the minimization GDPR principle. For example, mix-networks use mix nodes (proxy servers, relays) which gather messages from multiple transmitters to disrupt the relation between incoming and ongoing traffic. Onion routing[74] employs an onion encryption approach where a sender establishes a single encryption layer with each network node along the path, which is called an onion router. The messages are encapsulated by the sender in several layers of encryption, analogous to onion layers. Each onion router on the path decrypts its onion layer and relays data to the next onion router. When the final layer is decrypted, the data reach the destination (e.g. web server). The mixnets and onion routing mostly help individuals to communicate anonymously on the Internet.

Privacy-Enhancing Encryption Technologies. Encryption is considered as a security protection measure within the GDPR as well as a method for companies to comply with data protection requirements and avoid penalties in case of data breaches. However, the GDPR has no explicit encryption requirements[75].

---

[73] More about the privacy-preserving authentication techniques and concrete scheme examples can be found in the following works: J. CAMENISCH, and al. "Fast keyed-verification anonymous credentials on standard smart cards." *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, Cham, 2019; E. R. VERHEUL, "Practical backward unlinkable revocation in FIDO, German e-ID, Idemix and U-Prove." *IACR Cryptol*. ePrint Arch. 2016, 2016, p. 217.

[74] More about anonymous connections and onion routing can be found in the following works: M.G. REED, P. F.SYVERSON and D.M. GOLDSCHLAG. "Anonymous connections and onion routing." *IEEE Journal on Selected areas in Communications* 16, no. 4, 1998, pp. 482-494; D.M. GOLDSCHLAG, M.G. REED and P. F. Syverson, "Onion routing.", *Communications of the ACM* 42, no. 2, 1999, pp. 39-41.

[75] G. SPINDLER and P. SCHMECHEL, "Personal data and encryption in the European general data protection regulation", *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, *7*, 2016, p. 163.

Privacy-enhancing encryption technologies follow the principle of privacy by design and by default, as stored data remain encrypted, offering enhanced protection guarantees. Attribute based encryption, homomorphic encryption and searchable encryption are advanced cryptographic solutions that support enhanced functionalities. Each of these technologies can be used for outsourced data storage in encrypted form. Attribute based encryption offers advanced access control functionality, as the encrypted data can only be decrypted by an entity provided with a specific set of attributes[76]. Homomorphic encryption schemes enable performing computations over the encrypted data, while both the stored data and the computation results remain inaccessible to the hosting service[77]. Searchable encryption schemes enable search queries to be performed and statistics to be derived without the data being decrypted[78]. Additionally, it can be used to locate the data items that concern specific aspects and purposes. These data items can then be decrypted by authorized parties, allowing controlled access to only the relevant data items in the dataset. Searchable encryption and homomorphic encryption also facilitate user-centric approaches, as the outsourcing and management of the encrypted dataset can remain under the data owner's control.

While encrypted datasets are protected from unauthorized access by default, data minimisation still needs to be applied during the selection of data items to be included in the encrypted dataset, as parts of the data can be decrypted and re-identification could be possible if the decrypted data contains identifiable information. The fact that a dataset is stored in an encrypted form does not remove the need to avoid the collection and recording of any unnecessary identifying information. As far as security is concerned, encrypted data remains inaccessible to the storage service and, at the same time, is protected from information leakages in case of data breach. Data anonymity can also be maintained, as long as anonymous authentication techniques are used for accessing the storage service. The confidentiality of the encrypted data is protected, however additional protection techniques, such as digital signatures, need to be employed for the integrity and authenticity of the contents to be protected.

---

[76] V. Goyal, O. Pandey, A. Sahai and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data", *in Proceedings of the 13th ACM conference on Computer and communications security*, 2006, pp. 89-98.

[77] M. A. Will and R. K. Ko, "A guide to homomorphic encryption", *The Cloud Security Ecosystem: Technical Legal Business and Management Issues. Elsevier*, 2015, pp. 101-127.

[78] G. S. Poh, J. J. Chin, W. C. Yau, K. K. R. Choo and M. S. Mohamad, "Searchable symmetric encryption: designs and challenges", *ACM Computing Surveys (CSUR)*, *50*(3), 2017, pp. 1-37.

Privacy-Enhancing Computations. Privacy-enhancing computations can also be achieved using secure multi-party computations, which enable multiple parties to jointly compute a function on their inputs, while keeping those inputs private and without the need for a Trusted Third Party to be involved in the computation. In a secure multi-party computation protocol, no participant can learn anything further than their own entry, the public function being used for the computation and the result of the global computation. Therefore, secure multi-party computations mainly comply with the minimization GDPR principle.

Secure multi-party computation protocols are considered practical and applicable in contemporary problems, such as electronic voting and auctions, due to the dramatic increase in efficiency that has been achieved in the last decade[79]. A typical multi-party computation protocol achieves input privacy, i.e. no information about the private data of the participating parties can be inferred. Correctness of the output can also be ensured, depending on the security model of the protocol, either by guaranteed correct output, or by aborting the computation in case of an error[80].

General Anonymization Technologies. Both the traditional statistical disclosure control (SDC)[81] mechanisms and the recent differential privacy (DP)[82] methods can be regarded as measures, that can be taken by the data controller to fulfil the compliance requirements of data protection such as in the GDPR. Below, we first summarize the workflow of these mechanisms and methods, and then try to identify how they can help comply with the processing principles from the GDPR towards personal data protection.

For both SDC and DP, it is typically assumed that a data controller will possess a database which comprises the plaintext data from users. Then, the data controller can anonymize the data via a wide spectrum of techniques, e.g. suppressing and generalising attributes for SDC, calibrating noises into the querying results. One functional difference between SDC and DP is that SDC enables a data controller to publish an anonymized database while DP focuses more on the postprocessing of querying results from the plaintext database (i.e. no database will be published).

---

[79] D. Evans, V. Kolesnikov and M.Rosulek, "A pragmatic introduction to secure multi-party computation", *Foundations and Trends® in Privacy and Security*, *2* (2-3), 2017.

[80] D. W. Archer, D. Bogdanov, Y. Lindell, L. Kamm, K. Nielsen, J.I. Pagter, N. P. Smart and R. N. Wright, "From keys to databases–real-world applications of secure multi-party computation", *The Computer Journal*, *61*(12), 2018, pp. 1749-1771.

[81] A. Hundepool et al, *Statistical Disclosure Control*, John Wiley Sons, Inc., 2012.

[82] C.Dwork, F. McSherry, K.Nissim, and A.Smith, "Calibrating noise to sensitivity in private data analysis", *in Theory of Cryptog-raphy Conference*, Springer, 2006.

SDC mechanisms can contribute to complying with the purpose limitation principle, by processing the concerned attributes accordingly. For instance, if an attribute is not related to the predefined purpose of data processing, then these attributes can be suppressed before publishing the database. Similarly, DP can contribute in two ways. On the one hand, it can be configured to only answer queries that match with the predefined purpose. On the other hand, it can provide a fine-grained implementation of purpose limitation, e.g. by controlling the amount of calibrated noises with respect to the specific purpose of data processing. SDC mechanisms can contribute to the minimization of data exposure to some extent, by hiding the sensitive attributes in users' records. However, SDC may suffer from linkage attacks or the alike if some background knowledge is available to the attacker. In contrast, DP methods can rigorously minimize the data exposure by adjusting the calibrated noise, yet without suffering from the vulnerabilities of the SDC mechanisms. Adopting SDC and/or DP enables a data controller to act in an accountable manner when dealing with sensitive personal data. Particularly, with DP, a data controller can get high assurance regardless of the knowledge of the attacker.

SDC and DP do not directly address the storage limitation, transparency, integrity, and confidentiality principles. Particularly, they do not offer standard confidentiality guarantee on the data (e.g. those from Encryption techniques). However, they offer certain levels of privacy protections for the relevant human users. The accuracy principle does not closely connect with SDC and DP. However, we should note that if a data controller has published an anonymized database based on SDC, then it may need to renew the published database in order to fulfil this principle. One point worthy of note is that publishing the (almost) same database twice may lead to privacy risks. The data controller should be cautious if this occurs.

For both SDC and DP, we assume the data controller has access to the plaintext database. However, in practice, the users might be reluctant to share data directly with the data controller. In addition, the data controller might reduce its liability to protect the users' data. As such, distributed variants of SDC and DP might be employed. If this is done, then it provides a higher-level protection to the users.

3.2. Summary

The table below summarizes the privacy enhancing technologies explained above and its compliance with GDPR processing principles.

PETs area	PET methods/technology	Supported GDPR principles
Privacy-enhancing digital signatures	Group signatures Ring signatures Blind signatures	*Data minimization* *Integrity and partly with confidentiality*
Privacy-enhancing authentication	Attribute-based credentials Anonymous credentials Pseudonymous entity authentication	*Data minimization* *Partly purpose limitation*
Privacy-enhancing communication systems	IP security (IPsec) Transport Layer Security (TLS) Secure Shell (SSH) Mix-networks Onion Routing	*Integrity and confidentiality* *Partly data minimization*
Privacy-enhancing encryption technologies	Attribute-based encryption Homomorphic encryption Searchable encryption	*Confidentiality and partly integrity* *Data minimization and anonymization*
Privacy-enhancing computations	Secure multi-party computation	*Data minimization*
General anonymization technologies	Statistical Disclosure Control (SDC) Differential Privacy (DP)	*Purpose limitation* *Data minimization*

Table 1. Summary of PETs and its compliancy with GDPR processing principles

3.3. Illustration of Personal Data Managing in Parking Reservation Generation Scenario

In this section, we discuss how regulation compliance of business processes could be achieved. The application of the GDPR method consists of several steps[83]: first, the GDPR model is instantiated to produce an as-is compliance model. Second, the as-is model is compared to the GDPR compliance model to determine the missing essential classes and/or attributes. Third, these missing elements are used to define potential compliance violations that will identify areas of investigation for the DPO. If necessary, the process model can be corrected to address them as well.

[83] R. MATULEVIČIUS, J. TOM, K. KALA and E. SING, "A Method for Managing GDPR Compliance in Business Processes", *in Advanced Information Systems Engineering*, N. HERBAUT and M. LA ROSA (eds), Lecture Notes in Business Information Processing, vol 386. Springer, 2020, Cham. https://doi.org/10.1007/978-3-030-58135-0_9.

Both the GDPR model and the supporting method are implemented to the prototype tool[84] which could be used by the data protection officers of controllers to analyze and evaluate the complex intelligent infrastructure processes, where processes are characterized as a technical system.

Instantiation of the GDPR Model. To illustrate the instantiation of the GDPR model (see Figure 2) and generate the as-is compliance model, we consider the data object parkingReservation that holds the vehicle's location information. Based on the described parking scenario, the assignment of attributes and values are described below.

GDPR roles: In the scope of this process, the User Device as provider of the personal location information is considered with respect to the data subject. The PSP is the controller as it is the service provider who utilizes the location information over multiple processing operations. The PLT who receives the parking reservation (holding location information) to produce the permit is considered a recipient. There is no processor or third party in this process, so those classes are assigned a NotRequired stereotype.

Consent and purpose of processing: Personal data being considered here is the parkingReservation data object. As it does not fall under any of the special categories of personal data described by the enumeration, DATA CATEGORY, it is assigned the value general. Before checking whether the class Consent can be instantiated (meaning, whether the process shows the collection of data subject consent evidenced by a consent agreement), it is necessary to determine whether it needs to be instantiated (collected) at all[85]. Since the personal data in question does not fall under a special category, the SpecialPurpose class does not need to be instantiated as it describes the consent collection rules for special categories of personal data. However, the class Purpose must be instantiated and all its attributes are set to false except general in this case because the data is not being collected for any of the other purposes as described by Article 6 that exempt the controller from the collection of consent. This means that the collection of consent is necessary in this business process. Since there is no evidence of consent collection in the business process diagram, the class Consent and its corresponding artifact, ConsentAgreement cannot be instantiated. They are assigned the stereotypes MissingClass and MissingArtifact respectively.

Processing/filing system, processing task and technical measures: Since the parking reservation is an input to the SendParkingReservation activity, the class ProcessingTask is instantiated as the same. However, since there is no

[84] DPO: https://dpotool.cs.ut.ee/.

[85] For example, one should consider whether the contractual obligations could be fulfilled in the first place. But for the sake of the illustration and discussion, we exemplify how consent could be managed in this chosen case.

evidence of any processing task being recorded in the model, the artifact, ProcessingLog, cannot be instantiated and it is assigned the stereotype MissingArtifact. The PSP system is just used for verification and information transmission and does not store any of the information it processes, so it cannot be classified as a filing system. So the class FilingSystem is assigned a NotRequired stereotype. The remaining classes, ProcessingSystem and TechnicalMeasures, must be determined after consultation with the technical stakeholders of the PSP, as these characteristics cannot currently be captured in BPMN diagrams. However, their attributes are to be used by the DPO to characterize the essential privacy and security properties of the processing/filing system in operation specified by the GDPR. We assign the processing of false to indicate to the DPO that they must be investigated and clarified.

Defining Non-compliances and Related Regulation Principles. Now that we have an instantiated compliance model, we are ready to proceed with the business process evaluation. We compare the as-is model with the compliance model to determine whether any required attributes, classes and artifacts are missing. It is important to note that some missing classes may not be required at all (e.g., Processor, Third Party or Recipient) or in the particular process context (class SpecialPurpose in this case) but generally, both artifacts representing the consent agreement and the record of processing are required in a business process. In this scenario, we see that once the GDPR roles and personal data classes are instantiated, the SpecialPurpose class is not required and due to the nature of the Purpose class as general, consent is required but is not evident. Therefore, the class Consent cannot be instantiated and its corresponding artifact, ConsentAgreement is missing. Similarly, the artifact ProcessingLog associated with the processing task in class SendParkingReservation is a missing required artefact as there is no evidence of it in the process model. The ProcessingSystem and TechnicalMeasures classes have missing required attributes. Table 2 summarizes the identified non-compliances (NC) and the corresponding GDPR principles. In the next step the DPO must check each of them to ascertain whether these are modelling issues or actually existing operational issues.

ID	Description	Related GDPR Principle
NC#1	Record of processing is missing	Transparency
NC#2	Data subject consent is missing	Purpose limitation
NC#3	Processing system has missing attributes, no technical measures	Minimization, Anonymization, Pseudonymization, Integrity, Confidentiality,
NC#4	Missing Privacy policy	Transparency, Purpose limitation

Table 2. Non-compliances found in the scenario model and their correspondence to the regulation principles

Refining Business Process to Address Non-compliance and fulfill the GDPR principles. Now we look at how the business process model can be refined to address the non-compliances.

Record of processing (Art. 30): To indicate that a processing task is recorded appropriately, the process model must contain an activity that records the processing of the utilized personal data. The output of the recording can be considered the ProcessingLog artifact and can be represented by a data object. Alternatively, a data storage element can be used to show that the record is structured along with other records of processing activities. One way to model this is shown in Figure 4 by introducing activity Record processing of location information. With the inclusion of this process fragment, NC#1 is addressed.

Data subject consent (Art. 7): Data subject consent and its associated consent agreement can be represented in a process model by the pattern shown in Figure 4 (see activities Provide consent and Request consent to process personal data inserted at the start of the process). Here, there is a communication exchange where the PSP requests consent to process the user's personal data. The output of the pattern is the ConsentAgreement artefact. With the inclusion of this process fragment, NC#2 is addressed.

Security of processing (Art. 32): The class ProcessingSystem is composed of attributes that any processing or filing system must possess to be compliant with the GDPR. As depicting these attributes is out of the scope of the BPMN language, they are better represented as annotations that the DPO must investigate to confirm the essential properties of the system as seen in Figure 4. With the inclusion of this annotation, NC#3 is addressed from the perspective of the process model.

Technical measures (Art. 25): Figure 4 shows how the lack of technical measures present in the process model can be addressed using Privacy-Enhanced BPMN[86] with the addition of privacy enhancing stereotypes to activities. In this case, processing of the location data is secured by the addition of two privacy preserving technologies, Public Key Encryption and Computation that utilizes a public key on the user's device to encrypt the location data that the PSP then computes on, and an additional security layer on the communication channel itself via TLS to represent a secure communication channel. This is an example of how NC#4 can be addressed in the process model itself.

[86] P. Pullonen, J. Tom, R. Matulevičius and A. Toots, "Privacy-enhanced BPMN: Enabling data privacy analysis in business processes models", *Software and Systems Modeling*, 18(6), 2019, pp. 3235-3264.

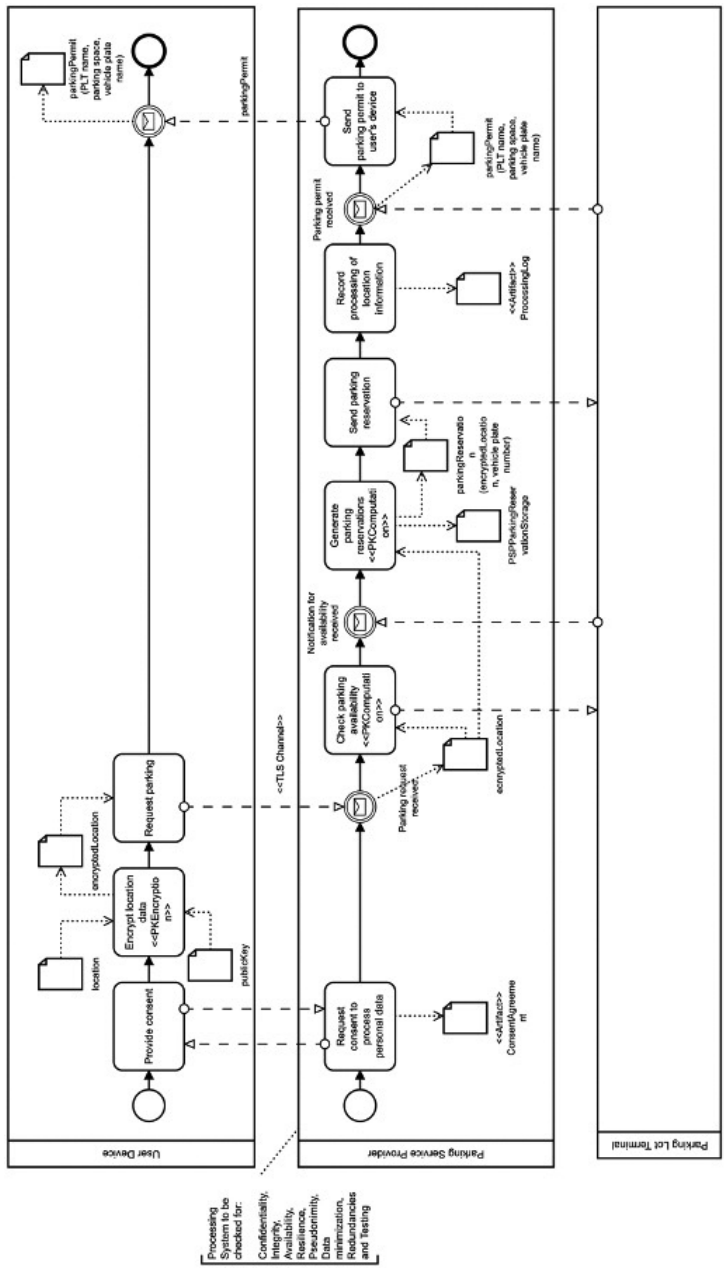

Figure 4 – Non-compliance resolution

3.4. Illustration for privacy policies: a language-based approach for Editing, Analysis, and Enforcement of privacy requirements

The collection, usage, and sharing of users' data is usually regulated by privacy policies, written in natural language terms, in which specific actions are authorized, obliged, or denied, under some contextual conditions. This is the case, for example, of the rules that each provider of social networking services publishes on its data processing pages[87]

Nevertheless, as we read in the previous sections, the GDPR intends to regulate any kind of electronic transaction that has to do with personal data. A type of scenario that lends itself to the dictates of GDPR – the Parking Ticket Generation – has been schematized through business processes in the previous section. Examples of rules that deal with, e.g., the record of processing (Art. 30) and the data subject consent (Art. 7) can be rendered, in natural language, by the following expressions:

Record of Processing: "The processing of the parking reservation must be recorded by the data controller".

Data Subject Consent: "If the data subject has not given consent to the parking reservation storage, then the data recipient cannot issue the parking permit".

Although the use of natural language (NL) enables end users to read and understand the allowed (or obliged, or denied) operations on their data, a key issue lies in the fact that NLs are not machine readable, and automatic controls on how the data are actually going to be used and processed by the entities that operate on them is not feasible. In particular, NLs cannot be used as the input language for a policy-based software infrastructure to be used for policy management. In fact, both automated policy analysis (the process to assure the lack of conflicting data policies[88] and policy enforcement (the actual application of the data policies, whenever a data access request takes place) require inputs in a machine-readable form, e.g. the *de facto* standard XACML.

With the aim of moving in the direction of managing and enforcing privacy policies automatically, here we show a language-based approach that leverages machine-oriented, English-based Controlled Natural Languages (CNLs).

[87] See, for example, the Facebook data policies pages https://www.facebook.com/full_data_use_policy or the Twitter privacy policies pages https://twitter.com/en/privacy.

[88] G. Costantino, F. Martinelli, Il. Matteucci and M. Petrocchi, "Efficient Detection of Conflicts in Data Sharing Agreements", *ICISSP* (Revised Selected Papers), 2017, pp. 148-172.

CNLs are a subset of natural languages, specifically conceived to make machine processing simpler. A CNL is, in essence, a developed language that is based on natural language, but it is more restrictive in terms of lexicon, syntax, semantics, while at the same time retaining most of its natural properties[89]. CNLs have a more contrived representation, in terms of grammar and vocabulary, and they thus reduce the ambiguity and complexity of a complete language[90], e.g., English, Spanish, French, Swedish, Mandarin, etc.[91].

CNLs have been proved to be effective in mitigating linguistic ambiguity challenges, as they can easily be translated into a formal language such as first-order logic or different version of description logic, automatically and mostly deterministically. Noticeably, a branch of CNLs conceived for expressing data privacy regulations are formal *per se*, being born with an associated formal syntax and semantics[92]. In general, these languages can conveniently express the kind of information that occurs for example in software specifications, formal ontologies, business rules, legal and medical regulations.

The proposed CNL has the purpose to reduce the barrier of adoption of legal contracts regulating data sharing, usually written in natural language, in terms of privacy guarantees, as well as to ensure contract mapping to formal languages that allow its automatic verification of the agreement and enforceable languages to permit its enforcement. A data sharing contract – or data sharing agreement – can be seen essentially as a policy to be followed between two or more parties, which agree on some terms and conditions with respect to data sharing and usage.

CNL4DSA. The CNL4DSA (Controlled Natural Language for Data Sharing Agreement) language supports the enforcement of privacy and security of electronic data exchange. CNL4DSA allows simple, yet formal, specifications of different classes of privacy policies, as listed below:

- **authorizations**, expressing the permission for subjects to perform actions on objects (e.g., on user's data), under specific contextual conditions;

[89] T. Kuhn, "A survey and classification of controlled natural languages," *Computational Linguistics*, vol. 40, no. 1, 2014, pp. 121-170.

[90] R. Schwitter, "Controlled natural languages for knowledge representation," *in Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010, pp. 1113-1121.

[91] A. Wyner, K. Angelov, G. Barzdins, D. Damljanovic, B. Davis, N. Fuchs, S. Hoefler, K. Jones, K. Kaljurand, T. Kuhn et al., "On controlled natural languages: Properties and prospects," *in International Workshop on Controlled Natural Language*. Springer, 2009, pp. 281-289.

[92] Il. Matteucci, M. Petrocchi and M. L. Sbodio, "CNL4DSA: a controlled natural language for data sharing agreements". *SAC*, 2010, pp. 616-620.

- **prohibitions**, referring to prohibit the fact that a subject performs actions on an object, under specific contextual conditions;
- **obligations**, defining that subjects are obliged to perform actions on objects, under specific contextual conditions.

Central to CNL4DSA is the capability to formally specify that "subject s performs action a on object o".

By adding the can/must/cannot constructs to the core constructs of the language, it is possible to express authorizations, obligations, and prohibitions. The authorizations, obligations, and prohibitions can then be evaluated according to properties of subjects and objects, for example in terms of users' roles, data categories, time, and geographical location.

To give the idea, the two rules regarding the parking reservation scenario, expressed above in natural language, are rendered in CNL4DSA as follows:

Record of Processing: "IF subject1 hasRole ParkingServiceProvider AND object1 hasCategory ParkingReservation THEN subject1 MUST record object1"

Data Subject Consent: "IF subject1 hasRole UserDevice AND subject2 hasRole ParkingServiceProvider AND subject3 hasRole ParkingLotTerminal AND object1 hasCategory ParkingPermit and object1 isRelatedTo subject1AND subject1giveConsent NoConsent THEN subject3 CANNOT issue object1"

CNL4DSA-based toolkit. Although born as a language to describe data sharing policies, CNL4DSA has proved suitable for expressing other kinds of requirements, such as software product lines specifications[93]. The language is not domain-specific, since it does not have a fixed associated vocabulary. Hence, it can be applied to various use cases, such as social networking[94] e-health[95], and emergency management[96] scenarios.

The strength of this language is that, over the years, a series of tools have been developed around it, each of which serves a precise purpose within the life cycle of a rule. Below, we describe each of these tools, and the role covered by CNL4DSA.

[93] St. Gnesi and M. Petrocchi; "Towards an executable algebra for product lines", *SPLC* (2), 2012, pp. 66-73.

[94] Ir. K. Tanoli, M. Petrocchi and R. De Nicola, "Towards automatic translation of social network policies into controlled natural language", *RCIS*, 2018, pp. 1-12.

[95] Il. Matteucci, P. Mori, M. Petrocchi and L. Wiegand, "Controlled data sharing in E-health", *STAST*, 2011, pp. 17-23.

[96] F. Martinelli, Il. Matteucci, M. Petrocchi and L. Wiegand, "A Formal Support for Collaborative Data Sharing", *CD-ARES*, 2012, pp. 547-561.

A textual rule, either written in CNL4DSA or in natural language, is managed by a CNL4DSA-based toolkit, originally proposed in[97] and successively renewed. Initially comprising a CNL4DSA Authoring Tool, a CNL4DSA Policy Analyser, and a CNL4DSA Mapper Tool, the toolkit has recently been enriched with a translator from natural language rules to CNL4DSA rules, the NL2CNL translation tool. We provide a brief description of the components hereunder:

- *NL2CNL Translator*: a user with no expertise of CNLs can edit rules in natural language (e.g., in English); with a minimal user's effort, the translator outputs the rules in CNL;
- *CNL4DSA Authoring Tool*: an author with expertise in CNLs can edit rules directly in CNL4DSA. The rules are constrained by CNL4DSA constructs (see Section 3.7.1) and the terms in the rules come from specific vocabularies;
- *CNL4DSA Analyser*: it analyzes a set of CNL4DSA rules, detecting potential conflicts among them (a conflict exists when two (or more) rules simultaneously allow and deny an access request to data under the same contextual conditions). In case a conflict is detected, a conflict solver strategy based on prioritization of rules is put in place to correctly enforce the priority rules;
- *CNL4DSA Mapper*: it translates the CNL4DSA rules into an enforceable language. The mapping process takes as input the analyzed CNL4DSA rules, translates them in a XACML-like language[98] and combines all the rules in line with the predefined conflict solver strategies. The outcome of this tool is an enforceable policy. Such policy will be evaluated at each request to access, use, and operate on the data specified in the policy itself.

A CNL4DSA Lifecycle Manager orchestrates all the previous components. When users log into the Lifecycle Manager, this enacts their specific functions, according to the user's role (e.g., end-data owner, data controller, data processor). Thus, users interact with the toolkit via the Lifecycle Manager.

Overall, CNL4DSA richness and flexibility, both in terms of describing specifications from different domains and of being equipped with different specifications processing tools, go into the direction to achieve an

[97] J. Fr. Ruiz, M. Petrocchi, Il. Matteucci, G. Costantino, C. Gambardella, M. Manea and A. Ozdeniz, "A Lifecycle for Data Sharing Agreements: How it Works Out", *APF*, 2016, pp. 3-20.

[98] OASIS XACML Technical Committee, "eXtensible Access Control Markup Language (XACML) Version 3.0," 2013.

integrated framework for the specification and analysis of safety, security, privacy, and trust in complex and dynamic scenarios.

Conclusion

In this paper, we focus mainly on one of these characteristics, which is the lawful processing of personal data into an Intelligent Infrastructure environment. Several obligations for the data controller frame the management of personal information. We focused on the requirements of purpose limitation, minimization, storage limitation, integrity and confidentiality of the data and, finally, transparency for the data subjects.

Information and communication technologies are sometimes represented as *per se* intrusive into the privacy. However, the GDPR gives to the ICT technologies an important role to reinforce the principles surrounding the protection of personal data. Indeed, the GDPR enshrines the privacy-by-design requirement. According to it, the data controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures. The Privacy Enhancing Technologies are then of the utmost importance as their function is to integrate and strengthen data protection principles into various ICT use cases.

Privacy-enhancing digital signatures and authentication methods are already well-established techniques that can preserve various privacy and security features for users in various ICT use cases that must address various GDPR privacy principles.

For their parts, privacy-enhancing encryption technologies and privacy-enhancing computations constitute valuable building blocks for privacy-preserving applications that follow the principle of privacy by design and by default.

Although the aforementioned PETs are already well-established techniques that can preserve various privacy and security features for users. Nevertheless, appropriate combination with other techniques such as privacy-enhancing communications and privacy-preserving data publishing, such as differential privacy, can increase GDPR compliace.

In addition to a technical reinforcement of the legal obligations, let's hope that these privacy enhancing technologies will be accompanied by transparency information towards users to strengthen the trust of the data subjects in intelligent infrastructures.