Intelligent Environments 2020
C.A. Iglesias et al. (Eds.)

doi:10.3233/AISE200030

The Unsecure Side of (Meta)Data in IoT Systems

Pier Giorgio CHIARA [0000-0002-9444-3480] [a,1]
[a] *University of Luxembourg, University of Bologna, University of Turin*

Abstract: The exponential spreading and deployment of emerging digital technologies such as the Internet of Things (IoT) has been remarkable: the IoT market is expected to triple, at least, from USD 170.57 billion in 2017 to USD 561.04 billion by 2022. IoT technologies collect, generate and communicate a huge amount of different data and metadata, through an increasing number of interconnected devices and sensors. Current EU legislation on data protection classifies data into personal and non-personal. The paper aims at charting the resulting entanglements from an interdisciplinary perspective. The legal analysis, integrated with a technical perspective, will address firstly the content of IoT communications, i.e. "data", and the underlying distinction between personal and non-personal. Secondly, the focus will shift on the metadata related to communications. Through a technical analysis of the highly sensitive nature of metadata, even when the content is encrypted, I will argue that metadata are likely to undermine even more the ontological and sharp division between personal and non-personal data upon which the European legal frameworks for privacy and data protection have been built. The incoming ePrivacy Regulation shall provide metadata, which should be considered always personal data, the same level of protection of "content" data. This interpretation might broaden the scope of application of GDPR and the connected obligations and responsibilities of data controllers and data processors too much.

KEYWORDS: Data Protection, Privacy, Security, IoT, Non-Personal Data, Metadata

I. Introduction: The Internet of things as a major source of threat for security, privacy and data protection.

The need of a theoretical premise is justified by the complex interrelationship between the concepts of privacy and data protection. Albeit they may often overlap, privacy and data protection are rights with different rationales [1–3], as outlined by the jurisprudence of the European Court of Justice (ECJ) and the European Court of Human Rights (ECHR). The classic formulation of the right to privacy follows a general principle of non-interference in one's private sphere. Pagallo, recalling Hanna Arendt, conceives privacy, in the digital era, as a movable degree of "opaqueness" [4]. The level of opacity that an individual may expect to attain is defined by the rules of the legal system[2]: public

[1] Corresponding Author, Law, Science and Technology, Rights of Internet of Everything – Horizon 2020 Marie Sklodowska-Curie ITN EJD; E-mail: piergiorgio.chiara@uni.lu.

[2] Cfr. *ex multis* with Floridi [44]: "the "ontological friction" consists in the amount of work and efforts required for a certain kind of agent to obtain, filter and/or block information (also, but not only) about other agents in a given environment".

interest criteria, defined by law, may ultimately interfere with the right to privacy. Data protection, on the other side, has been conceived as a more proactive right[3], underpinned by the principles of transparency, lawfulness and fairness, according to which personal data are collected, processed and therefore used[4].

For the purpose of this paper, it will be assumed that, in order to address privacy and data protection concerns raised by the IoT, privacy enablers embedded from the design stage of these devices are necessary in order to make these devices trustworthy, safe and reliable [5,6]. The need is thus to promote better privacy engineering practices. There is a large consensus in considering security[5] and data protection by design closely interlinked. Data protection norms seek to address the requirements regarding the protection of personal data, which are not only security-related; conversely, security settings refer to all types of products, systems and services [7][6].

IoT ecosystem is peculiarly designed and constrained when it comes to map cyber-security vulnerabilities and privacy and data protection threats. Firstly, the problem can be framed through an interconnected vertical and horizontal model: IoT has many possible applications (also called verticals) that range within a wide spectrum from eHealth and smart home scenarios to industrial and smart cities ecosystems. Horizontally, one can imagine the combinations of multiple technologies that enable IoT functionalities: radio-frequency identification (RFID), wireless sensor networks, cloud computing, etc.

Accordingly, security risks arise: each vertical has specific requirements, often different from the others and every part of the chain of all of those technologies needs to be secured, which is not a trivial task [8]. There is an urgent necessity of scalable decentralized defensive frameworks.

From a data protection standpoint, another crucial layer of complexity is represented by the huge amount of data generated or collected by the numerous sensors equipped on smart objects [9]. The scale of Big Data processing takes to a level where risks are unforeseen: “the scale is in terms of volume, variety, velocity and veracity, all the V’s of the big data definition, and their combination in analytics technologies” [10]. Users’ profiles can be easily inferred by the data collections of these data [11–13]. Moreover, these huge databases can be correlated thanks to multiple data mining techniques, resulting in “a continuous stream of profiles that can be tested and enhanced to better service those that ‘use’ them” [12]. Among these techniques, data aggregation is particularly relevant for the scope of this paper[7]. Thus, data from various IoT sources need to be grouped together from similar or diverse sources for further processes based on a well-defined data model (e.g., physical locations, device types, etc.)[8]. As stressed by the report of Article 29 Data Protection Working Group, “once the data is remotely

[3] Conclusion of the General Advocate Sharpston, Opinion delivered on 17.06.2010 about the joined cases C-92/09 and C-93/09.

[4] See Article 8(2) of the EU Charter of fundamental rights.

[5] The broad notion of security may be misleading since a proper distinction needs to be done between information security and cyber-security. The former underpins the protection of data and information, while the latter generally refers to the ability to protect or defend the use of cyber-space form cyber-attacks, encompassing therefore a broad range of risks governance.

[6] Enisa has developed a methodology which assists the development of applications in a secure manner, in order to decrease the number and severity of IoT vulnerabilities [45].

[7] Cfr. with Rajagopalan and Varsheney [46] for a specific case-study on wireless sensor networks.

[8] Many authors have begun to reason in terms of “group privacy” and “collective data protection”, since this kind of data collection and processing treat groups rather than individuals; cfr. with [47,48].

stored, it may be shared with other parties, sometimes without the individual concerned being aware of it. In these cases, the further transmission of his/her data is thus imposed on the user who cannot prevent it without disabling most of the functionalities of the device. As a result of this chain of actions, the IoT can put device manufacturers and their commercial partners in a position to build or have access to very detailed user profile" [14]. Leaving aside further reflections on consent which fall outside the scope of the present investigation, the focus lies in the risks underpinned by users' profiling [15]. Profiling is not illegal *per se*: recital 48 of the GDPR thus states that data subjects should be informed about the existence of profiling (that is, the construction of profiles), and the consequences of such profiling (that is, the consequences of applying such profiles). On a preliminary basis, it becomes vital to ascertain the nature of data that are going to be used to create users' profiles. A crucial step of a cautious risk-analysis in IoT sector should outline whether the data processed are personal or non-personal, in order to identify the applicable regulatory framework.

Finally, from a privacy preserving perspective, IoT systems pose severe concerns when it comes to address the confidentiality of communications[9]. It has been acknowledged by the proposal for the adoption of the so-called ePrivacy Regulation, by pointing out at recital 15 that the full protection of the rights to privacy and confidentiality of communications should apply to the transmission of machine-to-machine communications, in order to promote a trusted and secure Internet of Things.

II. The interrelation between non-personal data and metadata in the context of IoT

This section will address firstly the content of IoT communications, i.e. "data", and the underlying distinction in personal and non-personal data. Secondly, the focus will shift on the metadata accompanying the communications.

The notion of non-personal data is not clearly defined. Thus, even the European Regulation for the free flow of non-personal data provides a negative definition, i.e. "data other than personal data as referred to in Article 4(1) of Regulation (EU) 2016/679"[10]. Additionally, not only the definition of "personal data" is intentionally designed as broadly by the European legislator[11], but also the identifiability criterion is dynamic and context-dependent, which, similar to "relate to"[12], cannot be established in an absolute way[13]. It results that legal certainty is even more difficult to obtain.

The non-personal nature of such data might encompass a wide spectrum of information: machine-generated data and commercial data, whether they have never been personal, i.e. data not relating to an identified or identifiable person, or subsequently anonymized [16].

At the dawn age of IoT's hyper-connection, data reuse and data mining, it will become more and more arduous to discern whether a piece of information will not impact the

[9] See recital 15 of the Directive 2002/58/EC: "A communication may include any naming, numbering or addressing information provided by the sender of a communication or the user of a connection to carry out the communication".

[10] See article 3(1), EU Regulation 2018/1807 for the free flow of non-personal data.

[11] The ECJ has endorsed this broad understanding of the concept of personal data: it is not necessary that all the information allowing the identification of the individual must be in possession of one person (see Judgment of 19 October 2016, Patrick Breyer v Bundesrepublik Deutschland, C-582/14, paragraph 43).

[12] See article 4(1), EU Regulation 2016/679

[13] See Recital 26 of EU Regulation 2016/679

privacy and the right to data protection of an individual [17]. Furthermore, the possibility to single out an individual on the basis of very few data points has become easier [18]. This happens mainly due to Big Data analytics techniques [19], that rely upon huge datasets composed by personal or non-personal data or, as most often happens, mixed datasets combining the two[14].

Threats to privacy and data protection may as well arise from other data processing than personal, by combining various allegedly non-personal data to infer information related to a person or a group [20,21].

Current practices of market operators show the tendency to implement standardized procedures for the anonymisation of personal data [22], therefore making them non-personal[15]. There is a thriving literature[16] nonetheless aiming at implement algorithms in order to demonstrate how insecure most of the anonymization techniques could be. Organizing, crossing and correlating several datasets with few anonymized personal records may result in the emergence of patterns related to single individuals out [23,24]. In this regard, it is appropriate to consider the case of anonymised statistical data. The EDPS analysed the processing of statistical data, which usually consists of two different phases: "the initial phase while re-linking the data is still possible, and indeed, desired in order to enrich statistical data by linking various datasets[17]; a later phase when statistical data has been prepared, and the keys allowing linking the various datasets can be destroyed" [25]. However, the destruction of the keys does not necessarily convert such statistical data into non-personal data. Additionally, the EDPS clarifies that the input data may be processed in a twofold way: destroyed together with the identification-keys or stored as raw data. As noticed by Graef et al. [26] "even when organisations are only in possession of aggregate statistical data and the initial identifiers have been destroyed, such data might still be considered personal depending upon whether it has been subjected to anonymisation techniques, and whether the latter are considered adequate"[18]. The researches of Sweeney [27] and Narayanan and Shmatikov [28] are two ground-breaking analysis which might be better illustrate how data aggregation works in practice. Among the many merits, the above mentioned examples show how and to what extent data inference and re-identification could hamper the individual's rights to privacy and data protection, "for example leading to humiliation or even threat to life due to the disclosure of confidential information" [29].

Moving forward on the ambiguous and blurred distinction when it comes to categorize data, now the investigation considers the other face of the same coin, namely metadata. The proposal for the ePrivacy Regulation defines metadata at article 3(c) in a very detailed fashion, acknowledging their sensitive nature[19]: they may expose sensitive information and present significant risks [30].

[14] The mixed dataset, for practical purposes, has been referred "to a situation whereby a dataset contains personal data as well as non-personal data and separating the two would either be impossible or considered by the controller to be economically inefficient or not technically feasible" [49].

[15] See Recital 26 of EU Regulation 2016/679

[16] Cfr. *ex multis* with Ohm [50].

[17] The input data are pseudonymised and, through other technical and organisational measures (complaint with article 32 of EU Regulation 2016/679), the risks of re-identification of the individuals are minimised. The pseudonymisation process requires key coding the data: the keys, i.e. the information that links the datasets to the correct individuals, must be kept separately.

[18] For an in-depth analysis, see Article 29 Working Party on anonymisation techniques [51].

[19] See Recital 2 of the Proposal for the adoption of the ePrivacy Regulation

Recent research shows how metadata may be as sensitive as content, outlining individual's information exposure in IoT environments[20], in terms of destinations of network traffic. Even if encrypted. The data packets[21] communicated (transport layer) from the IoT system to the server are sent with so-called timestamps, i.e. the time and date of the action delivered by the device: in case of aggregated timestamps, one can retrieve timing patterns. Timestamps are metadata: they set the context in which the informational object is formed and shared. While data present elements of ambiguity and subjective interpretation, context metadata (e.g. timestamps) are more meaningful and objective: it is the human experience of the use of informational or physical objects, and their interconnection thanks to IoT, the most valuable resource and also the one less governable by the end-user [31].

In Ren et al. [32], to understand the information exposed during the interaction with an IoT device, the research started with collecting network trafficking, and labelling it afterwards, when devices are powered on. It resulted that diverse non-first-party destinations, e.g. Amazon, Google, Akamai, have received information from the majority of the IoT devices under investigation. Researchers then focused on whether devices send data securely by analysing the means of encryption adopted. Even though unencrypted traffic is a minority of all traffic, substantial information exposure has been identified via plaintext traffic for all devices, categories, interactions, and regions. The personal data -or even sensitive data- exposed in plaintext was limited. Moreover, the authors identified many cases of unexpected behavior, e.g. when a device generates network traffic corresponding to an interaction that either did not occur, or it was not intended by the user. There were cases of devices surprisingly sending audio or video in the uncontrolled experiments, highlighting that concerns about individuals' data exposition by IoT devices is warranted.

The three aforementioned researches are particularly relevant for this paper since they show that even when devices use encryption techniques, the timing patterns of the network traffic enable accurate identification of the interactions that caused the network traffic. "Put another way, an eavesdropper can reliably learn a user's interactions with a device across a wide range of categories, opening the potential for profiling and other privacy-invasive techniques" [32]. Acar et al. introduced a novel multi-stage privacy attack: through machine learning approaches, an adversary can automatically detect and identify types of devices, their actions, states, and related user activities by passively monitoring the wireless traffic of smart home devices. The intrusion works well on both encrypted and unencrypted communications, achieving very high accuracy (above 90%). To mitigate this privacy concern, they propose a new yet effective mitigation mechanism to hide the real activities of the users [33].

Similarly, Takbiri et al. investigate the fundamental limits of user privacy when both anonymization and obfuscation-based protection mechanisms are applied to users' time series of data [34].

However, the research of Ren et al. differs from the other two approaches by developing a machine learning approach which uniquely considers different possible interaction methods between user and device.

[20] Their methodology consisted in 34,586 repeatable experiments on 81 devices in two labs, one in the US at North-eastern University's Mon (IoT)r Lab and one in the UK at Imperial College London, over one month.

[21] Reference is made, here, to the data packets in the transport layer (TCP/IP) within the OSI/ISO model.

The three models have demonstrated that illicit profiling as well as privacy and data protection concerns arise, in the IoT domain, even when communications' content is not at stake. And even if it is encrypted. Data packets and thus related timestamps, *per se*, should not be considered *prima facie* personal data since they are the reading of the time-of-day clock when the structured query language (SQL) statement is executed at the application server[22]. These approaches have valuably demonstrated that an eavesdropper might be able to infer, through the recorded timestamp and the patterns they are able to create, valuable information regarding the user. Therefore, these patterns shall be categorized as personal data even when the underlying communication is encrypted: they can nonetheless relate to an individual, and/or the individual can be considered to be identified or identifiable [35].
Dumortier et al. noted that the European Court of Justice's Tele2 ruling (C-203/15) did not argue that metadata was sensitive by definition: "rather, the Court condemned the indiscriminate and universal collection of a very broad set of metadata, given that this data taken as a whole could establish a profile of the individuals concerned, in the context of potentially criminal activity. […] The processing of metadata, even in the context of electronic communications, can also have very limited data protection implications" [36]. Nevertheless, the above mentioned technical researches have been shown that even when GDPR's appropriate safeguards are established, like means of encryption, metadata have considerable privacy and data protection implications. Acknowledging that the ECJ has broaden the interpretation of personal data afterwards in the Nowak case (C-434/16) [23], the rationale behind this interpretation is in line with the objective reasoning of the Court of Justice of the European Union in Breyer v Germany[24]: the scope of personal data should be determined by assessing whether there are means reasonably likely to be used to identify individuals, and not merely a theoretical possibility of identification [37]. In our case, the means provided by timing patterns, which allow an adversary to "*reliably* learn a user's interactions with a device across a wide range of categories" [32], are more *reasonably likely to be used* for identification rather than remaining a *theoretical possibility*.
This interpretation must deal with two underlying issues. Firstly, it should be noted that even though the Breyer case offered the Court a chance "to confirm or reject the WP29 guideline 'implied identification = reasonably likely identification', the Court chose not to do so" [38]. The result of the evaluation of what the ECJ considered "identification measures reasonably likely to be taken" was elaborated in order to answer the narrow question posed, and in relation to the circumstances of the case.

[22] Cfr. with: https://www.ibm.com/support/knowledgecenter/SSFMBX/com.ibm.swg.im.dashdb.sql.ref.doc/doc/r0005886.html

[23] The Court ruled on the meaning of information relating to a person twice: in the Joint cases C-141/12 and C- 372/12 YS and M. and S. v Minister of Immigration, Integration and Asylum (2016) and in the Case C-434/16 Peter Nowak v Data Protection Commissioner (2017). The *relation* link of the former is interpreted by the Court narrowly, as information about an individual, thus rejecting the broader understanding of [36]. The ECJ endorsed the AG Sharpston's Opinion where, at line 56, considers as personal data "only information relating to facts about an individual". In the Novak case, the ECJ overturned this interpretation. Following the Opinion of AG Kokott, the Court stated that the notion personal data potentially encompasses any information, as long as it relates to the data subject, i.e. when the information is linked to a particular person by reason of its content, purpose or effect (Novak case, para 34-35).

[24] If the CJEU had applied the GDPR, rather than the Data Protection Directive, it would almost certainly have reached a similar conclusion [52].

Secondly, in considering what *means of identification* are reasonably likely to be used, the Court added the factor of legality to the conditions set out by WP29: ECJ ruled that identification would not be reasonably likely if prohibited by law. Purtova, however, rightly points out that the Court's reasoning shall be viewed in a more nuanced way, namely that "a legal prohibition to combine data for identification would make the means of identification 'less reasonably likely to be used', rather than 'not reasonably likely'"[38]. Interestingly enough, another recent research highlights the same privacy-risk scenario but in a context where timing patterns are publicly available: it shows that device identification in blockchain introduces privacy risks as the malicious nodes can identify users' activity pattern by analysing the temporal pattern of their transactions in the blockchain [39].

In conclusion, given that eavesdropping activities, such as sniffing, man-in-the-middle attack and spoofing, in relation to the timing patterns are considered contrary to the law[25], according to the interpretation endorsed by the Court (but contrary to what stated by WP29) these means of identification are less reasonably likely to be used, if we accept the reasoning of the Court. But, as a matter of fact, they can be used. And the resulting inference might lead to illegal profiling and other privacy-invasive techniques.

I argue therefore that the analysed case-study of IoT systems' timing patterns calls for a necessary reassessment of the abovementioned equation outlined by Article 29 Working Party: accordingly, such metadata are reasonably likely to identify.

III. Conclusion.

This paper has contributed to show that timestamps related to IoT devices' encrypted data packets can arguably be deemed personal data, even though, *prima facie,* they fall outside any definition of personal data. The case of illegal profiling from adversarial inferences of network timing patterns in IoT devices shows that serious privacy and data protection concerns may arise even though security techniques such as encryption are adopted.

Therefore, equating *every* kind of metadata with content, in terms of legal protection granted by the incoming ePrivacy Regulation, will be crucial. Nevertheless, it should be noted that Article 29 Working Party, commenting on the proposal, singles out as a point of "grave concern" the different level of protection accorded to content and metadata [40].

The necessary equalisation process might deal with the strong assumption made by Purtova, i.e. "if all data has a potential to impact people and is therefore personal, all data should trigger some sort of protection against possible negative impacts" [38]. It is certainly true that if nearly every aspect of the communication is personal data, and therefore data protection regime comes always into play, the "highly intensive and non-scalable regime of rights and obligations created by the GDPR will not simply be difficult but impossible to maintain in a meaningful way" [38]. Notwithstanding, Purtova claims that a broad understanding of personal data may still hold on for a while since the hyper-connectivity of our society has not yet reached a level of maturity such as to justify such a paradigm shift.

[25] Cfr. with, *inter alia,* the Italian Criminal Code: articles 615-bis and 615-ter consider the abusive access to a computer or telematic system, or the unlawful interference with individuals' privacy as criminal offences.

However, it may no longer be the case, as testified by the case of IoT devices' metadata. The European Commission released in February 2020 the first pillars of the new digital strategy, aiming at creating a single market for data that will ensure Europe's global competitiveness and data sovereignty. In "A European strategy for data", Brussels announces the creation of sector and domain-specific data spaces [41]. These common European data spaces will ensure that more data becomes available for use in economy and society, while keeping companies and individuals who generate the data under control.

However, the program still heavily relies on a sharp and non-problematic distinction between personal and non-personal data [26]. Thus, at the very first point of the appendix of the Data strategy document, it is argued that the intention of the Commission is to unleash the potential value of using non-personal data in industrial manufacturing, whereas "data generated by individuals are concerned, their interests should be fully taken into account in such a process and compliance with data protection rules must be ensured" [41]. It has been left unclear whether there is a boundary, in such a critical domain, in the relatability of data to individuals.

In this scenario, it is essential to create and consolidate a European data governance model[26], overcoming the ontological distinction of data, in order to complement and support the entire GDPR system.

IV. Future research.

Further research should aim at facing up and stimulating the debate on whether encrypted data could be deemed personal data or not in the context of IoT systems [42,43].

Acknowledgments:

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 814177.

References:

[1.] Kokott J, Sobotta C. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. Int Data Priv Law. 2013;3(4):222–8.

[2.] Docksey C, Hijmans H. The court of justice as a key player in privacy and data protection: An overview of recent trends in case law at the start of a new era of data protection law. Eur Data Prot Law Rev. 2019;5(3):300–16.

[3.] Lynskey O. Deconstructing data protection: The "added-value" of a right to data protection in the eu legal order. Int Comp Law Q. 2014;63(3):569–97.

[4.] Pagallo U. The Impact of Domestic Robots on Privacy and Data Protection. Data Prot Move [Internet]. 2016;24(October):387–410. Available from: http://link.springer.com/10.1007/978-94-017-7376-8

[5.] Chaudhuri A, Cavoukian A. The Proactive and Preventive Privacy (3P) Framework for IoT Privacy

[26] Cfr. with Pagallo et al. [53]: the authors suggest a middle way between bottom-up and top-down governance models, for the creation of a regulatory model based on hard law and self-regulation, also including the engagement of civil society.

by Design. EDPACS. 2018 Jan 2;57(1):1–16.

[6.] Cavoukian A. Privacy by design: origins, meaning, and prospects for assuring privacy and trust in the information era. IGI Glob [Internet]. 2012 [cited 2020 Feb 11];Privacy pr:170–208. Available from: https://www.igi-global.com/chapter/privacy-design-origins-meaning-prospects/61500

[7.] ENISA. Recommendations on shaping technology according to GDPR provisions: Exploring the notion of data protection by default. 2018.

[8.] Rayes A, Salam S. Internet of Things From Hype to Reality. Internet of Things From Hype to Reality. 2019.

[9.] EDPS. Meeting the challenges of big data A call for transparency, user control, data protection by design and accountability. 2015.

[10.] ENISA. Privacy by design in Big Data. An overview of privacy enhancing technologies in the era of big data analytics. 2015.

[11.] Eskens S. Profiling the European Citizen in the Internet of Things: How Will the General Data Protection Regulation Apply to this Form of Personal Data Processing, and How Should It? SSRN Electron J. 2016;

[12.] Hildebrandt M. Profiling and the identity of the European citizen. In: Profiling the European Citizen: Cross-Disciplinary Perspectives. Springer Netherlands; 2008. p. 303–43.

[13.] Pagallo U, Durante M, Monteleone S. What Is New with the Internet of Things in Privacy and Data Protection? Four Legal Challenges on Sharing and Control in IoT. In: Data protection and privacy: (in) visibilities and infrastructures. Springer, Cham; 2017. p. 59–78.

[14.] Article 29 Data Protection Working Party. Opinion 8/2014 on the on Recent Developments on the Internet of Things [Internet]. 2014 [cited 2020 Jan 31]. Available from: http://ec.europa.eu/justice/data-protection/index_en.htm

[15.] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2017.

[16.] European Parliament Research Service. Briefing on Free flow of non-personal data in the European Union [Internet]. 2017 [cited 2020 Feb 14]. Available from: http://www.europarl.europa.eu/RegData/etudes/BRIE/2017/614628/EPRS_BRI%282017%29614628_EN.pdf

[17.] Graef I, Gellert R, Purtova N, Husovec M. Feedback to the Commission's Proposal on a Framework for the Free Flow of Non-Personal Data. SSRN Electron J. 2018 Feb 7;

[18.] EDPS. Comments of the EDPS on a Proposal for a Regulation of the European Parliament and of the Council on a framework for the free-flow of non-personal data in the European Union [Internet]. 2018 [cited 2020 Feb 14]. Available from: www.edps.europa.eu

[19.] Tene O, Polonetsky J. Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. Northwest J Technol Intellect Prop [Internet]. 2013 [cited 2020 Mar 3];11(5):239–73. Available from: https://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

[20.] Big Data Public Private Forum. Big Data Technical Working Groups White Paper. 2014.

[21.] European Parliament's ITRE Committee by Blackman and Forge. Data Flows — Future Scenarios: In-Depth Analysis for the ITRE Committee [Internet]. 2017. Available from: http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/607362/IPOL_IDA(2017)607362_EN.pdf

[22.] Garante per la Protezione dei Dati Personali. Indagine Conoscitiva sui Big Data. 2020.

[23.] Zhao R, Ouyang W, Wang X. Unsupervised Salience Learning for Person Re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2013. p. 3586–93.

[24.] Zheng WS, Gong S, Xiang T. Person re-identification by probabilistic relative distance comparison. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society; 2011. p. 649–56.

[25.] EDPS. Opinion on safeguards and derogations under Article 89 GDPR in the context of a proposal for a Regulation on integrated farm statistics. 2017.

[26.] Graef I, Gellert R, Husovec M. Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation. SSRN Electron J. 2018 Oct 3;

[27.] Sweeney L. Simple Demographics Often Identify People Uniquely. Health (Irvine Calif). 2000;671:1–34.

[28.] Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets. In: 2008 IEEE Symposium on Security and Privacy. 2008. p. 111–25.

[29.] ENISA. Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics [Internet]. 2015 [cited 2020 Feb 18]. Available from: https://www.enisa.europa.eu/news/enisa-news/privacy-by-design-in-big-data-an-overview-of-privacy-enhancing-technologies-in-the-era-of-big-data-analytics

[30.] Conley C. Metadata: Piecing Together a Privacy Solution [Internet]. 2014. Available from: https://www.aclunc.org/sites/default/files/Metadata report FINAL 2 21 14 cover %2B inside for web %283%29.pdf

[31.] Palmirani M, Martoni M. Big data, governance dei dati e nuove vulnerabilità. POLITEIA. 2019;XXXV(136):9–22.

[32.] Ren J, Dubois DJ, Choffnes D, Mandalari AM, Kolcun R. Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. Proc Internet Meas Conf [Internet]. 2019 [cited 2020 Feb 26];267–79. Available from: https://doi.org/10.1145/3355369.3355577

[33.] Acar A, Fereidooni H, Abera T, Kumar Sikder A, Miettinen M, Aksu H, et al. Peek-a-Boo: I see your smart home activities, even encrypted! ArXiv Prepr arXiv180802741. 2018;

[34.] Takbiri N, Houmansadr A, Goeckel DL, Pishro-Nik H. Matching anonymized and obfuscated time series to users' profiles. IEEE Trans Inf Theory. 2019 Feb 1;65(2):724–41.

[35.] Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007.

[36.] Dumortier J, Somers G, Jacobs E, Graux H, Debusseré F, Van Camp S, et al. Legal memo with respect to the concept of metadata and its degree of sensitivity under future European e-privacy law. 2018.

[37.] Mourby M, Mackey E, Elliot M, Gowans H, Wallace SE, Bell J, et al. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. Comput Law Secur Rev. 2018 Apr 1;34(2):222–33.

[38.] Purtova N. The law of everything. Broad concept of personal data and future of EU data protection law. Law, Innov Technol. 2018 Jan 2;10(1):40–81.

[39.] Dorri A, Roulin C, Jurdak R, Kanhere S. On the Activity Privacy of Blockchain for IoT. IEEE 44th Conf Local Comput Networks [Internet]. 2019 Dec 21 [cited 2020 Mar 2];258–61. Available from: http://arxiv.org/abs/1812.08970

[40.] Article 29 Data Protection Working Party. Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation (2002/58/EC). 2017.

[41.] European Commission. A European Strategy for Data. 2020.

[42.] EDPS and AEPD. Introduction to the Hash Function as a Personal Data Pseudonymisation Technique. 2019.

[43.] Article 29 Data Protection Working Party. Statement of the WP29 on encryption and their impact on the protection of individuals with regard to the processing of their personal data in the EU [Internet]. 2018 [cited 2020 Feb 26]. Available from: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622229

[44.] Floridi L. The ontological interpretation of informational privacy. Ethics Inf Technol. 2005;7(4):185–200.

[45.] ENISA. How to implement security by design for IoT [Internet]. 2019 [cited 2020 Feb 13]. Available from: https://www.enisa.europa.eu/news/enisa-news/how-to-implement-security-by-design-for-iot

[46.] Rajagopalan R, Varshney PK. Data aggregation techniques in sensor networks: A survey [Internet]. 2006 [cited 2020 Feb 14]. Available from: https://surface.syr.edu/eecs/22

[47.] Pagallo U. The Group, the Private, and the Individual: A New Level of Data Protection? In: Group Privacy. Springer International Publishing; 2017. p. 159–73.

[48.] Taylor, L., Floridi, L., & Van der Sloot B (Eds. . Group Privacy: New Challenges of Data Technologies. Springer. 2016;126.

[49.] European Commission. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL: Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union [Internet]. 2019 [cited 2020 Feb 19]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:250:FIN

[50.] Ohm P. Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA l Rev. 2009;57:1701.

[51.] Article 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques [Internet]. 2014 [cited 2020 Feb 17]. Available from: http://ec.europa.eu/justice/data-protection/index_en.htm

[52.] Borgesius FZ. The Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition. Eur Data Prot Law Rev. 2017;3(1):130–7.

[53.] Pagallo U, Casanovas P, Madelin R. The middle-out approach: assessing models of legal governance in data protection, artificial intelligence, and the Web of Data. Theory Pract Legis. 2019;7(1):1–25.