Assuring Compliance of European Smart Tourist Destinations with the Principles of the General Data Protection Regulation: a roadmap[*] [*]

Manuel David Masseno[***]
Cristiana Santos[****]

[*] Paper accepted, following a Call for Papers, and presented at the *2nd UNWTO World Conference on Smart Destinations*, convened by the United Nation World Tourism Organization and the Government of the Kingdom of Spain, at Oviedo, the 26th June 2018.

[**] This paper was drafted within the framework of the Research Project: "*Big Data, Cloud Computing* y otros retos jurídicos planteados por las tecnologías emergentes; en particular, su incidencia en el sector turístico" – DER2015 – 63595 (MINECO/FEDER). Coordinated by Professor Apollònia Martínez Nadal at the *Universitat de les Illes Balears*, Spain.

[***] Professor Adjunto no Instituto Politécnico de Beja, onde também integra a Coordenação do Laboratório UbiNET – Segurança Informática e Cibercrime e do Mestrado em Engenharia de Segurança Informática. Membro do Grupo de Estudos Temático em Direito Digital e *Compliance* da FIESP – Federação das Indústrias do Estado de São Paulo, Brasil, desde 2016. Investigador no IJI – Instituto Jurídico Interdisciplinar da Faculdade de Direito da Universidade do Porto, desde 2005, Investigador no Projeto I+D "*Big Data, Cloud Computing y otros nuevos retos jurídicos planteados por las tecnologías emergentes; en particular, su incidencia en el sector turístico*" (DER2015-63595-R MINECO/FEDER), coordenado pela *Universitat de les Isles Balears*, desde 2016. Assessor Jurídico Independente no Projeto da União Europeia (Horizonte 2020) PoSeID-on – Protection and control of Secured Information by means of a privacy enhanced Dashboard (ID: 786713), desde 2018.

[****] Mestre em Direito dos Contratos e da Empresa pela Universidade do Minho (2012), Doutora em Direito, Ciência e Tecnologia pelo *Joint International Doctoral Program* (LAST-JD) – ERASMUS MUNDUS, das Universidades de Bolonha, Turim, Luxemburgo e Autónoma de Barcelona, e em Informática, pelas Universidades de Bolonha e do Luxemburgo (2017). Investigadora colaboradora no Centro de Investigação em Justiça e Governança – JusGov da Universidade do Minho, desde 2013; Advogada, desde 2011.

Abstract: This paper aims to provide a consistent answer to the concerns regarding privacy and data protection within the framework of Smart Tourism Destinations (STD) that tourism science has given rise to, given the applicability of the new General Data Protection Regulation of the EU (GDPR). Our main result provides a roadmap for compliance of STD design and management with the core principles embodied in the GDPR, providing guidelines both for public and private sectors and for other stakeholders, namely for citizens-tourists. With this work we intend to help achieve fully privacy-compliant STD, in Europe and elsewhere.

Keywords: *Privacy and Data Protection, GDPR, Regulation, Smart Tourism Destinations*

Resumo: Este estudo pretende dar uma resposta consistente às preocupações relativas à privacidade e à proteção de dados no contexto dos Destinos Turísticos Inteligentes (DTIs), as quais foram postas em evidência pela ciência do turismo, tendo em mente o novo Regulamento Geral sobre Proteção de Dados da UE (RGPD). O nosso principal resultado consiste em enunciar um esboço de roteiro para aferir da conformidade da conceção e gestão dos DTIs com os princípios fundamentais do RGPD, facultando orientações tanto ao setor público quanto ao privado e a outros interessados, como os cidadãos-turistas. Com este trabalho pretendemos ajudar a alcançar uma plena conformidade das STDs com a privacidade e a proteção de dados, na Europa e não só.

Palavras-chave: *Privacidade e Proteção de Dados, RGPD, Regulação, Destinos Turísticos Inteligentes*

Introduction

STD are an offspring of the technological foundations of *Smart Cities*. They benefit from the interplay between other technological environments based on the IoT and the *Cloud*, as enabled by *Big Data Analytics*. However, the connections between STD and Privacy & Data Protection did not receive significant attention within legal research[1],

[1] For the legal theoretical framework of this paper, see our recently published articles, such as Masseno, Manuel David; Santos, Cristiana. "Between Footprints: Balancing Environmental Sustainability and Privacy in Smart Tourism Destinations", in *Unitedworld Law Journal*, v. 1-II, 2017, p. 96-118, accessed 30/07/2018 <https://www.unitedworldschoollawjournal.com/wp-content/uploads/2018/05/Between-Footprints-Balancing-Environmental-

even if it was perceived and identified as an overlooked issue by tourism science[2].

Basically, these technology-enhanced tourism services allow tourists to get more from their travel and helps them to fulfil the experiential potential of their destination.

However, ICT embedded within STD also permit the collecting and analysis of large amounts of data (for example, to enable the identification of attitude patterns and to predict the behavior of tourists or travelers). This is achieved by identifying their potential needs and desires even at an unconscious level. Hence, these experiences are achieved through intensive personalization, context-awareness and real-time monitoring, which involve processes of information management that entail legal risks, demanding a careful analysis of the data protection framework. As a large spectrum of user-generated content processed in a STD concern personal data and human interaction, there is a direct impact on individuals and their rights regarding the processing of personal data.

Moreover, the application of the GDPR, which came into effect on 25 May 2018, renders urgent a review of the current conceptions and practices regarding privacy concerns STS to ensure compliance.

Nevertheless, while realizing the benefits of using big data analytics and being a competitive STD, addressing data protection issues supports good practice in information governance that organizations utilizing STD should closely assess. Therefore, data protection compliance should be

Sustainability-and-Privacy-in-Smart-Tourism-Destinations-by-Manuel-David-Masseno-and-Cristiana-Santos-1.pdf,> and Masseno, Manuel David; Santos, Cristiana. "Assuring Privacy and Data Protection within the Framework of Smart Tourism Destinations", *MediaLaws – Rivista di Diritto dei Media*, 2018, n. 2, p. 251-266, accessed 30/07/2018 <http://www.medialaws.eu/rivista/assuring-privacy-and-data-protection-within-the-framework-of--smart-tourism-destinations/>.

[2] Namely, Anuar, Faiz I.; Gretzel, Ulrike. "Privacy Concerns in the Context of Location-Based Services for Tourism", in *ENTER 2011 Conference. Accessibility of ICTs and Accessible Travel Information*, Innsbruck, Austria, 2001, accessed 30/07/2018 <http://agrilifecdn.tamu.edu/ertr/files/2013/02/13.pdf>; Buhalis, Dimitrios; Amaranggana, Aditya. "Smart Tourism Destinations", *Information and Communication Technologies in Tourism 2014 – Proceedings of the International Conference in Dublin, Ireland*, Heidelberg: Springer, 2014, pp. 553-564; or Gretzel, Ulrike; Sigala, Marianna *et al.* "Smart tourism: foundations and developments", *Electronic Markets*, v. 25, n. 3, 2015, p. 179–188.

an enabler of the success of STD and not a regulatory or administrative burden.

The paper is organized as follows. Section 2 outlines some of the most important risks attributable to STD regarding privacy and data protection. Section 3 describes the obligations of the organizations processing personal data, according to the GDPR[3], which constitute the current basis of the EU-wide legal obligations regarding privacy and data protection. Section 4 refers to the compliance tools which confirm to the above-mentioned legal obligations. Section 5 concludes the paper.

1. The Risks of Smart Tourism Destinations for Privacy and Data Protection

In this section we explain some of the potential risks STD technologies entail for privacy and data protection. As is increasingly appreciated, the use and combination of advanced techniques of *Big Data Analytics*, which include machine learning, data mining techniques, etc., enhance the common risks to privacy and data protection. The following are enhanced when information (*e.g.* mobility data) is connected and matched with data from other sources of publicly available information (e.g., *Facebook* or *Twitter* postings, reviews at *Booking* or at *TripAdvisor*, blogs entries, etc.) and analysis revealed users' social interactions and activities, as is the case with smart tourist travel cards.

1.1. Identification and re-identification of individuals from allegedly anonymized or pseudonymized data

These concerns stem from the fact that integrating large collections of data from distinct sources of available tourism datasets, even with

[3] Regulation (EU) 2016/679, of the EP and of the Council of 27/04/2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), accessed 30/07/2018 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG>.

apparently innocuous, non-obvious or anonymized resources, may enhance a jigsaw of indirect correlation of identification and re-identification; this scenario could escalate if massive information resources via the web are available[4]. Thereby, personal information set through re-identification intrinsically conforms with legal requirements, as identification not only means the possibility of retrieving a person's name and/or address, but also includes potential identifiability by singling out, *linkability* and inference[5].

As data collected by the ubiquitous computing sensors is, in principle, personal data[6] or personally identifiable information, the processing of non-sensitive data can lead, through data mining, to data that reveals personal or sensitive information, thus blurring the conventional categories of data.

1.2. Covert profiling of individuals and non-transparency of the processing

Profiling is an important feature in tourism destinations. Tourism service providers are adapting their approach to service by meeting the personalized expectations of customers. Data-processing scenarios collect user's input and feedback which are used to build fine-grained premium services and recommender systems in the form of trail packages. The richer the user profile, the higher the temptation for operators to target a user with unsolicited advertising or to engineer a pricing structure

[4] ART 29 WP – Article 29 Working Party of the European Union: Opinion 7/2003, on the re-use of public sector information, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2003/wp83_en.pdf>; Opinion 3/2013, on purpose limitation, accessed 30/07/2018<http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf>; and, Opinion 6/2013, on open data and public-sector information (PSI) reuse, accessed 30/07/2018<http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf>.

[5] ART 29 WP Opinion 05/2014, on anonymization techniques, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

[6] ART 29 WP Opinion 4/2007, on the concept of personal data, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf>, consulted on 15/06/2018.

designed to extract as much surplus from the user as possible[7]. Notably, "[...] analytics based on information caught in an IoT environment might enable the detection of an individual's even more detailed and complete life and behavior patterns."[8]. However, as a norm, the GDPR prohibits automated individual decision-making that significantly affect individuals in Art. 22 (1).

Indeed, development of consumer-tourist automated profiles, facilitated by Big Data Analytics, can *significantly affect* data subjects[9]. Covert profiling, in certain cases, may lead to unintended consequences: i. when based on incomplete data, profiling can lead to false negatives, depriving individuals from benefits that they would be entitled to; ii. the so called, "*filter bubbles*" effect, according to which data subjects will only be exposed to content which confirms their own preferences and patterns, without any door open to serendipity and casual discovery; iii. isolation and/or discrimination.

Besides, within a STD, machine learning decisions and profiling can lead to direct or indirect discrimination through the exclusion/denial of services/goods (e.g. denial of insurance, exclusion from the sale of tourist services or high-end products, shops or entertainment complexes of certain profiles of tourists and even decisions that impact upon health, creditworthiness, recruitment, insurance risk, etc). It can even lead to discrimination in relation to essential utilities for those unwilling to share personal data. As indicated, tourists may be discriminated against if they belong to a certain social group, but also this categorisation might be based on factors, identified by the analytics, that they share with members of that group. Therefore, to ensure a fair and transparent processing (as determined by the principle of fairness and transparency), automated

[7] ENISA – European Networks and Information Security Agency. 2015 Report on Privacy and Data Protection by Design – from policy to engineering, accessed 30/07/2018 <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design/at_download/fullReport>, .

[8] ART 29 WP Opinion 8/2014, on the recent developments on the Internet of Things, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf>.

[9] EDPS – European Data Protection Supervisor. Opinion 3/2015, Europe's big opportunity, EDPS Recommendations on the EU's options for data protection reform, accessed 30/07/2018 <https://edps.europa.eu/sites/edp/files/publication/15-10-09_gdpr_with_addendum_en.pdf>.

decisions should take account of all the circumstances concerning the data and not be based on merely de-contextualized information or on data processing results[10]. Moreover, the data controller should find ways to build discrimination detection into their machine learning systems, to prevent inaccuracies and errors being assigned to labeled profiles, as referred in Recital 71 of GDPR[11].

1.3. Repurposing of data

Data analytics can mine data for new insights and find correlations between apparently disparate datasets. Hence, automatic capture of big data can be frequently reused[12] for secondary unauthorized purposes, profiling, or for abusive marketing activities, undermining the purpose specification principle, which states that the purposes for which data is collected must be specified and lawful (Art. 5(1) (b)). As for repurposing, personal data should not be further processed in a way that the data subject might consider unexpected, inappropriate or otherwise objectionable[13] and, therefore, unconnected to the delivery of the service.

1.4. Surveillance under the disguise of service provision and its desensitizing effect

On the other hand, the data subject's interactions within a STD will be increasingly mediated by or delegated to (smart) devices and apps. Most of

[10] ART 29 WP Guidelines on Transparency under Regulation 2016/679, accessed 30/07/2018 <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227>.

[11] ART 29 WP Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, accessed 30/07/2018 <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>.

[12] ART 29 WP Opinion 3/2013, on purpose limitation, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> .

[13] COE – Council of Europe Guidelines on the Protection of individuals with regard to the processing of personal data in a world of Big Data, T-PD, 2017, accessed 30/07/2018 <https://rm.coe.int/16806ebe7a>.

the destinations are using video-surveillance systems as sensors to supply real-time information on public transportation, traffic (also in relation to emergency and personal safety), navigation, and access to tourist information on the go, all of which provide value to the user: safety, convenience, and utility in daily lives, as well as on vacation.

This information is transmitted via, for e.g., smart remote controllable digital Closed-Circuit Television cameras that can zoom, move and track individual pedestrians, Automatic Number Plate Recognition, GPS, Wi-Fi network tracking reliable facial recognition software, and location-based service apps.

It has been argued that such devices desensitize users to providing location-based information because of the ease with which it happens and the "coolness" factor that comes with it.

1.5. Failed consent

Within this sort of intelligent environment, it is problematic to give, or withhold, our prior consent to data collection, as it seems to be absent by design. These ubiquitous sensors are so embedded in the destination that there is little awareness of them, or none at all; thus, they literally "disappear" from the users' sight. Users will not even be conscious of their presence and hence the notion of consent to the collection of data is problematic. We may, at least to some extent, concede that obtaining such consent, in STD contexts, would be achieved in a mechanical or perfunctory manner, or as a "routinization".

We also perceive with regard to Closed-Circuit Television, Automatic Number Plate Recognition and Webcams whilst tracking and sensing that notice in the form of information signs in the area being surveilled, or on related websites, would not conform to the consent requirements. Thus, the main issue of the *IoT* embedded in STD is that its sensorization devices are explicitly designed to be unobtrusive and seamless, invisible in use and imperceptible to users and thereupon, users do not have the opportunity to give their unambiguous, informed, specific, explicit, and granular consent[14].

[14] ART 29WP Opinion 15/2011, on the definition of consent, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/

Therefore, the data controller might have difficulty in demonstrating that consent was given, and the data subject is not able to withdraw that consent. Still, consent is not yet part of a function specification of *IoT* devices, and thus, they do not have the means to "provide fine-tuned consent in line with the preferences expressed by individuals," because smart roads, trams, tourist office devices are usually small, screenless and lack an input mechanism (a keyboard or a touch screen)[15].

1.6. Imbalance

Smart technologies often produce situations of imbalance, where data subjects are not aware of the fundamental elements of data processing and related consequences, are unable to access and manage their information, which leads secondarily to an enhanced information asymmetry as consumers.

1.7. Tendency to collect and analyze all data

The tourism industry is inherently based on data-exchange: to generate massive databases, it is necessary to optimally exploit all information available and thus, datasets need to be as exhaustive and varied as possible in order to faithfully reflect tourist activity within a territory.

In substance, smart technology undertakes the extensive collection, aggregation and algorithmic analysis of all the available data for various reasons, such as understanding customer purchasing behaviour and patterns or remarketing based on intelligent analytics, hampering the data minimization principle (Art. 5 (1)(c)). In addition, irrelevant data is also being collected and archived, undermining the storage limitation principle (Art. 5 (1) (e)).

wp187_en.pdf> ; updated by its Guidelines on Consent under Regulation 2016/679, accessible at <http://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030>.

[15] ART 29 WP Opinion 8/2014, on the recent developments on the Internet of Things, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf>.

1.8. Inaccurate data

Where information sources are not trustworthy, results drawn from data analysis may also not be representative or accurate (*i.e.*, analysis based on social media sources are not necessarily representative of the whole population at issue).

Besides, machine learning itself may contain hidden bias which lead to inaccurate predictions and profiles of individuals. In any case, profiling involves creating derived or inferred data, occasionally leading to incorrect decisions (discriminatory, erroneous and unjustified), regarding their behaviour, health, creditworthiness, recruitment, insurance risk, etc.

Even exercising the "*right to be forgotten*", where data subjects have the right for their data to be erased in several situations (e.g., when the data is no longer necessary for the purpose for which it was collected, or based on inaccurate data (as set by the accuracy principle depicted in Art. 5 (1) (d)), it may in reality be difficult for a business to find and erase someone's data if it is stored across several different systems and jurisdictions.

2. The obligations of organizations while processing personal data within a STD

While realizing the benefits of using big data analytics and being a competitive STD, addressing data protection concerns supports best practices in information governance. Accordingly, it is in the interests of organizations intending to become an STD should pay careful attention to these issues. Data protection compliance should hence be viewed as an enabler of the success of an STD and not as a regulatory or procedural burden. As is by now widely known, infringement or non-compliance with the Regulation may lead to fines up to €20 million, or 4% of the worldwide annual revenue of the prior financial year, whichever is higher.

As stated in the tourism literature, tourism, by definition, is a service-intensive industry with a "*business network*", since it relies on a number of stakeholders for its ability to deliver products and services. In this network, each of the actors involved in the transportation, accommodation, gastronomy, attractions and ancillary services, potentially process personal data.

For a STD, the public or private organisations that decide the "whys" and "hows" by which the personal data is to be processed are called "data controllers". They may use other parties that process personal data on their behalf, called "data processors". Both data controllers and data processors must abide by the GDPR obligations.

However, Big Data Analytics can make it difficult to distinguish between controllers and processors; further, within the modern data value chain, organizations outsourcing analytics and artificial intelligence to specialized companies need to consider carefully who has control over the processing of any personal data (Art. 4 (7) (8)).

Therefore, if an organization chooses to store its customer data in the cloud, then the cloud provider is likely to be a data processor, as it is acting on the original organization's behalf, and it is not determining the purpose of the processing.

Hence, if an organization purports aims to conduct its analytics outsourcing in a data controller-data processor relationship, it is important that the contract includes clear instructions about how the data can be used and the specific purposes for which it is being processed. However, it does not follow from the existence of a contract of this type that the sub--contracted company performing data analysis is a data processor; if this company uses its discretion and expertise to decide what data to collect and how to apply its analytic techniques, then it is very likely to be a data controller as well; in fact a co-controllership[16] (Art. 24).

Under the accountability principle (Art. 24), data controllers shall be responsible for, and be able to demonstrate compliance with, all the obligations and principles contained in the regulation. Some of its most important obligations are explained below.

2.1. Appointing a Data Protection Officer

The GDPR mandates the appointment of a DPO within the organization whose responsibilities include: monitoring data governance and privacy,

[16] ICO, "Guide on Big Data, Artificial Intelligence, Machine Learning and Data Protection", 2017, accessed 30/07/2018 <https://ico.org.uk/for-organisations/guide-to-data-protection/big-data/>.

providing advice, monitoring data protection impact assessments, and acting as the point of contact with any supervisory authority. This is mandatory where the processing is carried out by a public authority or body, except for the courts; their core activities consist of processing operations which require regular and systematic monitoring of data subjects on a large scale, or processing on a large scale of special categories of data (Articles 37 to 39)[17].

2.2. *Algorithmic accountability*

Organizations should also check "algorithmic accountability", which means being able to check that the algorithms used and developed by machine learning systems are actually doing what we think they are doing and are not producing discriminatory, erroneous or unjustified results. Organizations using machine learning techniques in STD are obliged to assure data quality by checking the sources of the data, the accuracy of the data, whether is sufficiently up to date, how securely it is kept, and whether there are restrictions on how it can be used (anonymized data).

2.3. *Fair, lawful and transparent processing obligations*

STD organizations must process personal data "fairly, lawfully and in a transparent manner in relation to the data subject", i.e., when the data is collected, it must be clear as to why that data is being collected and how the data will be used. Whether the data is volunteered, observed, inferred, or collected from accessible sources, individuals are fully entitled to know what it is, from where and from whom the controllers obtained it, and how automated decisions were taken in relation to it. The GDPR prohibits automated individual decision-making that significantly affect individuals (Art. 22 (1)). Therefore, in order to ensure fair and transparent processing, automated decisions should take account of all the circumstances

[17] ART 29 WP Guidelines on Data Protection Officers ('DPOs'), accessed 30/07/2018 <http://ec.europa.eu/newsroom/document.cfm?doc_id=44100>.

surrounding the data and not be based on merely de-contextualized information or on data processed results. The controller should furthermore build discrimination detection into their machine learning systems, to prevent inaccuracies and errors being assigned to labeled profiles.

2.4. Lawfulness of processing

Processing personal data should be based upon certain conditions, namely: the consent of the tourist, a contract, a public interest, a legitimate interest, etc. In these intelligent environments, our ability to give or withhold our prior consent to data collection is questionable, as it seems to be absent by design. Within STDs, it should be acknowledged that ubiquitous sensors are so embedded in the destination that they literally "disappear" from the users' sight, meaning that users will not even be conscious of their presence and hence, by definition, do not consent to the collection of data. So, at least to some extent, the obtaining of consent in STD contexts can at best be mechanical, perfunctory, or routinized.

With reference to the remaining legal criteria, processing personal data relies on "public interest", which can sidestep the need for consent (health, national governmental agencies gather data – for e.g. e-Government systems, e-Health). Nevertheless, this possibility should not conceal any eventual "third-party interest".

Most commercial systems rely on the "legitimate interests" ground, even if they consist of "the vaguest ground for processing". This offers considerable scope for industry to process data by claiming any purportedly necessary "legitimate interest". In fact, the processing must be "necessary" for legitimate interests and not just *potentially* interesting for the operator. It follows that the processing is unnecessary if there is any other means of meeting that legitimate interest which interferes less with public privacy.

As for the contractual condition, it may be difficult to show that big data analytics in STD are strictly necessary for the performance of a contract, since the processing goes beyond what is required to sell a product or deliver a service.

2.5. Purpose limitation

The principle of purpose limitation is to ensure that the purpose for which the data is collected is specified and lawful. This principle also prevents arbitrary re-use, which means that personal data should not be further processed in a manner that the data subject might considered unexpected, inappropriate or otherwise objectionable and therefore unrelated to the delivery of the service. In other words, exposing data subjects to different/greater risks than those contemplated by the initial purposes may be considered to amount to the further processing of data in an unexpected manner[18].

2.6. Data Minimization, Collection and Retention obligations

Data minimization means that personal data shall be "*adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed*" (Art. 5 (1) (c)). This obligation means that STD entities should minimize the amount of data they collect and process, and the length of time they keep the data. Even if in practice, smart technology envisages the massive collection, aggregation and algorithmic analysis of all the available data to understand customer buying behavior and patterns, or remarketing based on intelligent analytics, organizations need to be clear about which data is deemed to be *necessary, excessive* and *relevant* for processing purposes.

As for data storage, personal data shall not be kept (stored) longer than necessary for the purpose for which it is being processed, as prescribed by the storage limitation principle (Art. 5 (1) (e)). This obligation is part of the lifecycle governance strategy retention policies of companies that defensibly dispose of irrelevant data rather than keeping data archived forever.

Regarding retention timeframes, retention schedules allow unnecessary data to be disposed of, as it is no longer of business value or needed to meet

[18] ART 29 WP Opinion 3/2013, on purpose limitation, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf>.

legal obligations. Data mapping techniques may permissibly identify where and what type of data is stored within an organization. Data management segmentation can also help to segregate EU data from data coming from other data subjects.

2.7. Accuracy and up to date processing obligations

If sources of data are reliable, accurate and representative, so too must be the results drawn from big data analysis employed in a STD environment (Art. 5 (1) (d)). For example, analysis based on social media sources are not necessarily representative of the population as a whole[19].

Organizations employing machine learning algorithms need to consider the distinction between correlation and causation[20], *i.e.*, when there is no *direct cause and effect* between two phenomena that show a close correlation. In these cases there is a risk of drawing inaccurate, but also – and when applied at the individual level – potentially unfair and discriminatory conclusions[21]. The potential accuracy (or inaccuracy) of any resulting decisions might cause discriminatory, erroneous and unjustified decisions regarding the data subject´s behavior in relation to their health, creditworthiness, recruitment, insurance risk, etc. The quality of the profiles and of the personal data upon which they are built, again, seem to matter for the prosperity of the industry.

2.8. Data breach reporting

EU data protection law requires controllers to promptly notify the relevant supervisory authority and the data subjects of potential data breaches in the event of causing a high risk to data subjects. The notification must include at least: the name and contact details of the DPO (or other relevant point of contact); the likely consequences of the data breach; and

[19] ICO, Guide on Big Data, cit.

[20] ICO, Guide on Big Data, cit.

[21] EDPS, Opinion 7/2015 on Meeting the challenges of big data, accessed 30/07/2018 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf>.

any measures taken by the controller to remedy or mitigate the breach. However, the controller may be exempt from this requirement if the risk of harm is remote because the affected data are protected (e.g., due to strong encryption). Most importantly, if the risks associated with the breach have been effectively resolved, then the organization may be exempt from the notification requirements[22].

2.9. Processing activities records

EU data protection law requires organizations involved in STD to keep records (written or electronic) of their data processing activities (art. 30). Examples of records to be kept include the purposes of the processing; the categories of data subjects and personal data processed; and the categories of recipients with whom the data may be shared. Upon request, these records must be disclosed to DPAs.

2.10. Codes of conduct and certification mechanism

In order to enhance transparency and compliance with this Regulation, associations and other institutional bodies representing both controllers and processors are obliged to elaborate codes of practice specifying how the GDPR should be applied. These bodies must then submit their draft codes of conduct to the relevant supervisory authority for approval. The GDPR introduced certification mechanisms and data protection marks, allowing data subjects to quickly assess the level of data protection employed by the products and services in question. A list of certified organizations will thus be publicly available. Codes of conduct and approved certification mechanisms will also assist controllers in identifying the risks related to their type of processing and in adhering to best practices.

[22] ART 29, Guidelines on Personal data breach notification under Regulation 2016/679, accessed 30/07/2018 <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612052>.

3. Compliance tools at the GDPR

Compliance tools enable STD organizations to meet their data protection obligations while protecting people's privacy rights in a STD context. These are: anonymization and pseudonymization techniques, privacy policies, DPIA, personal data stores, algorithmic transparency, privacy seals/certification, and PbD measures to mitigate identified legal risks and implications. STD managers may demonstrate commitment to compliance through internal documentation and employee training in relation the GDPR-related mandates, such as via written internal policies.

3.1. Anonymization

As a stated principle, when data is rendered *anonymous* (Recital 26 of the GDPR) all identifying elements have been irreversibly eliminated from a set of personal data, and allows no possibility to re-identify the person(s) concerned. Consequently, it is deemed to be no longer personal data. Later, anonymised data might be aggregated in order to be analysed and to gain insights about the population, as well as combined with data from any other sources. At this stage, *IoT* developers can analyse, share, sell or publish the data without any data protection requirements.

Conversely, de-anonymization strategies in data mining techniques entails that anonymous data is cross-referenced with other sources to re-identify the anonymous data. Thus, the processing of datasets rendered anonymous may never be absolutely ensured.

In what refers to *pseudonymized* personal data, identifiers are replaced by a pseudonym (through encryption of the identifiers). In turn, pseudonymized data continues to allow an individual data subject to be singled out and linkable across different datasets and therefore stays inside the scope of the legal regime of data protection[23].

[23] ART 29, WP Opinion 05/2014, on anonymization techniques, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

3.2. Privacy policies

Privacy policies consist of documents which set forth an organization's data practices on processing activities of personal data to its users, such as collection, use, sharing, and retention. They serve as a basis for decision-making and as a "tool for preference-matching" for consumers, as consumers tend to place a higher value on a product/service, after learning more about its attributes and tradeoffs. As such, Privacy Policies constitute the *locus* where consequences are produced, the "technically most feasible place to protect privacy and personal data"[24].

The GDPR states that information addressed to the data subject should be "concise, easily accessible and easy to understand, and that clear and plain language, and additionally, where appropriate, visualization is used" (Article 12(7) and Recital 60).

However, in a STD scenario, these requirements can be problematic, and it has been suggested that privacy notices are not feasible when Big Data Analytics are entailed, given that: travelers engaged in tourism are unwilling to read lengthy legalese such as privacy notices, since it would take significantly more time than they spend using the content or the app itself; the context in which data is collected (e.g., destination apps, wearable watches and glasses or IoT devices) is difficult to provide the information.

Regarding the amount and type of these interactions, it is just too onerous for each data subject to assess their privacy settings across dozens of entities in order to ponder the non-negotiable trade offs of agreeing to privacy policies without knowing how the data might be used now and in the future, and to assess the cumulative effects of their data being merged with other datasets. On the other hand, information can be delivered in a user-friendly form, namely by: videos or in-app notices; cartoons and standard icons applied to privacy notices, explaining their content. As for wearable devices, privacy information could be provided on the device itself, or by broadcasting the information via Wi-Fi or making it available through a QR code[25].

[24] President's Council of Advisors on Science and Technology, Big Data and Privacy: a Technological Perspective. Executive Office of the President, USA (2014), accessed 30/07/2018 <https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf>.

[25] ART 29, WP Opinion 8/2014, on the recent developments on the Internet of Things, accessed 30/07/2018 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf>.

3.3. Data protection impact assessment

A DPIA is a tool that can help to identify and mitigate privacy risks before the processing of personal data. This assessment involves description of the envisaged processing operations, an evaluation of the privacy risks and the measures contemplated to address those risks.

Art. 35 of GDPR indicates that when a type of processing which uses a systematic and extensive evaluation of individuals based on automated processing and profiling is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged operations on the protection of personal data. It is likely that general big data applications within an STD involving the processing of personal data will fall into this category[26].

3.4. Privacy by design

PbD is an approach in which IT system designers seek to adopt preemptive *technological* and *organizational* measures to protect personal data, when designing or creating new products and services. By design solutions are necessary at the early development stage (planning and implementation) of any new product or service that affects personal data. It aims to address privacy concerns attached to the very same technology that might create risks (Art. 25).

Besides anonymization techniques, PbD involves other engineering and organizational measures, including: security measures such as access controls, audit logs and encryption; data minimization measures, to ensure that only the personal data that is needed for a particular analysis or transaction is processed at each step (such as validating a customer); purpose limitation and data segregation measures so that, for example, personal data is kept separately from data used for processing intended to detect general trends and correlations; as well as sticky policies which

[26] ART 29, WP Guidelines on Data Protection Impact Assessment (DPIA), accessed 30/07/2018 <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236>.

record individual preferences, and corporate rules within the metadata that accompanies data.

Within a STD scenario, controllers and processors should test the adequacy of the above-mentioned solutions by design on a limited amount of data by means of simulations before they are used on a larger scale. Such a learn-from-experience approach makes it possible to assess the potential bias inherent in using different parameters in analyzing data, and provides a rational for minimising the use of information. However, there is a lack of a privacy mindset in IT system designers. As stated by ENISA: "[...] privacy and data protection features are, on the whole, ignored by traditional engineering approaches when implementing the desired functionality. This ignorance is caused and supported by limitations of awareness and understanding of developers and data controllers as well as lacking tools to realize PbD. While the research community is very active and growing, and constantly improving existing and contributing further building blocks, it is only loosely interlinked with practice."[27]

3.5. Personal data spaces

The European Data Protection Supervisor suggested that one way to increase an individual's control over the use of their data is through what are usually called personal data spaces, vaults or stores, which are often provided by personal information management services[28].

These are third-party services (intermediaries) that collect, manage and store people's personal data on their behalf and make it available to organisations as and when the individuals wish to do so. This tool aims to address criticisms related to the lack of control over how personal data is used in a big data environment, as tourists are not aware of how data is being collected or how it is used, and do not have the time to read privacy notices.

[27] ENISA, 2015, Report on Privacy and Data Protection by Design, cit.

[28] EDPS, Opinion 7/2015, cit.