Data Protection in the Era of Artificial Intelligence: Trends, Existing Solutions and Recommendations for Privacy-Preserving Technologies

Tjerk Timan and Zoltan Mann

Abstract This chapter addresses privacy challenges that stem particularly from working with big data. Several classification schemes of such challenges are discussed. The chapter continues by classifying the technological solutions as proposed by current state-of-the-art research projects. Three trends are distinguished: (1) putting the end user of data services back as the central focal point of Privacy-Preserving Technologies, (2) the digitisation and automation of privacy policies in and for big data services and (3) developing secure methods of multi-party computation and analytics, allowing both trusted and non-trusted partners to work together with big data while simultaneously preserving privacy. The chapter ends with three main recommendations: (1) the development of regulatory sandboxes; (2) continued support for research, innovation and deployment of Privacy-Preserving Technologies; and (3) support and contribution to the formation of technical standards for preserving privacy. The findings and recommendations of this chapter in particular demonstrate the role of Privacy-Preserving Technologies as an especially important case of data technologies towards data-driven AI. Privacy-Preserving Technologies constitute an essential element of the AI Innovation Ecosystem Enablers (Data for AI).

Keywords Data protection · Artificial Intelligence · Big data · Challenges · Future directions

T. Timan (✉)

Strategy, Analysis & Policy Department, TNO, The Hague, The Netherlands

e-mail: tjerk.timan@tno.nl

Z. Mann

paluno, University of Duisburg-Essen, Essen, Germany

© The Author(s) 2021

E. Curry et al. (eds.), *The Elements of Big Data Value*,

https://doi.org/10.1007/978-3-030-68176-0_7

1 Introduction

One of the challenges of big data analytics is to maximise utility whilst protecting human rights and preserving meaningful human control. One of the main questions in this regard for policymakers and lawmakers is to what extent they should allow for automation of (legal) protection in an increasingly digital society. This chapter contributes to this debate by looking into different technical solutions developed by the projects of the Big Data Value Public-Private Partnership (BDV PPP) that aim to protect both privacy and confidentiality whilst allowing for big data analytics. Such Privacy-Preserving Technologies are aimed at building privacy-by-design from the start into the back end and front end of digital services. They make sure that data-related risks are mitigated both at design time and run time, and they ensure that data architectures are safe and secure. In this chapter, we discuss recent trends in the development of tools and technologies that facilitate secure and trustworthy data analytics and provide recommendations based on the insights and outcomes of the projects of the BDV PPP and from the task forces of the Big Data Value Association (BDVA), combined with insights from recent debates and the literature.

1.1 *Aim of the Chapter*

The aim of this chapter is to provide an overview of trends in Privacy-Preserving Technologies and solutions as currently developed by research projects that are part of the Big Data Value Public-Private Partnership (BDV PPP). In the chapter, we focus on providing an overview of technical solutions for privacy and data protection challenges posed by Big Data and AI developments. The main particularity of big data is the number of data sources and the heterogeneousness of these sources. In many cases this leads to a mix of datasets that contain both personal and non-personal data. Combinations and aggregations of datasets in turn lead to new data. Mixing and reusing data on a large scale and at high velocity makes many forms of protection of data difficult, and enforcement of data protection laws challenging. In addition to legal, ethical, institutional and organisational checks and balances surrounding privacy rights, technological solutions to mitigate privacy issues caused by large-scale use of personal data are multiple, and rapidly developing. This chapter provides a selection of the many technologies aimed at protecting privacy while upholding the benefits of big data analytics. We hope the chapter serves policymakers, technology developers and other relevant audiences interested in Privacy-Preserving Technologies.

A note: Many solutions deal with mitigating risks of personal data breaches as a result of big data analytics. However, many of these solutions are equally applicable

to the case of sharing non-personal data between parties.¹ As such, there is a difference between “privacy preservation” when talking about personal data, and “confidentiality preservation” when dealing with non-personal yet confidential data, although the techniques for the two can be the same. For the sake of simplicity, we will refer to solutions as “Privacy-Preserving Technologies”, irrespective of whether they are applied to personal or non-personal data.

1.2 Context

Recent news about data leaks,² (the lack of) control over content and political influence of social networks has provided an increasing awareness of how social media platforms (mis)use personal data, which in turn has had an effect on the level of trust users have in such platforms and digital services (Newman et al. 2017). Many social media platforms get their (economic) value from capturing visitors’ behaviour either directly (via services offered) or indirectly (by tracking users’ online behaviour). With the migration from laptop- or PC-based browsing via web browsers to consuming media on mobile devices and via dedicated apps, it has become possible to collect far more types of data surrounding this behaviour in a far more targeted manner, even in near real time (Patent No. 9,720,569 2017). Combining places where people go digitally with where they are physically offers many possibilities, but also brings about many new privacy risks. Although location data is explicitly categorised as personal data in the GDPR (e.g. De Hert et al. 2018), it is not always clear what kinds of risks such data poses, specifically in combination with other types of personal or non-personal data. Debates on what personal data exactly entails (Purtova 2018) and how to apply personal data protection in the context of large-scale data analytics are even more pressing in the current landscape of data protection regulation.³ Slowly but surely, companies and governments deploying big data analytics and processing personal data are applying (and complying with) the GDPR. Beyond the growing awareness of the need to comply (the first case of a

¹Which can lead to personal data afterwards. For example, by processing data from a machine, an algorithm could identify the operator based on the consumption of electrical power of the machine. This then becomes related to personal data and could therefore be relevant to the EU General Data Protection Regulation (GDPR).

²While there are many data breaches on a corporate level that are often not mentioned or don’t make headline news, a rather (in)famous one was the data breach of a company whose secrecy and data protection were part of its core value proposition: <https://www.theguardian.com/technology/2016/feb/28/what-happened-after-ashley-madison-was-hacked>

³For an overview of the current data regulatory landscape, see a recent deliverable by the LeMo project: <https://lemo-h2020.eu/newsroom/2018/11/1/deliverable-d22-report-on-legal-issues>

GDPR fine was issued in 2018⁴), there is a wider societal need for trust in digital environments.⁵

The question of how to foster trust in digital systems is a complex and multifaceted one. Many recent research projects are engaged directly or indirectly in (re)-building trust in digital environments, via different approaches, ranging from technical to social, ethical and organisational. Going beyond mere compliance with the GDPR and other data privacy laws (Gellert *n.d.*) (sometimes dubbed “phase 1” of privacy protection in data analytics), the main aim of many current research projects that deal with Privacy-Preserving Technologies is to explore how privacy can be utilised as an asset, as a competitive advantage or as a unique selling point (sometimes dubbed “phase 2”). One of the challenges of arriving at a fully functional digital single market is to take human rights as a starting point while also offering a unique environment for innovation, to offer framework conditions that allow companies to reach this phase 2. In this chapter, we highlight projects that are developing solutions to bridge the gap between utility and privacy and that offer a positive-sum outcome, instead of a zero-sum outcome (Cavoukian 2008), when it comes to privacy and security of data. We provide recommendations for policy concerning the development of Privacy-Preserving Technologies and the uptake of such technologies by different markets or sectors. Scalability of solutions is marked as one of the main barriers in this regard, especially when cryptographic techniques are used at any point along the analysis pipeline.

2 Challenges to Security and Privacy in Big Data

What is it about big data that makes for specific data protection challenges that need addressing, and how can we address them? The challenges of protection of personal data in the context of big data analytics (BDA) mainly connect to concepts such as profiling and prediction based on large datasets of personal data. A secondary result of big data analytics is that combinations of non-personal data (according to the definition provided in the GDPR (Zarsky *n.d.*)) can still lead to the identification of persons and/or other sensitive information (Kerr 2012), rendering many current pseudonymisation and anonymisation approaches insufficient. A dilemma put forward by data science is that data protection and data-driven innovation have diverging, even opposite, premises: the former requires a clear and defined purpose for any type of processing, whereas the latter is often based on exploration of data in order to find a purpose. While this dichotomy is not new, the increasing scale, speed and

⁴<https://iapp.org/news/a/portugal-fines-hospital-400k-euros-for-gdpr-violation/>

⁵See, for instance, <https://medium.com/ipg-media-lab/how-tech-companies-are-failing-the-trust-test-1f1057de9317>

complexity of current data analytics reinforce it.⁶ We need to look for new ways to guarantee the protection of personal data while retaining the potential benefits of big data analytics. The BDVA subgroup on Data Protection and Pseudonymisation Mechanisms summarised current challenges in the most recent BDVA Strategic Research and Innovation Agenda (SRIA) (Zillner et al. 2017), including:

- A general, easy-to-use and enforceable data protection approach suitable for large-scale commercial processing⁷
- Maintaining robust data privacy with utility guarantees, also implying the need for state-of-the-art data analytics to cope with encrypted or anonymised data^{8,9}
- Risk-based approaches calibrating data controllers' obligations regarding privacy and personal data protection¹⁰
- Combining different techniques for end-to-end data protection (Mann et al. 2018; Stojmenovic et al. 2016)

The last point has also been observed by the E-SIDES project, who have investigated a wide range of technologies for privacy preservation in big data: "In practice, the technologies need to be combined to be effective and there is no single most important class of technologies".¹¹

Another challenge when designing privacy solutions for big data is the number of data sources, which can result in different settings where stakeholders can have varying degrees of access to the processed data. In the case of a single data owner, the data owner may encrypt their data with their own keying material and may apply data analytics on the encrypted data either locally or by offloading to a third-party platform. On the other hand, nowadays data is being collected by a vast range of

⁶See E-SIDES Deliverable D4.1, section 3.2. See also the ENISA report on privacy in the era of big data (<https://www.enisa.europa.eu/publications/big-data-protection>), in which the novelty is described as follows: "Therefore, the new thing in big data is not the analytics itself or the processing of personal data. It is rather the new, overwhelming and increasing possibilities of the technology in applying advanced types of analyses to huge amounts of continuously produced data of diverse nature and from diverse sources. The data protection principles are the same. But the privacy challenges follow the scale of big data and grow together with the technological capabilities of the analytics" (p. 22).

⁷For an elaborate overview of different types of measures, both technical and non-technical, see E-SIDES project Deliverable D4.1, section 4 and D3.2, section 4.4: <https://e-sides.eu/assets/media/e-sides-d4.1-ver.-1.0-1540563562.pdf>

⁸This is one of the goals of the MOSAICrOWN project, a recently started H2020 project which aims to enable data sharing and collaborative analytics in multi-owner scenarios in a privacy-preserving way, ensuring proper protection of private/sensitive/confidential information. <https://mosaicrown.eu>

⁹See e-sides Deliverable 3.2, in which a Privacy-Preserving Technologies uptake gap analysis is provided. <https://e-sides.eu/resources/deliverable-d32-assessment-of-existing-technologies>

¹⁰A risk-based tool featuring a didactic interface to carry out Data Protection Impact Assessment according to GDPR is available from the French data protection authority CNIL at: <https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment>

¹¹See E-SIDES Deliverable D3.2, conclusions. <https://e-sides.eu/resources/deliverable-d32-assessment-of-existing-technologies>

applications and services, by different kinds of organisations. This data is often subject to deep analysis in order to infer valuable information for these organisations. Nevertheless, restrictions on data access and sharing (such as using traditional encryption techniques) can render data analytics less effective, in the sense that without access to high volumes of data, applications that rely on analytics cannot maintain a good level of accuracy of their analytical models.

The ability to train an accurate model depends on the diversity of training data. With more diverse data collected from different sources, analytical models can be increasingly accurate. However, recent privacy-related regulations or business interests inhibit data producers from sharing (sensitive) data with third parties. As a consequence, organisations are not benefiting from employing collaborative large-scale analytics and from deriving more accurate global analytical models. Privacy-preserving data analytics should consider the case of data coming from multiple sources while enabling collaborative analytics without compromising the privacy of the different data subjects involved.¹²

In this regard, two main approaches can be identified. The first one aims at providing means to protect the data, establishing trust among partners (e.g. possibly by encrypting the data or adding a perturbation under Differential Privacy principles), such that data can be outsourced and processed elsewhere, even by third parties. This approach requires a very strong level of protection, since the variety of manipulations/attacks is potentially very large. Such strong protection also imposes strong restrictions: limited types of operations on the data (possibly enforced by a usage control policy), presence of distortions that may bias the results, very high computational requirements and loss of control on the ultimate data usage. A second approach relies on the deployment of a controlled processing environment where the participants are expected, or forced, to operate under specific predetermined rules and protocols. In this scenario, the data does not leave the owner facilities, and the process of training relies on secure operations on the data following pre-specified protocols. Instances of this approach are the environments known as Industrial Data Platforms (IDP) and Personal Data Platforms (PDP). This approach has been adopted, for instance, in the Musketeer project,¹³ as described in the next section. Several techniques of pseudonymisation and anonymisation have also been utilised in the Transforming Transport project in the context of an e-commerce pilot, the urban pilot in the city of Tampere (Finland) and several airport pilots.¹⁴ Finally, one may also allow an authorised third party to make analytical queries over the collected data.

¹²This is the main goal of the Musketeer project, an H2020 project that has recently started, which aims at developing an Industrial Data Platform (IDP) facilitating the combination of information from multiple sources without actually exchanging raw data (thereby protecting privacy/confidentiality) such that, eventually, better machine learning models are obtained.

¹³Machine Learning to Augment Shared Knowledge in Federated Privacy-Preserving Scenarios. EU H2020 Research and Innovation Action – grant No. 824988. <http://musketeer.eu>

¹⁴See Transforming Transport newsletters here: <https://transformingtransport.eu/downloads/newsletters>

In short, the role of Privacy-Preserving Technologies is to establish trust in a digital world, in a digital way. Although some of the above-mentioned challenges also require non-technical solutions (organisational measures, ethical guidelines on data analytics and AI,¹⁵ increased education, etc.), in the following we focus mostly on the technical solutions in the making.

3 Current Trends and Solutions in Privacy-Preserving Technologies

Different activities in Europe on data protection, such as works on privacy standards, privacy engineering and awareness-raising events, have been developed over recent decades.¹⁶ However, while the field of privacy engineering is ever-evolving in research labs and universities, for the translation into applications and services their maturity level (sometimes also referred to as Technology-Readiness Level – TRL) is important. We need to better understand the current maturity levels and types of solutions available for a specific challenge or issue (sometimes referred to as Best Available Techniques), but also an overview in general about the available technological solutions. Companies, governments or other institutions might require different levels of maturity for a particular Privacy-Preserving Technology, depending on what kind of big data processes they are involved in. ENISA, the EU Agency for Cybersecurity, developed a portal¹⁷ that provides an assessment methodology for determining the readiness of these solutions for certain problems or challenges.¹⁸ For the classification of Privacy-Preserving Technologies, a first point of departure can be found in Jaap-Henk Hoepman's Blue Book on privacy-by-design strategies (Hoepman 2020). Here, an overview is provided in terms of how and where different privacy-by-design strategies can be applied. He distinguishes the following strategies, divided into data-related and process-related tasks around privacy protection (Gürses et al. 2006) (Table 1):

There are some parts of this structure that might overlap when it comes to Privacy-Preserving Technologies, especially if the notion of Privacy-Preserving Technologies is taken broadly, to include any technology that can aid in the protection of privacy or support Privacy-Preserving Data Processing activities.

¹⁵See, for instance, <https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>

¹⁶See https://edps.europa.eu/data-protection/ipen-internet-privacy-engineering-network_en and https://ipen.trialog.com/wiki/Wiki_for_Privacy_Standards

¹⁷<https://www.enisa.europa.eu/events/personal-data-security/pets-maturity>

¹⁸Sometimes also referred to as “best available technique”, or BAT. The EDPS (European Data Protection Supervisor) describes BATs for data protection as follows: “the most effective and advanced stage in the development of activities and their methods of operation, which indicates the practical suitability of particular techniques for providing the basis for complying with the EU data protection framework. They are designed to prevent or mitigate risks to privacy, personal data and security” (see EDPS opinion, p. 10).

Table 1 Privacy strategies according to Hoepman

Data-related tasks	
Minimise	Limit as much as possible the processing of personal data.
Separate	Separate the processing of personal data as much as possible from the data itself.
Abstract	Limit as much as possible the detail in which personal data is processed.
Hide	Protect personal data, or make it unlinkable or unobservable. Make sure it does not become public or known.
Process-related tasks	
Inform	Inform data subjects about the processing of their personal data in a timely and adequate manner.
Control	Provide data subjects adequate control over the processing of their personal data.
Enforce	Commit to processing personal data in a privacy-friendly way, and enforce this adequately.
Demonstrate	Demonstrate that you are processing personal data in a privacy-friendly way.

Privacy-Enhancing Technologies, which precede the use of Privacy-Preserving Technologies as a term, are somewhat different: Privacy-Enhancing Technologies are aimed at improving privacy in existing systems, whereas Privacy-Preserving Technologies are mainly aimed at the design of novel systems and technologies in which privacy is guaranteed. Therefore, Privacy-Preserving Technologies adhere more strongly to the principle of “privacy-by-design”.¹⁹ When looking at some of the organisational aspects, we see that developments in big data and AI have also opened new avenues for pushing forward new modes of automated compliance, for instance via sticky policies and other types of scalable and policy-aware privacy protection.^{20,21,22.}

Other attempts have recently been made to create meaningful overviews or typologies of Privacy-Preserving Technologies, mainly with the goal to create clarity for the industry itself (e.g. via ISO standards) and/or to aid policymakers and SMEs.²³ Approaches are data-centred (“What is the data and where is it?”), actor-centred (“Whose data is it, and/or who or what is doing something with the data?”) or risk-based²⁴ (“What are the likelihood and impact of a data breach?”). The ISO 20889 standard, which strictly limits²⁵ itself to tabular datasets and the

¹⁹We thank Freek Bomhof (TNO) for this point.

²⁰This is one of the main aims of the SPECIAL project.

²¹The BOOST project is developing a European Industrial Data Space based on the IDSA framework, which promotes trust and sovereignty based on certification and usage control policies attached to datasets: <https://boost40.eu/>

²²The RestAssured project uses sticky policies to capture user requirements on data protection, which are then enforced using run-time data protection mechanisms. More details can be found at <https://restassuredh2020.eu/>

²³See, for instance, the E-SIDES project and the recently started SMOOTH platform project.

²⁴See E-SIDES D3.2, page 10.

²⁵See ISO standard 20889, introduction (p. VI).

de-identification of personally identifiable information (PII), distinguishes, on the one hand, privacy-preserving techniques such as statistical and cryptographic tools and anonymisation, pseudonymisation, generalisation, suppression and randomisation techniques, and, on the other hand, privacy-preserving models, such as differential privacy, k-anonymity and linear sensitivity. The standard also mentions synthetic data as a technique for de-identification.²⁶ In many such classifications, there are obvious overlaps, yet we can see some recurring patterns, for example in terms of when in the data value chain certain harms or risks can occur.²⁷ Such classifications aim to somehow prioritise and map technological and non-technological solutions. Recently, the E-SIDES project has proposed the following classification of solutions to data protection risks that stem from big data analytics: anonymisation, sanitisation, encryption, multi-party computation, access control, policy enforcement, accountability, data provenance, transparency, access/portability and user control.²⁸ When looking at technical solutions, they are aimed at preserving privacy at the source, during the processing of data or at the outcome of data analysis, or they are necessary at each step in the data value chain (Heurix et al. 2015).

Acknowledging both the needs and the challenges in making such solutions more accessible and implementable (Hoepman et al. 2016), we want to show how some current EU projects are contributing to both the state of the art and to the accessibility of their solutions. A number of research projects in the Horizon 2020 funding programme are working on technical measures that address a variety of data protection challenges. Among others, they work on the use of blockchain for patient data, homomorphic encryption, multi-party computation, privacy-preserving data mining (PPDM²⁹), and non-technical measures and approaches such as ethical guidelines, and the development of Data Privacy Vocabularies and Controls Community Group (see W3C working group DPVCG).³⁰ Moreover, they explore ways of making use of data that are not known to the data provider before sharing them, based on usage policies and clearing house concepts.³¹ Table 2 gives an overview of the types of challenges recognised by the BDV PPP projects and the BDVA Strategic Research and Innovation Agenda (SRIA), and the (technological) solutions connected to these challenges.

The following overview provides an insight into current trends and developments in Privacy-Preserving Technologies that have been or are being explored by recent

²⁶See also <https://project-hobbit.eu/mimicking-algorithms/#transport>

²⁷Although the assumption that data processing activities take place in a sequential way is contestable.

²⁸E-SIDES D3.2, page 21.

²⁹See, for example, <https://web.stanford.edu/group/mmds/slides/mcsherry-mmds.pdf>

³⁰<https://www.w3.org/community/dpvcg/>

³¹See IDSA Reference Architecture Model: <https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf>

Table 2 Challenges identified by BDVA members

Type of challenge	Solutions
Contradiction between big data innovation and data protection	Linked data, sticky policies
Secure and trusted personal data sharing	(Secure) multi-party computation, self-sovereign identity management, data governance
Processing sensitive (health) data	Blockchain, multi-party computation
(Limits of) anonymisation and pseudonymisation	Homomorphic encryption, differential privacy, data wrapping
Dealing with multiple data sources and untrusted parties	Multi-party computation, data sanitisation and wrapping techniques
Maintaining robust data privacy with utility guarantees	Multi-party computation, federated learning approaches, distributed ledger technologies
Risk-based approaches calibrating data controllers' obligations	Automated compliance, risk assessment tools
Combining different techniques for end-to-end data protection	Integration of approaches, toolboxes, overviews and repositories of Privacy-Preserving Technologies (such as ENISA's self-assessment kit)

research projects and that we see as being key for the future research and development of Privacy-Preserving Technologies.

3.1 *Trend 1: User-Centred Data Protection*

For many years, the main ideas of what data is or who it belongs to and who controls access to it have been predominantly aimed at service providers, data stores and sector-specific data users (scientific and/or commercial). The end user and/or data subject was (and predominantly still is) taken on board merely by ticking a consent box on a screen, or is denied a service if not complying or if personal data is not provided, via, for instance, forcing users to make an account or to accept platform lock-in conditions. An increasing data-scandal-fed dissatisfaction can be witnessed in society, which in turn also demands different models or paradigms on how we think about and deal with personal data. Technologically, this means that data architectures and logics need to be overhauled. Some of the trends we see revolve around (end) user control. The notion of control in itself is a highly contested concept when it comes to data protection and ownership, as it remains unclear what “exercising control” over one’s personal data should actually entail (Schaub et al. 2017). Rather, novel approaches “flip” the logic of data sharing and access, for instance by actualising dynamic consent and by introducing self-sovereign identity schemes based on distributed ledger technologies.³² Moreover, there are

³²See, for instance, International Data Spaces Association. <https://www.internationaldataspaces.org/publications/infographic/>

developments to make digital environments more secure by making compliance with digital regulation more transparent and clear. Within the Transforming Transport³³ project, the pilot studies suggested that extra training or assistive tools (i.e. an electronic platform or digital service) should be utilised. These tools and training material will be characterised by a user-friendly natural language on the provided definitions on questions raised. Moreover, the explanations to be offered to everyday users should be easily digestible in comparison to the current legalistic and lengthy documents offered by national authorities, which still do not cover cases extensively. For example, the SPECIAL project aims to help data controllers and data subjects alike with new technical means to remain on top of data protection obligations and rights. The intent is to preserve informational self-determination by data subjects (i.e. the capacity of an individual to decide how their data is used), while at the same time unleashing the full potential of big data in terms of both commercial and societal innovation. In the SPECIAL project, the solution lies in the development of technologies that allow the data controller and the data subject to interact in new ways, and technologies³⁴ that mediate consent between them in a non-intrusive manner. MOSAICrOWN is another H2020 project that aims at a user-centred approach for data protection. This project aims to achieve its goal of empowering data owners with control over their data in multi-owner scenarios, such as data markets, by providing both a data governance framework, able to capture and combine the protection requirements that can possibly be specified by multiple parties who have a say over the data, and effective and efficient protection techniques that can be integrated in current technologies and that enforce protection while enabling efficient and scalable data sharing and processing. Another running H2020 project, MyHealthMyData (MHMD), aims at fundamentally changing the way sensitive data is shared. MHMD is poised to be the first open biomedical information network, centred on the connection between organisations and individuals, encouraging hospitals to make anonymised data available for open research, while prompting citizens to become the ultimate owners and controllers of their health data. MHMD is intended to become a true information marketplace, based on new mechanisms of trust and direct, value-based relationships between citizens, hospitals, research centres and businesses. The main challenge is to open up data silos in healthcare that are sealed at the moment for various reasons, one of them being that the protection of privacy of individual patients cannot be guaranteed otherwise. As stated by the research team, the “MHMD project aims at fundamentally changing this paradigm by improving the way sensitive data are shared through a decentralised data and transaction management platform based on blockchain technologies”.³⁵ Building on the underlying principle of smart contracts, solutions are being developed that can connect different stakeholders of medical data,

³³See <https://transformingtransport.eu>

³⁴https://www.specialprivacy.eu/images/documents/SPECIAL_D1.7_M17_V1.0.pdf, p 36.

³⁵http://www.myhealthmydata.eu/wp-content/themes/Parallax-One/deliverables/D1.1_Initial-List-of-Main-Requirements.pdf, p 6.

allowing for control and trust via a private ledger.³⁶ The idea behind using blockchain is that it allows for a shared and distributed trust model while also allowing for more dynamic consent and control for end users about how and for which (research) purposes their data can be used.³⁷ By interacting intensively with the different stakeholders within the medical domain, the MHMD project has developed an extensive list of design requirements for the different stakeholders (patients, hospitals, research institutes and businesses) to which their solutions should (in part) adhere.³⁸ While patient data is particular, both in sensitivity and in the fact that it also falls under specific healthcare regulations, some of these developments also allow for more generic solutions to alleviate user control. The PAPAYA project is developing a specific component to alleviate user control, named Privacy Engine (PE).³⁹ The PE provides the data subject with mechanisms to manage their privacy preferences and to exercise their rights derivative from the GDPR (e.g. the right to erasure of their personal data). In particular, the Privacy Preferences Manager (PPM) allows the data subject to capture their privacy preferences on the collection and use of their personal data and/or special categories of personal data for processing in privacy-preserving big data analytics tasks. The Data Subject Rights Manager (DSRM) provides to the data subjects the mechanism for exercising their rights derivative from the current legislation (e.g. GDPR, Article 17, Right to erasure or “right to be forgotten”). In order to do so, the PE allows data controllers to choose how to react to data subject events (email, publisher/subscriber pattern, protection orchestrator). For data subjects, the PE provides a user-centric Graphical User Interface (GUI) to easily exercise their rights. A related technical challenge is how to furnish back-end Privacy-Preserving Technologies with usable and understandable user interfaces. One underlying challenge is to define and design meaningful human control and to find a balance between cognitive load and opportunity costs. This challenge is a two-way street: on the one hand, there is a boundary to be sought in terms of explaining data complexities to wider audiences, and on the other hand there is a “duty of care” in digital services, meaning that technology development should aid human interaction with digital systems, not (unnecessarily) complicate them. Hence, the avenue of automating data regulation (Bayamlioğlu and Leenes 2018) is of relevance here.

³⁶[http://www.myhealthmydata.eu/wp-content/themes/Parallax-One/deliverables/D6.8_Blockchainanalytics\(1\).pdf](http://www.myhealthmydata.eu/wp-content/themes/Parallax-One/deliverables/D6.8_Blockchainanalytics(1).pdf)

³⁷http://www.myhealthmydata.eu/wp-content/uploads/2018/06/ERRINICTWGBLOCKCHAIN_130618_MHMD_AR_FINAL.pdf

³⁸http://www.myhealthmydata.eu/wp-content/themes/Parallax-One/deliverables/D1.1_Initial-List-of-Main-Requirements.pdf from page 15 onwards.

³⁹https://www.papaya-project.eu/sites/default/files/papaya/public/content-files/deliverables/PAPAYA_D4_1_Platform_Design_and_Development.pdf, p 113 and onwards.

3.2 Trend 2: Automated Compliance and Tools for Transparency

Some legal scholars argue that the need to automate forms of regulation in a digital world is inevitable (Hildebrandt 2015), whereas others have argued that hardcoding laws is a dangerous route, because laws are inherently argumentative, and change along with society's ideas of what is right, or just (Koops and Leenes 2013). While the debate about the limits and levels of techno-regulation is ongoing, several projects actively work on solutions to harmonise and improve certain forms of automated compliance. When working with personal data, or sharing personal data, different steps in the data value chain (Curry 2016) can be automated with respect to preserving privacy. Data sharing in itself should not be interpreted as unprotected raw data exchange, since there are many steps to be taken in preparing the exchange (such as privacy protection). Following this premise, there are three main possible scenarios for sharing of personal data. The first one proposes to share data to be processed elsewhere, possibly protected using a Privacy-Preserving Technology (e.g. outsourced encrypted data to be processed in a cloud computing facility under Fully Homomorphic Encryption (FHE) principles). The second scenario proposes an information exchange, without ever communicating any raw data, to be gathered in a central position to build improved models (e.g. interaction among different data owners under Secure Multi-party Computations to jointly derive an improved model/analysis that could benefit them all). The third scenario relies on data description exchange at first. Then, when two stakeholders agree on exchanging data upon the description of a dataset (available in a broker), the exchange occurs directly between the two parties in accordance with the usage control policy (e.g. applying restrictions and pre-processing) attached to the dataset as presented by the International Data Spaces Association (IDSA) framework, for instance.⁴⁰ Furthermore, it is important to be aware of the trade-offs among data utility, privacy risk, algorithmic complexity and interaction level. The Best Available Technique concept cannot be defined in absolute terms, but rather in relation to a particular task and user context.

One of the challenges in automating compliance is the harmonisation of privacy terminology, both in the back end and the front end of information systems. The SPECIAL project focuses on sticky policies, developing a standard semantic layer for privacy terminology in big data, and dynamic user consent as a solution domain for dealing with the intrinsic challenge of obtaining consent from end users when dealing with big data. Basing their project on former work on architectures for big, open and linked data, they propose a specific architecture. Their approach to user control is via managing lifted semantic metadata⁴¹: "SPECIAL tries to leverage existing policy information into the data flow, thus recording environmental

⁴⁰<https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf>

⁴¹See https://www.specialprivacy.eu/images/documents/SPECIAL_D21_M12_V10.pdf

information at collection time with the information. This is more constraint than the semantic lifting of arbitrary data in the data lake. SPECIAL will therefore not only develop the semantic lifting further, but also develop ways how to register, augment and secure semantically lifted data”.⁴² The project is investigating the use of blockchain as a ledger to check and verify data(sets) on their compliance to several regulations and data policies. As they state: “The SPECIAL transparency and compliance framework needs to be realised in the form of a scalable architecture, which is capable of providing transparency beyond company boundaries. In this context, it would be possible to leverage existing blockchain platforms [...] each have their own strengths and weaknesses”.⁴³ Building on existing platforms and solutions, they contribute by looking into automation and formalisation of policy and the coupling of different formal policies semantically. The challenge is, on the one hand, to make end-user rights (rights of companies or individuals) manageable in the context of big data, and, on the other hand, to explore the limits of policy formalisation and machine-readable policies (technically, legally and semantically). Other solutions for automated compliance can be found in, for instance, the PAPAYA project mentioned earlier, in which a privacy engine transforms high-level descriptions to computer-oriented policies, allowing their enforcement in subsequent processes to only permit the processing of the data already granted by the data subject (e.g. filtering and excluding certain personal attributes). BOOST is another example of a project developing automated compliance (once stakeholders are certified) and transparency tools (dynamic management of participant attributes, clearing house) based on the IDSA framework. BOOST aims to construct a European Industrial Data Space (EIDS), enabling companies to use and exchange more industrial data to foster the introduction of big data in the factory.⁴⁴ The EIDS relies on secured and monitored connectors deployed on every participant’s facilities where data is hosted and made available for exchange.

All such solutions aim to translate and automate legal text into computer language, and then back again to some form of human control or intervention to tweak parameters in the computer language translation of legal requirements of compliance. This is a highly complex task, and, as we have seen with the cookie-law example (Leenes and Kosta 2015), not always easily implemented or well received. Yet we need to keep pushing such efforts in order to better understand the interaction between big data utility, human experience and interpretation of what personal data and privacy mean, and current and future privacy regulation.⁴⁵

⁴²https://www.specialprivacy.eu/images/documents/SPECIAL_D3.1_M6_V10.pdf, p. 12.

⁴³See https://www.specialprivacy.eu/images/documents/SPECIAL_D2.4_M14_V10.pdf, p. 8.

⁴⁴https://boost40.eu/wp-content/uploads/2018/02/boost_leaflet.pdf

⁴⁵See also the DECODE project: <https://decodeproject.eu/>

3.3 Trend 3: Learning with Big Data in a Privacy-Friendly and Confidential Way

Several projects are working on ways to cooperate without actually sharing data. Projects such as Bigmedilytics, SODA (Scalable Oblivious Data Analytics) and Musketeer are developing and/or applying approaches to data analytics that fall under the header of (secure) Multi-party Computation. Although multi-party computation is not one technology, but rather a toolbox of different technologies, the main idea of multi-party computation is to share analytics or outcomes of analytics rather than to share data. This can be achieved by developing trust mechanisms based on encryption or data transformation to create a shared computational space that acts as a trusted third party. Where formerly such a third party needed to be some form of a legal entity, now this third party can be a computational, transformed space. The advantage of such a space is that only aggregated data or locally computed analyses are shared; this makes it possible to work together with trusted and less trusted parties without sharing one's data. There are downsides as well at the moment: multi-party computation does not work well for all data manipulations and it negatively affects performance.

One of the projects working on multi-party computation is PAPAYA. The main aim of the PAPAYA project is to make use of advanced cryptographic tools such as homomorphic encryption, secure two-party computation, differential privacy and functional encryption, to design and develop three main classes of big data analytics operations. The first class is dubbed privacy-preserving neural networks, in which PAPAYA makes use of two-party computation and homomorphic encryption to enable a third-party server to perform neural network-based classification over encrypted data. The underlying neural network model is customised in order to support the actual cryptographic tools: the number of neurons is optimised and the underlying operations mainly consist of linear operations and some minor comparison. Although the developed model differs from the original one, it is ready to support cryptographic tools in order to ensure data privacy while still keeping a good accuracy level. Furthermore, the project also focuses on the training phase and investigates a collaborative neural network training solution based on differential privacy. A second proposed solution is privacy-preserving clustering: PAPAYA investigates algorithms that consist of regrouping data items in k clusters without disclosing the content of the data. The project specifically focuses on trajectory clustering algorithms. Partially homomorphic encryption and secure two-party computation are the main building blocks to develop privacy-preserving variants of such clustering algorithms. The third area is privacy-preserving basic statistics. The project is developing privacy-preserving counting modules which make use of functional encryption to enable a server to perform the counting without discovering the actual numbers. The result can only be decrypted by authorised parties.

The SODA (Scalable Oblivious Data Analytics) project⁴⁶ aims to enable practical privacy-preserving analytics of information from multiple data assets, also making

⁴⁶<https://www.soda-project.eu/>

use of multi-party computation techniques. The main problems addressed include privacy protection of personal data and protection of confidentiality for sensitive business data in analytics applications. This means that data does not need to be shared, only made available for encrypted processing. So far, SODA has been working on pushing forward the field of multi-party computation. In particular, they work on enabling practical privacy-preserving data analytics by developing core multi-party computation protocols and multi-party computation-enabled machine learning algorithms. The project also considers the combination of multi-party computation and Differential Privacy to enable the protection of (intermediate) results of multi-party computation. The aforementioned innovations are incorporated in multi-party computation frameworks and proof of concepts. They address underlying challenges such as compliance with privacy legislation (GDPR) requirements, willingness of individuals and organisations to share data, and reputation and liability risk appetite of organisations. SODA analyses user and legal aspects of big data analytics, using multi-party computation as a technical security measure under the GDPR, whereby encrypted data is to be considered de-identified data.

The Musketeer project aims at developing an open-source Industrial Data Platform (IDP) instantiated in an interoperable, highly scalable, standardised and extendable architecture, efficient enough to be deployed in real use cases. It incorporates an initial set of analytical (machine learning) techniques for privacy-preserving distributed model learning such that the usage of every user's data fully complies with the current legislation (such as the GDPR) or other industrial or legal limitations of use. Musketeer does not rely on a single technology; rather, different Privacy Operation Modes will be implemented. Machine learning algorithms will be developed on the basis of different Privacy Operation Modes. These Privacy Operation Modes have been designed to remove some privacy barriers and each one describes a potential scenario with different privacy preservation demands and with different computational, communication, storage and accountability features. To develop the Privacy Operation Modes, a wide variety of standard Privacy-Preserving Technologies will be used, such as federated machine learning, homomorphic encryption, differential privacy or multi-party computation, also aiming at developing new ones or incorporating others from third parties in the future. Upon definition of a given analytic task, the platform will help to identify the Best Available Technique to be selected among the Privacy Operation Modes, thereby facilitating the usage of the platform especially for non-expert users and SMEs. Security and robustness against attacks will be ensured, not only with respect to threats external to the data platform, but also internal threats, through early detection and diminishment of the potential misbehaviours of IDP members. To further foster the development of a user data economy based on the data value (ultimately enabling data- and AI-driven digital transformation in Europe), the project will explore reward models capable of estimating the contribution of a user's data to the improvement of a given task, such that a fair monetisation scheme becomes possible.

Having provided an overview of cutting-edge trends and directions of the field of Privacy-Preserving Technologies, we will now mention some key challenges regarding the development, scaling and uptake of solutions developed by these projects.

3.4 *Future Direction for Policy and Technology Development: Implementing the Old & Developing the New*

Looking at the origins of Privacy-Preserving Technologies, they are technologies to re-establish trust that was broken by technology in the first place. There are inherent risks in technological “solutionism”, such as getting into an arms race between novel harm-inducing technologies and trying to find remedies. Also, many technological solutions for data protection themselves need personal data or some form of data processing in order to protect that same data and/or data subject. This bootstrapping problem is well known, and hence other solution domains have gained traction (such as organisational, ethical and legal measures⁴⁷). Yet here there is also an increased interaction with, and demand for, novel remedying technologies: the GDPR has placed unique demands on implementing privacy-by-design and privacy-by-default solutions, which are entirely or in part technological. In the wake of AI, we also see the field of explainable AI (XAI⁴⁸) turning to technical measures to explain or make apparent automated decision-making. In short, we need technical solutions to fix what is broken in present-day information societies, and/or to prevent novel harm. In the wake of recent H2020 calls, the timing seems adequate to take stock of what is already available and what is being developed for the near future. Moreover, the work needed in the research, development, implementation and maintenance of Privacy-Preserving Technologies reflects a growing market and an increased number of stakeholders working in the field of privacy and data protection.

The GDPR requires national data protection authorities from every EU member state to consult and agree as a group on cases for using specific datasets required by big data technologies. Several pilots that are running in the Transforming Transport project came across fragmented policies regarding GDPR across Europe, and thus they experienced an imbalance between the different interpretations of (the protection of) privacy rights. It is currently difficult for the industry to define personal data and the appropriate levels of privacy protection needed in a sample dataset. Such pilots provide the opportunity to give feedback to policymakers and influence the next version of the GDPR and other data regulations. Uncertainty about the interpretation of the GDPR also affects service operators in acquiring data for accurate situational awareness, for example. For instance, vehicle fleet operators may be

⁴⁷See also E-Sides deliverable 3.2: <https://e-sides.eu/resources/deliverable-d32-assessment-of-existing-technologies>

⁴⁸See, for instance, <https://www.darpa.mil/program/explainable-artificial-intelligence>

reluctant to provide data on their fleet to service operators since they are not certain which of the data is personal data (e.g. truck movements include personal data when the driver takes a break).⁴⁹ Due to such uncertainties, many potentially valuable services are not developed and data resources remain untapped.

There is an inherent paradox in privacy preservation and innovation in big data services: start-ups and SMEs need network effects, and thus more (often personal) data, in order to grow, but also have in their start-up phase the fewest means and possibilities to implement data protection mechanisms, whereas larger players tend to have the means to properly implement Privacy-Preserving Technologies, but are often against such measures (at the cost of fines that, unfortunately, do not scare them much so far). In order to make the Digital Single Market a space for human values-centric digital innovation, Privacy-Preserving Technologies need to become more widespread and easier to find, adjust and implement. Thus, we need to spend more effort in “implementing the old”. While many technological solutions developed by the projects mentioned above are state of the art, there are Privacy-Preserving Technologies that have existed for longer and that are at a much higher level of readiness.

Many projects aim to develop a proof of principle within a certain application domain or case study, taking into account the domain-specificity of the problem, also with the aim of collecting generalisable experience that will lead to solutions that can be taken up in other sectors and/or application domains as well. The challenges of uptake of existing Privacy-Preserving Technologies can be found in either a lack of expertise or a lack of matchmaking between an existing tool or technology for privacy preservation and a particular start-up or SME looking for solutions while developing a data-driven service. A recent in-depth analysis has been made by the E-SIDES project on the reasons behind such a lack of uptake, and what we can do about it.⁵⁰ They identify two strands of gaps: issues for which there is no technical solution yet, and issues for which solutions do exist but implementation and/or uptake is lagging behind.⁵¹ In addition to technical expertise, budget limitations or concerns that may prevent the implementation of Privacy-Preserving Technologies play a major role, as well as cultural differences in terms of thinking about privacy, combined with the fact that privacy outcomes are often unpredictable and context-dependent. The study of E-SIDES emphasises that the introduction of privacy-preserving solutions needs to be periodically reassessed with respect to their use and implications. Moreover, the ENISA self-assessment kit still exists and should perhaps be overhauled and promoted more strongly.⁵²

When it comes to protecting privacy and confidentiality in big data analytics without losing the ability to work with datasets that hold personal data, the group of

⁴⁹See, for example, <https://www.big-data-value.eu/transformingtransport-session-and-policy-workshop-at-the-ebdvf-2018/>

⁵⁰<https://e-sides.eu/assets/media/e-sides-d4.1-ver.-1.0-1540563562.pdf>

⁵¹See <https://e-sides.eu/resources/white-paper-privacy-preserving-technologies-are-not-widely-integrated-into-big-data-solutions-what-are-the-reasons-for-this-implementation-gap>

⁵²<https://www.enisa.europa.eu/publications/pets-controls-matrix/pets-controls-matrix-a-systematic-approach-for-assessing-online-and-mobile-privacy-tools>

technologies that falls under multi-party computation seems a fruitful contender. However, at the moment, the technology remains at the lower ends of TRL levels. As one SODA project member outlined, uptake of multi-party computation solutions in the market is slow. Many activities in the project are aimed at increasing uptake of multi-party computation solutions: “To bring results to the market we incorporate them in the open source FRESCO multi-party computation framework⁵³ and other software and we use them in our SME institute consulting business or spinoff thereof. Thirdly, we adopt them internally in our large medical technology enterprise partner, and we advocate multi-party computation potential and progress in the state of the art to target audiences in areas of data science, business, medical and academia”. The main barriers the project sees for adoption of multi-party computation solutions on a large commercial scale relate to, among others, “the relative newness of the technology (e.g. unfamiliarity, software framework availability and maturity) as well as the state of the technology that needs to develop further (e.g. performance, supported programming constructs and data types, technology usability)”. As a main message to policymakers, they state that: “Policy makers should be aware that different Privacy-Preserving Technologies are in different phases of their lifecycle.⁵⁴ Many traditional Privacy-Enhancing Technologies are relatively mature and benefit mostly from actions to support adoption whereas others (e.g. multi-party computation) would benefit most from continuing the strengthening of the technology next to activities to support demonstration of its potential and enable early adoption”.⁵⁵ This connects to the call made by ENISA to (self-)assess Privacy-Preserving and Privacy-Enhancing Technologies via a maturity model in order to develop a better overview of the different stages of development of the different technologies.

4 Recommendations for Privacy-Preserving Technologies

From the three trends mentioned above we formulate the following recommendations.

Development of Secure Data Storage Spaces

The growing use of digital services is pressing technologists to find privacy engineering solutions to alleviate the general concerns on privacy. The GDPR, among others, aims at providing legal assurances concerning the protection of personal data,

⁵³<https://github.com/aicis/fresco>

⁵⁴This point has been acknowledged by ENISA, who have developed a “Privacy-Enhancing-Technology self-assessment” toolkit in order to self-assess the market-readiness, or maturity, of a particular technical solution – see https://www.enisa.europa.eu/publications/pets-maturity-tool/at_download/fullReport

⁵⁵Based on an interview with SODA researcher Paul Koster, Senior Scientist, Digital Security, Data Science, Philips Research.

while an increasing number of frameworks, tools and applications demand personal data. On the one hand, laws and regulations for guaranteeing privacy, for protecting personal data and for ensuring usable digital identities have never been so rigorous, but on the other hand, compliance with the GDPR and other relevant data regulation remains challenging with today's threat landscape, making the risks of data breaches larger than ever. The GDPR imposes a number of onerous cybersecurity and data breach notification obligations on organisations across Europe, with strong enforcement power for data protection authorities, and this generates a frightening situation for companies when it comes to working with (big) data. Beyond engineering solutions, which already exist, another business opportunity is opening up: secure data storage environments (which may be part of personal, industrial or even hybrid data platforms). These are digital environments that are topic oriented, linked and certified by data protection authorities, offering the possibility to train algorithms that need to be trained on real data while offering guarantees of IPR protection and making sure that databases in these environments are accurate. Within experiments and testing phases, such secure environments would exempt the enterprises that need data from the responsibility to prove that they have all the necessary security measures in accordance with the legal precepts. Combined with such approaches, lessons learnt from cases and best practices should feed into the updating of current data policies according to the use cases in the different industrial sectors. This would allow Europe to move forward in making business from AI/ML taking into account Privacy-Preserving Technologies.

Continued Support for Research, Innovation and Deployment of Privacy-Preserving Technologies

As stated above, the E-SIDES project has performed an in-depth gap analysis concerning the uptake of Privacy-Preserving Technologies. One of the main challenges identified and broadly underlined by the BDV PPP stakeholders that participated in this chapter is that of scalability. The main argument here, as also posed earlier by the E-SIDES project, is that the uptake of Privacy-Preserving Technologies suffers from a bootstrapping problem: the more certain solutions are used, the better they become; but in order for companies and SMEs to start using them, they need to be good (i.e. robust, verified, standardised, known in the industry, etc.). Many types of solutions emerge from research and development communities in privacy engineering. Within privacy engineering, solutions can come from community-identified problems that emerge during the development of digital services; they can come from dedicated programmes in which solutions are pitched for known and existing problems in society; or they can originate from demands posed by regulation of a certain digital technology. Without active developer communities and without support to get solutions and ideas from these communities into the real world, many potential solutions will never come to fruition. As such, more efforts into community building and support is necessary, combined with strengthened research and innovation actions to develop solutions that address the communities' requirements. There are already many efforts to strengthen the connection between large enterprises, SMEs and R&D in privacy engineering and the

implementation of Privacy-Preserving Technologies.⁵⁶ However, this still requires significant knowledge and awareness about data processing, big data analytics and data protection issues. Already existing infrastructures such as Digital Innovation Hubs⁵⁷ and Big Data Centres of Excellence⁵⁸ could also act as knowledge transfer centres for education, implementation and expertise for Privacy-Preserving Technologies, although for now Privacy-Preserving Technologies are not their main focus. Continuous efforts should be provided to develop training material, tutorials and tool support (e.g. libraries, open-source components, testbeds) and to incorporate them into formal and non-formal education. Highlighting and following best practices of implementation of Privacy-Preserving Technologies per sector would be a good way to allow companies to learn from – and improve – Privacy-Preserving Technology uptake.

Support and Contribution to the Formation of Technical Standards for Preserving Privacy

Different applications of big data technologies lead to different types of potential harm that require different responses and technological measures. Whereas we have provided a high-level overview of privacy (and confidentiality) threats and corresponding technical solution areas, more work is needed to capture, understand and communicate which type of solution fits a particular problem. This would benefit data-driven companies, start-ups and SMEs tremendously. The work done by ISO standardisation bodies and others that tackle the challenge of classification of technologies is crucial in understanding, shaping and prioritising challenges and solutions in the field of privacy engineering. The sanitisation efforts by projects mentioned earlier also push forward the creation of a common privacy language and semantics between machine and human language. This is a necessary step for automating compliance and for preparing good data for AI.⁵⁹ We need to continue work on maturity modelling and to support an EU-driven marketplace for Privacy-Preserving Technologies. Moreover, we need to keep supporting efforts to increase the development and implementation of technological standards around Privacy-Preserving Technologies. In terms of privacy regulation, despite the complexities and difficulties regarding its implementation, the GDPR can still be seen as a major step to strengthen protection of personal data for individuals. However, there is still uncertainty about the practical implications of the GDPR, also in combination with other data-related regulation (as such, the GDPR is merely one piece in the data-regulation puzzle). If risks to Europe's technology industry and big data strategy materialise in a significant way and aspects of the GDPR weaken competition and

⁵⁶See, for instance, the SMOOTH project: <https://smoothplatform.eu/>

⁵⁷<https://ec.europa.eu/digital-single-market/en/digital-innovation-hubs>

⁵⁸<http://www.bdva.eu/node/544>

⁵⁹See <https://www.mckinsey.com/featured-insights/europe/ten-imperatives-for-europe-in-the-age-of-ai-and-automation>

competitiveness, lawmakers should not hesitate to make necessary adjustments, wherever possible.⁶⁰

Acknowledgements We thank the following contributors: Rosa Araujo (Eurecat, Spain), Alberto Crespo Garcia (Atos Spain S.A., Spain), Ariel Farkash (IBM), Antoine Garnier (IDSA, Germany), Akrivi Vivian Kiouisi (INTRASOFT Intl, Greece), Paul Koster (Philips, the Netherlands), Antonio Kung (Trialog, France), Giovanni Livraga (Università degli Studi di Milano, Italy), Roberto Díaz Morales (Tree Technology S.A., Spain), Melek Önen (EURECOM, France), Ángel Palomares (Atos Spain S.A., Spain), Angel Navia Vázquez (Univ. Carlos III de Madrid, Spain) and Andreas Metzger (paluno, Germany). Research leading to these results received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 732630 (BDVe).

References

- Bayamlioglu, E., & Leenes, R. (2018). The 'rule of law' implications of data-driven decision-making: a techno-regulatory perspective. *Law, Innovation and Technology*, 1–19. doi:<https://doi.org/10.1080/17579961.2018.1527475>
- Cavoukian, A. (2008). Staying one step ahead of the GDPR: Embed privacy and security by design. *A Peer-Reviewed Journal*, 2(2).
- Curry, E. (2016). The big data value chain: Definitions, concepts, and theoretical approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*. New York: Springer. https://doi.org/10.1007/978-3-319-21569-3_3
- De Hert, P., Papakonstantinou, V., Malignieri, G., Beslay, L., & Sanchez, I. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, 34(2), 193–203.
- Gardner, K. C., Broda, T., Jackson, T. C., Solnit, M., Sharma, M., Bubenheim, B., & Cosby, K. (2017). *Patent No. 9,720,569*. USA: U.S. Patent and Trademark Office.
- Gellert, R. (n.d.). For an explanation and discussion on the risk-based approach in the GDPR, see Gellert, R. (2018). Understanding the notion of risk in the General Data Protection Regulation. *Computer Law & Security Review*, 34(2), 279–288.
- Gürses, S., Berendt, B., & Santen, T. (2006). Multilateral security requirements analysis for preserving privacy in ubiquitous environments. *Proceedings of the UKDU Workshop*, 51–64.
- Heurix, J., Zimmermann, P., Neubauer, T., & Fenz, S. (2015). A taxonomy for privacy enhancing technologies. *Computers & Security*, 53. <https://doi.org/10.1016/j.cose.2015.05.002>
- Hildebrandt, M. (2015). *Smart environments and the end(s) of law. Novel entanglements of law and technology*.
- Hoepman, J.-H. (2020). *Privacy design strategies*.
- Hoepman, J.-H., Jensen, M., & Schiffner, S. (2016). *Readiness analysis for the adoption and evolution of privacy enhancing technologies*.
- Kerr, O. S. (2012). *The Mosaic Theory of the Fourth Amendment*. *Rev.* 311.
- Koops, B.-J., & Leenes, R. (2013). Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law. *International Review of Law, Computers & Technology*, 28, 1–13. <https://doi.org/10.1080/13600869.2013.801589>
- Leenes, R., & Kosta, E. (2015). Taming the cookie monster with Dutch law – A tale of regulatory failure. *Computer Law & Security Review*, 31. <https://doi.org/10.1016/j.clsr.2015.01.004>

⁶⁰See also the recent policy briefs by the Transforming Transport project mentioned earlier.

- Mann, Z. Á., Salant, E., Surridge, M., Ayed, D., Boyle, J., Heisel, M., et al. (2018). *Secure data processing in the cloud. Advances in service-oriented and cloud computing: Workshops of ESOCC 2017* (pp. 149–153). New York: Springer.
- Newman, N. Fletcher, R. Kalogeropoulos, A., Levy, D., & Nielsen, R. K. (2017). *Reuters Institute Digital News Report 2017*.
- Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, 10(1), 40–81.
- Schaub, F., Balebako, R., & Cranor, L. (2017). Designing effective privacy notices and controls. *IEEE Internet Computing*, 21, 70–77. <https://doi.org/10.1109/MIC.2017.75>
- Stojmenovic, I., Wen, S., Huang, X., & Luan, H. (2016). An overview of Fog computing and its security issues. *Concurrency and Computation: Practice and Experience*, 28(10), 2991–3005.
- Zarsky, T. Z. (n.d.). *See the personal data definition in the GDPR and its incompatibility as described in, for example: Zarsky, T. Z. (2016). Incompatible: The GDPR in the age of big data. Seton Hall L. Rev., 47, 995.*
- Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European big data value strategic research & innovation agenda*. Retrieved from Big Data Value Association website www.bdva.eu

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.