

Bird & Bird

Data-related legal, ethical and social issues

Privacy & Data Protection 	(Cyber)-Security 	Breach-related obligations 
Supply of Digital Content & Services 	Free Flow of Data 	Intellectual Property Rights 
Data Sharing Obligations 	Data Ownership 	Data Sharing Agreements 
Anonymisation & Pseudonymisation 	Open Data 	Liability 
Competition 	Trust, Surveillance, and Free Will 	Discrimination 
Transparency, Consent, Control and Personal Data Ownership 		

August 2019

Updated
version

Contents

EU-funded projects	1
Foreword	2
Glossary	3
General Overview	6
Privacy and Data Protection	10
Anonymisation/ pseudonymisation	17
(Cyber-)security	24
Breach-related obligations	31
Supply of digital content	36
Free flow of data	40
Liability	46
Intellectual property rights	52
Open data	59
Sharing obligations	65
Data ownership	71
Data sharing agreements	77
Competition	84
Trust, Surveillance and Free Will	88
Discrimination	94
Transparency, Consent, Control and Personal Data Ownership	99
Looking beyond	105

EU-funded projects

This publication was written in the context of the LeMO Project (www.lemo-h2020.eu). Certain chapters have also been based on the findings of the DEFEND, THREAT-ARREST and TOREADOR projects under the Horizon 2020 programme, of which Bird & Bird LLP is also a partner.



The LeMO (Leveraging Big Data to Manage Transport Operations) project aims to provide recommendations on the prerequisites of effective big data implementation in the transport sector. Transport researchers and policy makers today face various challenges including legal and ethical ones as they work to build tomorrow's transportation systems. LeMO addresses these issues by investigating the implications of the use of big data to enhance the economic sustainability and competitiveness of the European transport sector.

Grant agreement number 770038.



The DEFEND (Data Governance for Supporting GDPR) project aims to deliver a unique organisational data privacy governance platform to empower organisations from different sectors to assess and comply with the General Data Protection Regulation. The project's main focus is on building a platform driven by market needs and based on a new paradigm called Model-Driven Privacy Governance.

Grant agreement number 773701.



The THREAT-ARREST (Cyber Security Threats and Threat Actors Training - Assurance Driven Multi-Layer, end-to-end Simulation and Training) project aims to develop an advanced training platform to adequately prepare stakeholders from different sectors in defending high-risk cyber systems and organisations to counter advanced, known and new cyber-attacks. The platform will deliver security training based on a model driven approach, incorporating among others emulation and simulation.

Grant agreement number 786890.



The TOREADOR project was set up bearing in mind that many companies and organisations in Europe have become aware of the potential competitive advantage they could get by timely and accurate Big Data analytics, but lack the IT expertise and budget to fully exploit that potential. To overcome this hurdle, TOREADOR provides a model-based big data analytics-as-a-service (MBDAaaS) approach, providing models of the entire Big Data analysis process and of its artefacts.

Grant agreement number 688797.



The above projects have received funding from the European Union's Horizon 2020 research and innovation programme.

The information given in this document concerning technical, legal or professional subject matter is for guidance only and does not constitute legal or professional advice.

The content of this publication reflects only the authors' views. The European Commission and Innovation and Networks Executive Agency (INEA) are not responsible for any use that may be made of the information it contains.

Foreword

This publication presents the main legal, ethical and social issues predominantly relevant in an EU data-driven context. It is particularly pertinent in the context of big data, but the findings also apply to other disruptive technologies that heavily rely on data, such as the Internet of Things or Artificial Intelligence.

In order to provide practical examples or to contextualise our theoretical observations, several illustrations related to the transport sector are included throughout this publication. The findings however apply to many other sectors.

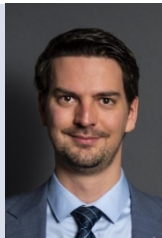
The issues related to data in the EU are constantly developing and will continue to evolve as the data economy remains at the centre of attention in the next EU legislature. Therefore, although the different Chapters of this publication are likely to require regular updates, they nonetheless aim to serve as a basis for further discussions on the 16 topics covered, with the ultimate aim of improving the general framework related to data.

We would like to thank the other authors who have contributed to this publication, as well as Marie Thiot who coordinated its content and without whom this tremendous work would have been impossible.

Julien Debussche

Senior Associate

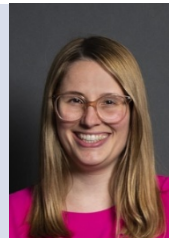
Julien.Debussche@twobirds.com



Jasmien César

Associate

Jasmien.Cesar@twobirds.com



Glossary

AI	Artificial Intelligence
AIS	Automatic Identification System
B2B	Business-to-Business
B2C	Business-to-Consumer
B2G	Business-to-Government
BMVI	Bundesministerium für Verkehr und digitale Infrastruktur
C-ITS	Cooperative Intelligent Transport System
CJEU	Court of Justice of the European Union
CMA	UK's Competition & Market Authority
Critical Infrastructure Directive	Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection
CSIRT	Computer Security Incident Response Team
Data Breach Notification Regulation	Commission Regulation (EU) No 611/2013 of 24 June 2013 on the measures applicable to the notification of personal data breaches under Directive 2002/58/EC of the European Parliament and of the Council on privacy and electronic communications
DPIA	Data Protection Impact Assessment
Digital Content Directive	Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L 136/1
DSA	Data Sharing Agreement
DSM Directive	Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market
DSP	Digital Service Provider
e-Commerce Directive	Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market
EEA	European Economic Area
ENISA	European Union Agency for Network and Information Security
e-Privacy Directive	Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector
EU	European Union

EULF	European Union Location Framework
EU Charter	EU Charter of Fundamental Rights
FOT	Field Operational Test
Free Flow Regulation or the Regulation	Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union
GDPR	General Data Protection Regulation (Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 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 [2016] OJ L 119/1)
ICO	Information Commissioner's Office
IEC	International Electrotechnical Commission
INSPIRE Directive	Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community
IoT	Internet of Things
ISO	International Standards Organisation
ITS	Intelligent Transport System
LINC	Laboratoire d'Innovation Numérique de la CNIL
MaaS	Mobility as a Service
NCA	National Competent Authority
NIS Directive	Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union
OES	Operators of Essential Services
PECS providers	Publicly Available Electronic Communication Service Providers
Platform-to-Business Regulation	Proposal for a regulation of the European Parliament and of the Council on promoting fairness and transparency for business users of online intermediation services
PSI	Public Sector Information
PSI Directive	Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information
Recast Proposal	Proposal for a directive of the European Parliament and of the Council on the re-use of public sector information
RMI	Repair and Maintenance Information
Trade Secrets Directive	Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure

Unfair Commercial Practices Directive

Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council

General Overview

This introductory Chapter is looking into the legal, ethical and social issues and opportunities surrounding big data, which were brought to the forefront by the LeMO Project (www.lemo-h2020.eu).

The publication aims to provide a summary of the findings from our research conducted in the LeMO Project concerning legal, ethical and social challenges and opportunities pertaining to big data in the transport sector, which was published in one report entitled 'Report on Legal Issues' available online at www.lemo-h2020.eu/deliverables/. The Chapter will also, where relevant, provide illustrations from the transport sector.

Key questions will be raised throughout this publication, such as "*do the privacy concepts of the GDPR fit with big data?*", "*can anonymisation techniques be applied while keeping an acceptable level of predictability and utility of big data analytics?*", "*is the current legal framework in relation to data ownership satisfactory ?*", "*what are the main areas in which competition law may have an impact on the use of big data?*", or also "*can social differences in access to technology and education or skills lead to data-driven discrimination?*".

More particularly, each Chapter will look at a specific topic pertaining to big data, namely:

- 1 privacy and data protection
- 2 anonymisation and pseudonymisation
- 3 security and cybersecurity
- 4 breach-related obligations
- 5 supply of digital content and services
- 6 the free flow of data
- 7 liability
- 8 intellectual property rights
- 9 open data
- 10 data sharing obligations
- 11 data ownership
- 12 data sharing agreements
- 13 competition
- 14 trust, surveillance and free will
- 15 discrimination
- 16 privacy, transparency, consent, control and data ownership

Below, we provide some background information useful to bear in mind while reading the upcoming Chapters.

The concept of "big data"



Although this publication does not aim to delve into the technical aspects of big data, it nonetheless emphasises, where needed, some of the particularities of big data and each of the legal, ethical and social issues mentioned above will be examined with big data analytics technologies in mind.

There is no real consensus on a definition of "big data". An oft-heard description however is that of large datasets comprising different types of data that have grown beyond the ability to be managed and analysed with traditional tools.¹ Handling such vast numbers of variable (un-)structured data in real-time requires the adoption and use of new methods and tools (e.g., processors, software, algorithms, etc.).²

One cannot discuss the notion of big data without highlighting some of the key characteristics of big data, usually expressed with a series of "V's", and in particular:

- **Volume:** refers to the vast amount of data acquired, stored, searched, shared, analysed, visualised, generated and/or managed. Big data technologies have notably enabled the storage and use of large datasets with the help of distributed systems, where parts of the data are

¹ Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3

² Commission, 'Towards a Thriving Data-Driven Economy' (Communication) COM(2014) 442 final, 4

stored in different locations, connected by networks and brought together by software.³

- **Velocity:** refers to the speed of processing, which is of the essence in a big data context. More particularly, it refers to the speed with which data is stored and analysed, as well as the speed at which new data is generated.⁴
- **Variety:** refers to the heterogeneous types of data that can be analysed, combining structured but also unstructured datasets. There are unanimous findings that most of the data being generated and analysed today is unstructured.

In addition to these three key features, several authors also refer to "**Veracity**" which relates to the ability of analysing datasets that comprise less controllable and accurate data.⁵ Accuracy is being challenged by some key features of big data. Indeed, "*big data applications typically tend to collect data from diverse sources, and without careful verification of the relevance or accuracy of the data thus collected.*"⁶ This typically poses legal issues but also ethical ones related to trust, privacy or transparency.

The "V" of "**Value**" has also been highlighted to refer to the possibility of turning data into value.⁷ While it could be argued that data, per se, has no value, processing it creates value. In other words, data that is merely collected and stored is not likely to generate any value unless it is used by some "intelligent" software algorithms, which analyse data, learn from data, and make or suggest decisions or predictions. Moreover, the value in data may also lie with the time spent by humans organising the data, creating the algorithms or training such algorithms with human-generated examples and answers. Similarly, the (personal) data provided by individuals in their day-to-day life (for instance by using social media platforms or using an itinerary application), also has value. In

fact, the European Commission explicitly recognised in a proposed Directive in 2015 concerning contracts for the supply of digital content that "*information about individuals is often and increasingly seen by market participants as having a value comparable to money.*"⁸ It further finds that "*digital content is often supplied not in exchange for a price but against counter-performance other than money i.e. by giving access to personal data or other data.*"⁹ On such basis, the Commission proposed to harmonise certain aspects of contracts for supply of digital content, taking as a base a high level of consumer protection.¹⁰

Finally, when looking into the legal, social and ethical issues related to big data, it is worth considering other disruptive technologies such as Artificial Intelligence ("**AI**") and its sub-branches, including Machine Learning, Deep Learning, or Neural Networks, which are all algorithm-based. Such algorithmic methods rely on vast amounts of data (big data) to find trends, patterns and predictions and to produce desired results.



Big data in the transport sector

In the transportation industry, vast volumes of data are generated every day, for example through sensors in passenger counting and vehicle locator systems and through ticketing and fare collection systems, to name a few.

Big data opens up new opportunities to define "intelligent" mobility and transportation solutions. By leveraging big data tools and predictive analytics, data analytics can help transportation stakeholders to make better decisions, improve operations, reduce costs, streamline processes and eventually better serve travellers and customers.¹¹

³ Bernard Marr, 'Why only one of the 5 Vs of Big Data really Matters' (*IBM Big Data & Analytics Hub*, 19 March 2015) <<http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters>> accessed 27 December 2018

⁴ James R. Kalyvas and Michael R. Overly, *Big Data: A Business and Legal Guide* (Auerbach Publications 2014) 5

⁵ Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3

⁶ European Data Protection Supervisor, 'Opinion 7/2015 Meeting the Challenges of Big Data: A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 27 December 2018

⁷ Bernard Marr, 'Why only one of the 5 Vs of Big Data really Matters' (*IBM Big Data & Analytics Hub*, 19 March 2015) <<http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters>> accessed 27 December 2018

⁸ Commission, 'Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content' COM(2015) 634 final

⁹ COM(2015) 634 final, Recital 13. See also Gianclaudio Malgieri and Bart Custers, 'Pricing Privacy – The Right to Know the Value of Your Personal Data' (2017) CLSR 289-303

¹⁰ COM(2015) 634 final, Recital 2

¹¹ Deliverable D1.1 of the LeMO Project, entitled "Understanding and mapping big data in transport sector", offers an introduction to big data in the transport sector (downloadable at: <https://lemo-h2020.eu/deliverables/>). It notably identifies untapped opportunities and challenges and describes numerous data sources. Deliverable D1.1 covers six transportation modes (i.e. air, rail, road, urban, water and multimodal) as well as two transportation sectors (passenger and freight). It further identifies several opportunities and challenges of big data in transportation, based on several subject matter expert interviews, applied cases, and a literature review. Finally, it

Policy framework



Legislators at EU and/or national levels have adopted policies in order to regulate several aspects related to data or the transport sector, but also to combat the most conspicuous and persistent ethical issues or to set social norms.

While there are no policies specific to big data, lawmakers have adopted some legislations aimed at protecting the privacy of their citizens, encouraging data sharing among private and public sector entities, and developing policies that support the digitalisation of the transport sector. Some of the key areas of recent policies in the transport sector are for instance the implementation of Intelligent Transport Systems, the increased Open Data policies, Automated Driving, and Smart Mobility.¹²

In addition to those public policies, companies – including in the transport sector – have adopted or decided to adhere to private sector policies. More particularly, the private sector has moved ahead to incorporate policies on the use of big data techniques into their own business models as process or product innovations. With digitalisation being a major trend in the transport sector, the potential applications are diverse and manifold.¹³

Despite the existence of public and private policies, the use of new technologies, such as in this case big data-driven technologies, creates new ethical and policy issues that require the adoption of new policies or the replacement of existing ones.

Assigning responsibilities

The data value cycle can be rather complex and involves numerous stakeholders. Many of these stakeholders are likely to have some kind of responsibility because, for instance, they create or generate data or algorithms, or because they use, compile, select, structure, re-format, enrich,

concludes that the combination of different means and approaches will enhance the opportunities for successful big data services in the transport sector, and presents an intensive survey of the various data sources, data producers, and service providers.

¹² Deliverable D1.2 of the LeMO project reviews current public policies implemented in the EU, its Member States and internationally, which support or restrict the (re-)use, linking of and sharing of data, in the context of big data techniques and in the transport sector (downloadable at: <https://lemo-h2020.eu/deliverables/>).

¹³ Deliverable D1.2 of the LeMO project illustrates in selected examples of transport-related private companies, the types of private sector policies that have been adopted or promoted (downloadable at: <https://lemo-h2020.eu/deliverables/>).

analyse, purchase, take a licence on, or add value to the data.

This complexity increases the difficulties in determining who could be legally, ethically or socially responsible and liable for any wrongdoing and damage, or who could be required to integrate legal, ethical and social principles in their processes. Does responsibility lie with computer system designers (e.g. software developers, software engineers, data scientists, data engineers), data providers (e.g. data brokers and marketplaces, individuals, public authorities), or even different actors?

Identifying legal issues related to big data in the transport sector

Not many legislations currently in force at EU and Member States level were made keeping disruptive technologies, such as big data, in mind. Indeed, legislative processes tend to be lengthy and often seem to end up lagging behind technological evolution. Consequently, the uptake of big data in any industry, including the transport industry, will inevitably be confronted with legal hurdles.

This publication therefore addresses how the use of (big) data and the deployment of new data-driven technologies may raise discussions in relation to the legal intricacies. While a particular emphasis is put on big data in the transport sector, the presented challenges and opportunities may also be valid for other domains.

More specifically, the research conducted in the context of the LeMO Project has enabled identifying the following key legal issues, deemed to be particularly relevant to big data, including in the transport sector: (i) privacy and data protection; (ii) (cyber-)security; (iii) breach-related obligations; (iv) anonymisation and pseudonymisation; (v) supply of digital content and services (and specifically, personal data as counter-performance); (vi) free flow of data; (vii) intellectual property in a big data environment; (viii) open data; (ix) data sharing obligations; (x) data ownership; (xi) data sharing agreements; (xii) liability; and (xiii) competition.

Identifying ethical and social issues related to big data in the transport sector

The discussions related to ethical (and social) issues in transportation are not new. Already in 1996, Professor Barbara Richardson suggested the need for the establishment of a new field of study and method of analysis that would become known as “Transportation Ethics”. Since then, what has changed in the transport sector is the huge technological development, notably in big data and artificial intelligence. Consequently, today more than ever, there is a need to look at the ethical and social implications of the use of data-driven technologies, including big data and AI, in the transportation sector.¹⁴

The second part of this publication therefore addresses how the use of (big) data and the deployment of new data-driven technologies may have a strong impact on the ethical and soci(et)al discussions. While a particular emphasis is put on big data in the transport sector, the presented challenges and opportunities may also be valid for other domains.

More specifically, the research conducted in the context of the LeMO Project has enabled identifying the following key ethical and social issues, deemed to be particularly relevant to big data, including in the transport sector: (i) trust; (ii) surveillance; (iii) privacy (including transparency, consent and control); (iv) free will; (v) personal data ownership;

(vi) data-driven social discrimination and equity; and (vii) environmental issues.

Conclusion

The next Chapters will focus on the 16 topics listed above. This, however, does not mean that other legal, ethical and social issues are not relevant. Indeed, the development of new services (such as in the transport sector) that rely on data-driven technologies raises a myriad of technical, economic, legal, ethical and social issues.



¹⁴ Rob Smith, '5 Core Principles to Keep AI Ethical' (World Economic Forum, 19 April 2018) <<https://www.weforum.org/agenda/2018/04/keep-calm-and-make-ai-ethical/>> accessed 27 December 2018



Privacy and Data Protection

In this second Chapter, we focus on some of the privacy and data protection aspects in a big data context. Where relevant, illustrations from the transport sector will be provided.

The analysis of privacy and data protection aspects in a big data context can be relatively complex from a legal perspective. Indeed, certain principles and requirements can be difficult to fit with some of the main characteristics of big data analytics, as will be demonstrated in this Chapter. In this respect, it is important to note that *“the process of aggregation implies that data is often combined from many different sources and that it is used and/or shared by many actors and for a wide range of purposes.”*¹⁵ This multitude of sources, actors and purposes cannot always be reconciled with the legal requirements related to data protection and security. Despite the intricacies of the legal analysis, it is still important to carefully examine how the legal requirements can be implemented in practice.

The legal assessment requires taking into consideration the newly adopted EU legal framework, and notably the new General Data Protection Regulation¹⁶ (hereinafter the **"GDPR"**), which became applicable on 25 May 2018, introducing a raft of changes to the existing data protection regime in the EU. While some of the data protection principles, obligations and rights pre-existed, some of them have been enhanced and others newly created by the GDPR.

In the remainder of this Chapter, we will not delve into all rights and obligations included in the GDPR. We will however examine some of the core principles and concepts put forward by the GDPR that many actors active in the field of big data analytics at European level will be confronted with, and how these may be difficult to reconcile with disruptive technologies.

Privacy and Data Protection in a Big Data Context: Challenges & Opportunities

This section dedicated to the analysis of some of the relevant challenges and opportunities related to privacy and data protection intends to show some of the intricacies that some concepts, principles and obligations may cause in relation to a disruptive technology such as big data.

The main findings, categorised by different topics, may be summarised as follows:

The concepts of "personal data" and "processing"

The GDPR applies to the "processing"¹⁷ of "personal data"¹⁸. As these definitions and the interpretation thereof are very broad, numerous obligations under the GDPR will apply in many circumstances when performing big data analytics.

¹⁵ Gloria González Fuster and Amandine Scherrer, 'Big Data and Smart Devices and Their Impact on Privacy. Study for the LIBE Committee' (European Parliament, Directorate-General for Internal Policies, Policy Department C Citizens' rights and constitutional affairs, 2015) 20 <[http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU\(2015\)536455_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf)> accessed 4 January 2019

¹⁶ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 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) [2016] OJ L 119/1

¹⁷ Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction (GDPR, art 4(2))

¹⁸ Any information relating to an identified or identifiable natural person (GDPR, art 4(1))

Moreover, in the context of big data, it cannot be excluded that the data analysis concerns "sensitive data"¹⁹ – the processing of which is restricted and prohibited in most cases – or that it will have a “transformational impact” on data. For instance, the processing of non-sensitive personal data could lead – through data mining, for instance – to the generation of data that reveals sensitive information about an individual.²⁰



Illustration in the transport sector:

The Article 29 Working Party observed in its Opinion 3/2017 on Cooperative Intelligent Transport Systems (hereinafter "C-ITS")²¹ that personal data processed through such systems may also include special categories of data as defined in Article 10 of the GDPR. More specifically, it finds that sensitive data may be collected through and broadcasted to other vehicles, such as criminal data in the form of speeding data or signal violations. It notably concludes that "*as a consequence [such C-ITS] applications should be modified to prevent collection and broadcast of any information that might fall under Article 10*".

The broad scope of application of the GDPR and the possible processing of sensitive data may require limiting certain processing activities or technical developments to tackle the stringent rules included in the GDPR.

Various actors, roles and responsibilities

In case personal data is being processed (as it is the case in data analytics), it is important to examine the concrete situation so as to determine precisely the exact role played by the different actors involved in such processing. The various concepts enshrined under EU data protection law and in particular the difference between “data controller” and “data processor”, as well as their interaction, is of paramount importance in order to determine the

responsibilities. In the same vein, such concepts are also essential in order to determine the territorial application of data protection law and the competence of the supervisory authorities.

The qualification of actors and the distinction between “controller” and “processor” can quickly become complex in a big data context. This is especially true taking into account additional data protection roles such as joint-controllership, controllers in common, and sub-processors. This is mainly due to the fact that many actors may be involved in the data value chain, the mapping of which can be rather burdensome.

Hence, additional guidance and template agreements, compliant with the strict requirements of the GDPR, are more than welcome to clarify the relationships in the big data value cycle.

Data protection principles

The GDPR outlines six data protection principles one must comply with when processing personal data²², most of which are being challenged by some key features of big data.

- The principle of "lawfulness" implies each processing of personal data should be based on a legal ground (see next section).
- The principle of “fairness and transparency” means that the controller must provide information to individuals about its processing of their data, unless the individual already has this information. The transparency principle in a big data context – where the complexity of the analytics renders the processing opaque – can become particularly challenging and implies that “*individuals must be given clear information on what data is processed, including data observed or inferred about them; better informed on how and for what purposes their information is used, including the logic used in algorithms to determine assumptions and predictions about them.*”²³

¹⁹ Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data or data concerning a person’s sex life or sexual orientation (GDPR, art 9).

²⁰ Gloria González Fuster and Amandine Scherrer, 'Big Data and Smart Devices and Their Impact on Privacy. Study for the LIBE Committee' (European Parliament, Directorate-General for Internal Policies, Policy Department C Citizens' rights and constitutional affairs, 2015) 20 <[http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU\(2015\)536455_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf)> accessed 4 January 2019

²¹ Article 29 Data Protection Working Party, 'Opinion 3/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS)' (2017) WP252, 8

²² Pursuant to Article 6 GDPR, these principles relate to: (i) lawfulness, fairness and transparency; (ii) purpose limitation; (iii) data minimisation; (iv) accuracy; (v) storage limitation; and (vi) integrity and confidentiality.

²³ European Data Protection Supervisor, 'Opinion 7/2015. Meeting the Challenges of Big Data. A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) 4

<https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 3 January 2019; See also Paul De Hert and Gianclaudio Malgieri, 'Making the Most of New Laws: Reconciling Big Data Innovation and Personal Data Protection within and beyond the GDPR' in Elise Degraeve, Cécile de Terwangne, Séverine Dusollier and Robert Queck (eds), Law,



Illustration in the transport sector:

In its guidelines on automated individual decision-making and profiling adopted on 3 October 2017, the Article 29 Working Party takes the example of car insurances to illustrate the possible issues of fair, lawful and transparent processing of personal data in the transport sector.²⁴ It indicates that some insurers offer insurance rates and services based on an individual's driving behaviour. The data collected would then be used for profiling to identify bad driving behaviour (such as fast acceleration, sudden braking, and speeding). The Article 29 Working Party concludes that in such cases, controllers must ensure that they have a lawful basis for this type of processing. They must also provide the data subject with information about the collected data, the existence of automated decision-making, the logic involved, and the significance and envisaged consequences of such processing.

- The principle of "purpose limitation" requires personal data to be collected and processed for specified, explicit and legitimate purposes. Foremost, this requires any processing of personal data to have a clearly defined purpose in order to be permitted. This may be particularly difficult in a big data context because *"at the time personal data is collected, it may still be unclear for what purpose it will later be used. However, the blunt statement that the data is collected for (any possible) big data analytics is not a sufficiently specified purpose."*²⁵
- The principle of "data minimisation" provides that personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. It is clear that the concepts of "data minimisation" and big data are at first sight antonymic. Indeed, *"the perceived opportunities in big data provide incentives to collect as much data as possible and*

*to retain this data as long as possible for yet unidentified future purposes."*²⁶

- Furthermore, personal data must be "accurate" and, where necessary, kept up-to-date. Similarly to others, the accuracy principle is being challenged by some key features of big data. Indeed, *"big data applications typically tend to collect data from diverse sources, and without careful verification of the relevance or accuracy of the data thus collected."*²⁷
- The principle of "storage limitation" requires personal data to be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. The GDPR does not specify the exact data retention periods given that these are necessarily context-specific. Big data analytics is a good illustration of the possibilities of processing personal data for a longer period and the difficulties that may arise in relation to the storage limitation principle. For instance, the principle may undermine the ability of being predictive, which is one of the opportunities rendered possible by big data analytics. Indeed, if big data analytics is allowing predictability, it is precisely because algorithms can compare current data with stored past data to determine what is going to happen in the future.

It follows from the above that the core data protection principles are, for the most part, in contradiction with some of the key features of big data analytics, and thus difficult to reconcile. Nevertheless, rethinking some processing activities but also IT developments may help complying with such principles, notably by having well-managed, up-to-date and relevant data. Ultimately, this may also improve data quality and thus contribute to the analytics.

Legal grounds to process personal data

In case the GDPR applies, any processing of personal data must be based on one of the grounds listed in Article 6(1) of the GDPR. In other words, in order for a processing activity to be lawful, from the outset and throughout the activity, it must always be based on one of the six grounds exhaustively

Norms and Freedoms in Cyberspace / Droit, Normes et Libertés dans le Cybermonde (Larcier 2018)

²⁴ Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and profiling for the purposes of regulation 2016/679' (2017) WP251, 15

²⁵ Nikolaus Forgó, Stefanie Hänold and Benjamin Schütze, 'The Principle of Purpose Limitation and Big Data' in Marcelo Corrales, Mark Fenwick and Nikolaus Forgó (eds), *New Technology, Big Data and the Law (Perspectives in Law, Business and Innovation)*, Springer 2017)

²⁶ European Data Protection Supervisor, 'Opinion 7/2015. Meeting the Challenges of Big Data. A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) 8
<https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 3 January 2019

²⁷ Ibid

listed in the GDPR.²⁸ Only four of them, however, seem to be able to be applied in a big data context.

- **Consent:** While "consent" is the first ground that can permit the processing of personal data, it can quickly become a difficult concept to comply with in light of its definition and the many conditions that must be met. More precisely, consent under the GDPR must be freely given, specific, informed and unambiguous.²⁹ Furthermore, the controller should be able to demonstrate that the data subject has given consent to the processing operation and should allow the data subject to withdraw his or her consent at any time.³⁰ The various conditions of consent are stringent and may be particularly difficult to meet. Therefore, relying on consent may prove to be unpractical or even impossible in a big data context, especially in its more complex applications.
- **Performance of or entering into a contract:** The processing ground provided under Article 6(1)(b) GDPR can be relied upon by the data controller when it needs to process personal data in order to perform a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; e.g., in case of purchase and delivery of a product or service. It follows that this ground for processing will be generally difficult to apply in a big data context, because it is unlikely that the processing of personal data for specific big data analytics purposes is "necessary" for the performance of a contract with the individual. Indeed, although big data analytics implies a complex chain of actors and multiple contracts, there is little interaction directly with the data subjects themselves.
- **Legal obligation:** Under Article 6(1)(c), the GDPR provides a legal ground in situations where "processing is necessary for compliance with a legal obligation to which the controller is subject". Generally, it is unlikely that personal data processing in a big data analytics context can be based on a "legal obligation". This being said, according to the Article 29 Working Party, such

legal ground should not automatically be set aside in a technology context.



Illustration in the transport sector:

In its Opinion on C-ITS, the Article 29 Working Party concludes that the long-term legal basis for this type of processing is the enactment of an EU-wide legal instrument. Indeed, the Article 29 Working Party considers it likely, given the projected prevalence of (semi-)autonomous cars, that the inclusion of C-ITS in vehicles will become mandatory at some point in time, comparable to the legal obligation on car manufacturers to include e-call functionalities in all new vehicles.³¹

- **Legitimate interests:** The protection of privacy and personal data is not absolute and often requires a balance of interests. Given the difficulties to rely on the abovementioned processing grounds in a big data context, the legitimate interests of an organisation may pose a good alternative.³² The GDPR includes Article 6(1)(f), which permits the processing of personal data where it is necessary "for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data." However, in an Opinion on the recent developments on the Internet of Things (hereinafter "**IoT**"), the Article 29 Working Party warns that a processing will not always be justified merely by the economic interests of the IoT stakeholder in the processing, taking into account the potential severity of interference into the privacy of the data subject.³³ A similar reasoning could be transposed to a big data context. Therefore, when trying to rely on legitimate interests, a careful balancing test between the interests of the big data stakeholder

²⁸ These are (i) the consent of the data subject; (ii) the necessity for the performance of a contract with the data subject or to take steps prior to entering into a contract; (iii) the necessity for the purposes of legitimate interests of the controller or a third party; (iv) the necessity for compliance with a legal obligation to which the controller is subject; (v) the necessity for the protection of the vital interests of a data subject or another person where the data subject is incapable of giving consent; and (vi) the necessity for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.

²⁹ GDPR, art 4(11)

³⁰ GDPR, art 7

³¹ Article 29 Data Protection Working Party, 'Opinion 3/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS)' (2017) WP252, 11

³² "Legitimate interests may provide an alternative basis for the processing, which allows for a balance between commercial and societal benefits and the rights and interests of individuals." Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 34 <<https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>> accessed 3 January 2019

³³ Article 29 Data Protection Working Party, 'Guidelines on the Recent Developments on the Internet of Things' (2014) WP223,

15

and the data subject will remain of the utmost importance.

Finding the most adequate legal ground to permit the processing of personal data in the context of big data analytics may prove difficult. Indeed, the conditions associated to the grounds exhaustively listed in the GDPR are stringent and may limit or prohibit certain processing activities. Nonetheless, thorough assessments, such as in the context of a legitimate interests assessment, are likely to enable finding the most appropriate processing ground, while at the same time having the evidence to demonstrate the reasoning that lies behind, in accordance with the accountability principle.³⁴

Core obligations under the GDPR

Some of the core obligations of the GDPR applicable to controllers (and processors) may be particularly relevant in the context of big data. This is surely the case for the requirements to conduct data protection impact assessments (hereinafter "DPIAs") and to implement privacy by design and privacy by default measures.

DPIAs are required to be conducted in certain cases only, i.e. when processing is "*likely to result in a high risk*", taking into account the nature, scope, context and purposes of the processing. While Article 35(1) GDPR clearly indicates that processing "*using new technologies*" is likely to result in a high risk, Article 35(3) and Recital 91 of the GDPR provide a non-exhaustive list of occasions when DPIAs are required. For other processing activities, the organisation should determine whether the processing activity poses a high risk to individuals. In such context, Recital 75 of the GDPR provides some relevant elements that may help determining whether a (high) risk exists. In addition to the abovementioned illustrations and elements provided by the GDPR to determine whether a DPIA may be required, Article 35(4) of the GDPR requires national supervisory authorities to establish a list of processing operations that are necessarily subject to the requirement to conduct a DPIA ("black list") whereas Article 35(5) allows national supervisory authorities to establish a list of processing activities for which no DPIA shall be required ("white list").

An analysis of the various lists and guidance published by the different authorities easily leads to the conclusion that new technologies, and in

particular big data analytics, will almost systematically require carrying out a DPIA. Indeed, some of the key characteristics of big data appear to be targeted, such as "large scale processing", "systematic monitoring", "automated decision-making with legal or similar significant effect", and "matching or combining datasets". Similarly, the use of data to analyse or predict situations, preferences or behaviours, or the systematic exchange of data between multiple actors, or the use of devices to collect data (and in particular relying on IoT) should lead to the requirement to carry out a DPIA.

Furthermore, the requirement to adopt "privacy by design" measures³⁵ entails that the controller must implement appropriate technical and organisational measures (e.g. pseudonymisation techniques) designed to implement the data protection principles (e.g. data minimisation). As for compliance with the "privacy by default" requirement³⁶, the controller must implement appropriate technical and organisational measures to ensure that, by default, only personal data necessary for each specific purpose of the processing are processed. This applies to the amount of data collected as well as to the extent of processing, period of storage and accessibility of the data. The measures adopted by the controller must guarantee that, by default, personal data are not made accessible to an indefinite number of individuals without the data subject's intervention.

These requirements to implement dedicated "by design" and "by default" measures are particularly relevant in IT environments, and thus also to big data. In practice, it requires organisations to ensure that they consider privacy and data protection issues at the design phase and throughout the lifecycle of any system, service, product or process. The requirements can therefore be far-reaching and apply to all IT systems, services, products and processes involving personal data processing, but also require looking into organisational policies, processes, business practices and/or strategies that have privacy implications, and rethinking physical design of certain products and services as well as data sharing initiatives. Moreover, organisations must take technical measures to meet individuals' expectations in order to notably delimit what data will be processed for what purpose, only process the data strictly necessary for the purpose for which they are collected, appropriately inform individuals

³⁴ GDPR, art 6(2): *The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability').*

³⁵ GDPR, art 25(1)

³⁶ GDPR, art 25(2)

and provide them with sufficient controls to exercise their rights, and implement measures to prevent personal data from being made public by default.



Illustration in the transport sector:

The past decade has seen the rise of new transportation modes such as ridesharing. Ridesharing services allow car owners to fill the empty seats in their cars with other travellers. Ridesharing services however come with certain privacy and data protection implications for the users of such services. Indeed, users wanting to rely on a ridesharing service need to share their location data with the ridesharing operators in order to determine a point where drivers and riders can meet. Aïvodji et al.³⁷ have developed a privacy-preserving approach to compute meeting points in ridesharing. Taking into account the privacy-by-design principle, they have been able to integrate existing privacy-enhancing technologies and multimodal routing algorithms to compute in a privacy-preserving manner meeting points that are interesting to both drivers and riders using ridesharing services.

Rights of individuals

The GDPR aims to protect natural persons in relation to the processing of their personal data and therefore grants several rights to such persons.³⁸ In addition to these rights, the GDPR further provides for strict procedures to respond to any data subject request in exercise of their rights, notably regulating issues with respect to the timing and format of the response, or the fees that may be requested. It also regulates the right for individuals to lodge a complaint with a supervisory authority, the rights to an effective judicial remedy against a supervisory authority, a controller or a processor, and the possibility for data subjects to mandate a not-for-profit body, organisation or association to lodge a complaint on their behalf.

The numerous rights granted by the GDPR to individuals can be particularly challenging in

³⁷ Ulrich Matchi Aïvodji, Sébastien Gambs, Marie-José Huguet and Marc-Olivier Killijian, 'Meeting Points in Ridesharing: A Privacy-preserving Approach' (2016) 72 Transportation Research Part C: Emerging Technologies 239

³⁸ These include: (i) the right of access (Article 15 GDPR); (ii) the right to rectification (Article 16 GDPR); (iii) the right to erasure (Article 17 GDPR); (iv) the right to restriction of processing (Article 18 GDPR); (v) the right to data portability (Article 20 GDPR); (vi) the right to object (Article 21 GDPR); (vii) the right not to be subject to automated decision-making, including profiling (Article 22 GDPR); and (viii) the right to withdraw consent (Article 7(3) GDPR).

relation to complex processing activities. Indeed, generally speaking, such rights can be overreaching and thus difficult to integrate in the context of big data analytics. It is nonetheless important to carefully consider the various rights and anticipate their concrete application. This being said, technology can also provide a means to individuals to exercise their rights in a more innovative way, such as through privacy enhancing technologies.



Illustration in the transport sector: in its guidelines on the right to data portability³⁹ of 5 April 2017, the Article 29 Working Party notably advocates for a broad interpretation, whereby “*raw data processed by a smart meter or other connected objects, activity logs, history of website usage or search activities*” fall within the scope of the portability right.⁴⁰ Therefore, in a big data analytics context, the exercise of the right to portability of data collected through intelligent cars (e.g., by various sensors, smart meters, connected objects, etc.) or related to C-ITS might turn out to be almost impossible namely from an engineering perspective, particularly in view of the Article 29 Working Party's far-reaching interpretation of this right.

International data transfers

The GDPR maintains the general principle that the transfer of personal data to any country outside the European Economic Area (hereinafter the “**EEA**”)⁴¹ is prohibited unless that third country ensures an adequate level of privacy protection. Accordingly, transfers of personal data to “third countries” (i.e.

³⁹ Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (2017) WP 242

⁴⁰ By contrast, “inferred” personal data, such as “the profile created in the context of risk management and financial regulations (e.g. to assign a credit score or comply with anti-money laundering rules)” are outside the scope of the portability right.

⁴¹ The European Economic Area includes the 28 EU countries and Iceland, Liechtenstein and Norway.

to countries outside the EEA not ensuring an adequate level of protection) are restricted. In such cases, the data flow must be based on a particular instrument to allow the data transfer to take place, such as Standard/Model Contractual Clauses (SCCs)⁴², Binding Corporate Rules (BCRs)⁴³, codes of conduct and certifications, or derogations.⁴⁴

The provision of big data analytics services may entail that the personal data collected and processed will be transferred outside the EEA. This can be particularly true when relying on cloud computing services. It follows that the GDPR requirements related to the transfer of personal data must be taken into account in order to determine the most adequate solution to permit such international flow.

Any data flows should therefore be carefully assessed and mapped, notably as part of the mapping of the different actors, in order to determine the data location and put in place the adequate (contractual) instruments.

Conclusion

The present Chapter undeniably only looks into and provides illustrations of the most topical issues, without claiming exhaustiveness. It however demonstrates that finding a balance between the various interests at stake is of paramount importance. It is therefore essential to keep in mind Recital 4 of the GDPR which stipulates that the right to the protection of personal data is not an absolute right, that it must be considered in relation to its function in society and be balanced against other fundamental rights, and that this must be done in accordance with the principle of proportionality.

Accordingly, any guidance or administrative/judicial decision should carefully take into account all interests at stake. Failing to do so would necessarily impede the development of disruptive technologies and prohibit the emergence of a true data economy.

⁴² A contract between the importer and exporter of the personal data containing sufficient safeguards regarding data protection.

⁴³ A binding internal code of conduct through which multinational corporations, international organisations and groups of companies wishing to transfer data within their corporate group comprising members established outside the EEA provide safeguards with respect to data protection.

⁴⁴ Derogations include: (i) explicit consent; (ii) contractual necessity; (iii) important reasons of public interest; (iv) legal claims; (v) vital interests; and (vi) public register data. The GDPR also provides for a limited derogation for non-repetitive transfers involving a limited number of data subjects where the transfer is necessary for compelling legitimate interests of the controller (which are not overridden by the interests or rights of the data subject) and where the controller has assessed (and documented) all the circumstances surrounding the data transfer and concluded there is adequacy. The controller must inform the supervisory authority and the data subjects when relying on this derogation.



Anonymisation/ pseudonymisation



In this third Chapter, we look, on the one hand, at the impact of anonymisation and pseudonymisation in a personal data protection context and, on the other hand, into the possible use of anonymisation and pseudonymisation techniques as a way to protect non-personal data.

First and foremost, it shall be noted that a discrepancy may exist between the legal and technical definitions of certain anonymisation and pseudonymisation techniques discussed in this Chapter. For the purpose of our legal analysis, this Chapter will rely on the legal definitions as outlined below.

Anonymisation, nowadays used as a common denominator for different types of techniques, can be described as a process by which information is manipulated (concealed or hidden) to make it difficult to identify data subjects.⁴⁵ The Oxford English Dictionary defines it as the act of removing identifying particulars or details for statistical or other purposes.

In its Opinion 05/2014 on Anonymisation Techniques, the Article 29 Working Party (the predecessor of the European Data Protection Board) discusses two different families of anonymisation techniques:⁴⁶

- *Randomisation*: anonymisation techniques that alter the veracity of the data in order to remove the strong link between the data and the individual. This family includes techniques such as noise addition, permutation, and differential privacy.

⁴⁵ Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701, 1707

⁴⁶ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 11-20

- *Generalisation*: anonymisation techniques that generalise, or dilute, the attributes of data subjects by modifying the respective scale or order of magnitude. This family includes techniques such as aggregation or K-anonymity, L-diversity, and T-closeness.

Pseudonymisation as a specific technique has gained attention more recently with its explicit codification into the GDPR. Indeed, the GDPR now specifically defines pseudonymisation as a technique of processing personal data in such a way that it can no longer be attributed to a specific individual without the use of additional information, which must be kept separately and subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.⁴⁷

The Article 29 Working Party had however already discussed it in its Opinion 05/2014 on Anonymisation Techniques, and notably gave the following examples of pseudonymisation techniques:⁴⁸

- *Encryption with secret key*: a technique whereby plain text is changed into unintelligible code and the decryption key is kept secret.
- *Deterministic encryption with deletion of the key*: a technique whereby a random number is selected as a pseudonym for each attribute in a

⁴⁷ GDPR, art 4(5)

⁴⁸ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 20-23

database and the correspondence table is subsequently deleted.

- *Hashing*: A technique that consists of irreversibly mapping input of any size to a fixed size output. In order to reduce the likelihood of deriving the input value, salted-hash functions or keyed-hash functions with stored or deleted key may be used.
- *Tokenisation*: A technique that consists of replacing card ID numbers by values that have reduced usefulness for an attacker.

The techniques and their respective definitions discussed above demonstrate the techniques' importance in a personal data protection context. However, on the basis of our research, we believe that anonymisation and pseudonymisation techniques may prove to be apt instruments to protect non-personal information in a technical manner.

Anonymisation and pseudonymisation of personal data

By their very nature, anonymisation and pseudonymisation perform different functions in the framework of data protection law. A major difference between the two concepts relates to the goals of the techniques. The goal of anonymisation is primarily to remove linking attributes and to avoid or impede the identification of individuals.⁴⁹ Pseudonymisation, however, is not aimed at rendering a data subject unidentifiable, given that – at least in the hands of the data controller – the original data are either still available or deducible. The different functions are discussed below.

Anonymisation and pseudonymisation as a processing subject to data protection law

The Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques emphasises that "*anonymisation constitutes a further processing of personal data.*"⁵⁰ The same reasoning can be applied to pseudonymisation, which is apparent from the definition of pseudonymisation included in the GDPR.⁵¹

This entails that, when applying an anonymisation or pseudonymisation technique to personal data, one must comply with the data protection principle

of purpose limitation, and notably with the requirement of compatibility with the purpose for which the data were initially collected (see also our previous Chapter Privacy and Data Protection).⁵² In other words, anonymising or pseudonymising personal data for purposes not compatible with the original purpose amounts to a violation of data protection rules unless there are other lawful grounds for processing.⁵³

Such strict application is criticisable as it may discourage data controllers from applying such techniques in the first place. Furthermore, as will be demonstrated below, anonymisation and pseudonymisation may serve as a means to comply with certain data protection rules, such as data protection by design, security of processing, and the purpose limitation principle itself. Therefore, on the premise that anonymisation and pseudonymisation techniques are applied to appropriately secure personal data and comply with other aspects of the GDPR, this should be considered to be compatible with – or even an inherent part of – the original processing purpose.

Anonymisation as a means to avoid the applicability of data protection law

Recital 26 of the GDPR specifies that data protection principles should not apply to anonymous information or to personal data rendered anonymous in such a way that the data subject is no longer identifiable. The Recital further explicitly excludes anonymous information from the GDPR's scope.⁵⁴

The same Recital however specifically states that personal data which have undergone pseudonymisation, but which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person and thus falling within the scope of GDPR.⁵⁵ In a big data context, this may be a preferred approach given that some level of identifiability may be needed, notably to achieve predictability in the analytics. It does imply

⁴⁹ Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data' (2007) WP 136, 29
⁵⁰ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 3
⁵¹ GDPR, art 4(5): 'Pseudonymisation' means the processing of personal data [...].

⁵² GDPR, art 6(4); Gerald Spindler and Philipp Schmechel, 'Personal Data and Encryption in the European General Data Protection Regulation' (2016) 7(2) JIPITEC 163 <<https://www.jipitec.eu/issues/jipitec-7-2-2016/4440>> accessed 9 January 2019

⁵³ Samson Y Esayas, 'The Role of Anonymisation and Pseudonymisation under the EU Data Privacy Rules: Beyond the 'all or nothing' Approach' (2015) 6(2) EJLT 4

⁵⁴ GDPR, Recital 26

⁵⁵ GDPR, Recital 26

however that pseudonymised data remains subject to data protection rules.⁵⁶

This Chapter therefore further examines whether and, if so, how the use of anonymisation techniques may provide a way out of the scope of data protection law.

In the context of the Data Protection Directive (repealed by the GDPR), the Article 29 Working Party highlighted in its Opinion 05/2014 that only when data is anonymised to the effect that it is no longer possible to associate it to an individual taking into account *all the means likely reasonably to be used* either by the data controller or a third party, it will not constitute personal data.⁵⁷ According to the Working Party Opinion, an effective anonymisation technique prevents all parties from singling out an individual in a dataset, from linking two records within a dataset (or between two separate datasets), and from inferring any information in such dataset.⁵⁸ In the opinion of the Working Party, this would require anonymisation as permanent as erasure, i.e. irreversible anonymisation.⁵⁹ The Working Party examines, in the third and substantial section of Opinion 05/2014, various anonymisation practices and techniques, none of which meet with certainty the criteria of effective anonymisation. Consequently, a case-by-case approach, in combination with a risk analysis, should be favoured in order to determine the optimal solution. Combinations of different anonymisation techniques could be used to reach the required (high) level of anonymisation, in which case data protection law would not apply.⁶⁰

Some commentators have been critical of the Article 29 Working Party's proposition on the basis that the Article 29 Working Party applies an absolute definition of acceptable risk in the form of

zero risk.⁶¹ They argue that data protection law itself does not require a zero risk approach and that, if the acceptable risk threshold is zero for any potential recipient of the data, there is no existing technique that can achieve the required degree of anonymisation.⁶² This might encourage the processing of data in identifiable form, which in fact presents higher risks. Therefore, such commentators claim that, when one assesses identifiability taking into account all means reasonably likely to be used, one should focus on whether identification has become "reasonably" impossible. This would be measured mainly in terms of time and resources required to identify the individual, while taking into consideration the available technology as well as technological developments.⁶³

A judgment from the Court of Justice of the European Union (hereinafter "**CJEU**") of 19 October 2016 in the *Breyer* case, though still rendered under the Data Protection Directive, might indicate a more practical mind-set. In that judgment, which dealt with the question whether dynamic IP addresses may constitute personal data, the CJEU held that the possibility to combine a dynamic IP address with the additional data held by the Internet service provider does not constitute a means likely reasonably to be used to identify the data subject "*if the identification of the data subject is prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.*"⁶⁴ This seems to indicate that the CJEU prefers to steer towards a risk-based approach and away from the Article 29 Working Party's absolute approach.

In conclusion, although the Working Party Opinion and the GDPR provide a clarification of the legal status of anonymisation and pseudonymisation techniques, they regrettably do not contain any guidance for data controllers or data processors on how to effectively anonymise or pseudonymise data.⁶⁵ Pursuant to the GDPR, however, associations and other bodies representing categories of data controllers or processors may

⁵⁶ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 10. See also p.29 noting that pseudonymisation reduces the linkability of a dataset with the original identity of a data subject; as such, it is a useful security measure but not a method of anonymisation.

⁵⁷ Data Protection Directive, Recital 26 (now GDPR, Recital 26); Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 3. The Article 29 Working party emphasises that data subjects may still be entitled to protection under other provisions (such as those protecting confidentiality of communications).

⁵⁸ *Ibid* 9; see also Commission de la protection de la vie privée, 'Big Data Rapport' (CPVP 2017) 20 on the concept of singling out <https://www.gegevensbeschermingsautoriteit.be/sites/privacy-commission/files/documents/Big_Data_Rapport_2017.pdf> accessed 9 January 2019

⁵⁹ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 6

⁶⁰ *Ibid* 24

⁶¹ Khaled El Emam and Cecilia Alvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques' (2015) 5(1) IDPL 73

⁶² *Ibid*

⁶³ GDPR, Recital 26

⁶⁴ Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:339, paras 45-46

⁶⁵ Khaled El Emam and Cecilia Alvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques' (2015) 5(1) IDPL 73

prepare codes of conduct regarding the pseudonymisation of personal data.⁶⁶ We believe such codes of conduct are indispensable to the uptake of pseudonymisation techniques in a big data context, including in the transport sector.



Illustration in the transport sector:

In its Code of Practice on Anonymisation⁶⁷, the UK Information Commissioner's Office ("ICO") looks into a case study involving the use of mobile phone data to study road traffic speeds. In such hypothesis, a telecommunications provider would share subscriber records with a research body, which would try to derive information about traffic speeds by looking at the speed with which individual phones are moving between particular locations. This would entail the processing of potentially intrusive personal information, i.e. geo-location data. According to the ICO, such processing can be avoided by replacing the mobile phone numbers with dummy values. The telecommunications provider could achieve this either through encryption of the individual data records or through tokenisation. In both instances, it is essential that the encryption key, respectively the mapping table are kept secret.

Anonymisation and pseudonymisation as a means to avoid the applicability of specific data protection obligations

Even if data protection law applies in general, anonymisation and pseudonymisation may serve as mechanisms to release data controllers or processors from certain specific data protection obligations related to personal data breach (such obligations will be further addressed in the next Chapter on Breach-related Obligations).

Anonymisation and pseudonymisation as a means to comply with data protection law

Anonymisation and pseudonymisation may also constitute a means to comply with certain data protection rules. Thus, even when the application of data protection law cannot be bypassed, some techniques may facilitate complying with it. In this respect, Recital 28 of the GDPR explicitly provides that *"the application of pseudonymisation to personal data can [...] help controllers and*

processors to meet their data protection obligations."

- *Data protection by design and by default*: As discussed in our previous Chapter Privacy and Data Protection, controllers must implement 'appropriate technical and organisational measures' to ensure the data protection principles under Article 5 of the GDPR are complied with in an effective way and to integrate the necessary safeguards into the processing in order to meet the requirements of the GDPR. Such measures may result, for example, from pseudonymisation techniques.⁶⁸
- *Security of processing*: Controllers (and processors) are required to implement appropriate technical and organisational measures.⁶⁹ Such measures shall take into account several factors such as (i) the state of the art; (ii) the costs of implementation; (iii) the nature, scope, context, and purposes of the processing; and (iv) the risk of varying likelihood and severity for the rights and freedoms of natural persons. The GDPR goes further than the former Data Protection Directive as it provides specific – yet limited – suggestions for what types of security measures might be considered "appropriate to the risk". The first of these suggested measures is *"the pseudonymisation and encryption of personal data"*.⁷⁰
- *Purpose limitation (further processing of personal data)*: According to the purpose limitation principle⁷¹, personal data must be collected for specified, explicit and legitimate purposes and not further processed in a manner incompatible with those purposes. In order to ascertain whether such processing for another purpose is compatible with the purpose for which

⁶⁸ Recital 78 of the GDPR adds that to demonstrate compliance with the GDPR, the data controller should adopt internal policies and implement measures to meet the principles of data protection by design and by default. It expressly recognises that such measures could include the pseudonymisation of personal data as soon as possible.

⁶⁹ GDPR, art 32

⁷⁰ Recital 83 of the GDPR further specifies that, in order to maintain security and to prevent processing in infringement of the GDPR, the data controller or processor should assess the risks inherent in the processing and implement measures to mitigate those risks, such as encryption. Similarly, in its 'Statement on encryption and their impact on the protection of individuals with regard to the processing of their personal data in the EU', the Article 29 Working Party highlights the necessity of using encryption techniques to guarantee confidentiality and integrity of personal data, and encourages the use of end-to-end encryption for data transfers (Article 29 Data Protection Working Party, 'Statement of the WP29 on encryption and their impact on the protection of individuals with regard to the processing of their personal data in the EU' (11 April 2018))

⁷¹ GDPR, art 5(1)(b)

⁶⁶ GDPR, art 40(2)(d)

⁶⁷ Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (ICO 2012) 68 <<https://ico.org.uk/media/1061/anonymisation-code.pdf>> accessed 9 January 2019

the personal data were initially collected, the GDPR requires the data controller to take into account the existence of appropriate safeguards, including pseudonymisation and encryption.⁷²

- *Storage limitation*: The storage limitation principle⁷³ requires personal data to be kept in a form permitting identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. This would call for either the deletion or the (effective) anonymisation of such data.⁷⁴

Techniques of anonymisation as a way to protect non-personal data

It cannot be excluded that certain stakeholders participating in big data analytics, including in the transport sector, engage in the disclosure of their trade secrets⁷⁵. The big data analytics lifecycle may also include the analysis of confidential information, which for some reasons may not qualify as a trade secret. Any disclosure of such confidential information may be potentially harmful to the commercial interests of the stakeholder involved.

Considering the commercial value of trade secrets and/or confidential information to any given company, it is essential to prudently protect them. This may be done by solely providing access to such information on a strict need-to-know basis or by putting in place non-disclosure agreements with anyone who needs to have access to the information. Such practical and contractual considerations may well be a good basis for protection, but they are not always sufficient. For instance, a contract cannot be enforced against third parties to that contract. Moreover, a breach of a non-disclosure agreement inevitably entails the loss of the "secret" character of a trade secret and is therefore usually irreversible. In such sense, only financial compensation is available as a remedy.

⁷² GDPR, art 6(4)(e)

⁷³ GDPR, art 5(1)(e)

⁷⁴ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 7

⁷⁵ Information which meets all of the following requirements: (i) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (ii) it has commercial value because it is secret; and (iii) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret. (Directive (EU) 2016/943 of the European Parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L 157/1, art 2(1))

Finally, practical and contractual solutions do not cover the situation of loss of information through theft or leaks, when the company was not willing to share the information in the first place.

It may therefore prove useful to implement a technical protection to supplement the practical and contractual protection and to render theft or leaks of non-personal information difficult or even impossible. The requirements related to the technical protection of data may then be reflected in the contractual terms, such as in the parties' obligations and warranties. From a legal perspective⁷⁶, anonymisation and pseudonymisation techniques may prove to be good protection mechanisms, given that their legal significance has already been recognised in the context of data protection legislation, and most recently by the GDPR.⁷⁷

Using anonymisation for the protection of non-personal information, notably in a big data analytics context, may yield the following benefits:

- The implementation of anonymisation techniques may qualify as a reasonable step "*under the circumstances, by the person lawfully in control of the information, to keep it secret*" in order to have one's information fall within the scope of the Trade Secrets Directive and thus to be able to invoke legal protection.
- More in general, by implementing a technical protection like anonymisation, one may be able to demonstrate, *e.g.* in court, that one has acted as a *bonus pater familias*⁷⁸ in protecting one's own or another's assets.
- Duplicating the mechanisms of protection (i.e. implementing a combination of legal, practical, contractual and technical protections) equates to a greater protection altogether.
- Sufficiently anonymised or pseudonymised information will not be compromised in case of a data leak or breach. The same would be true for encrypted information, provided that the key to the encrypted information does not reside with a third party.

⁷⁶ But also from a technical perspective.

⁷⁷ In the same vein, one might also want to consider full database encryption (i.e. zero-knowledge databases). However, a thorough assessment is needed of the impact of such encryption on the usability of the data contained in the database.

⁷⁸ A legal fiction developed through case law and legal doctrine, which represents the standard of care that can be reasonably expected from someone in any given circumstance (also called the "reasonable person").

- The implementation of a technical protection can be a means to strengthen contracts between the stakeholders involved in the big data analytics; i.e. by increasing the data importer's liability in case it does not adequately anonymise the imported information or in case it does not sufficiently protect the key to pseudonymised information.
- Whereas a legal framework for the ownership of data is currently lacking, a more pragmatic solution may be found for the ownership of the key to pseudonymised data in the existing legal framework on software protection.⁷⁹ Hence, it may be possible for companies to frame the sharing of pseudonymised information with a copyright-type software licence over said key, thus adding an extra layer of (both legal and contractual) protection.

anonymisation technique, preserving however research essential information, would facilitate the access to and re-use of valuable data.

Taking into account the advantages anonymisation offers in protecting non-personal information, it is commendable to apply anonymisation techniques to such sensitive non-personal information shared in a big data analytics context. Indeed, if companies can be reassured about the technical protection of their information in a big data environment, they will be more willing to share that information with big data analytics service providers or with big data analytics platforms.



Illustration in the transport sector:

In their paper on Anonymization of Data from Field Operational Tests⁸⁰, Y. Barnard et al. discuss the use of anonymisation and other data processing techniques to strip logs of personal and confidential information in order to encourage data sharing for transport research and innovation projects, with a particular focus on field operational tests ("FOTs"). FOTs involve the collection of large amounts of data to study driving behaviour interacting with intelligent transport systems ("ITS"), including cooperative intelligent transport systems ("C-ITS") and automated vehicles. The data gathered in such context may be personal, commercial, and/or research sensitive. Y. Barnard et al. therefore advocate the use of anonymisation techniques, while pointing out the potential risk of losing essential information in the process. According to them, an effective

⁷⁹ Directive 2009/24/EC of the European Parliament and of the Council on the legal protection of computer programs [2009] OJ L 111/16

⁸⁰ Barnard, YF, Gellerman, H, Koskinen, S et al. (2016) Anonymization of Data from Field Operational Tests. In: Congress Proceedings. 11th ITS European Congress, 06-09 Jun 2016, Glasgow, Scotland, UK

Conclusion

Anonymisation and pseudonymisation techniques generally provide fertile ground for opportunities with respect to big data applications, including in the transport sector. In this respect, it shall be noted that the use of anonymisation is specifically encouraged by Recital 13 of the ITS Directive⁸¹ as "*one of the principles of enhancing individuals' privacy*". In addition, this Chapter explored the possibility of applying anonymisation and pseudonymisation techniques to non-personal information.

Nevertheless, account must be taken of the challenges that may arise in this respect. Most importantly, a balance will need to be struck between, on the one hand, the aspired level of anonymisation (and its legal consequences) and, on the other hand, the desired level of predictability and utility of the big data analytics.



Illustration in the transport sector:

The CabAnon Project run by the 'Laboratoire d'Innovation Numérique de la CNIL' ("**LINC**") aims to assess the utility of properly anonymised data. For this purpose, the LINC team analyses records of taxi rides in New York City. While recognising that anonymisation entails a certain loss of information and, hence, a loss in terms of accuracy and utility, LINC aims to quantify such loss. It notably looked at the NYC taxi dataset's utility with respect to the following applications: (i) allowing taxi users to identify spots in their vicinity where they are likely to quickly find a taxi using density of traffic; (ii) allowing city planners to conceive other solutions to organise mobility based on the number of passengers per taxi; (iii) allowing people to determine the best moments to commute and city planners to identify places with traffic congestion on the basis of traffic speed; and (iv) providing insights to city planners on how people move through the city and how to improve public transportation based on the direction of traffic. LINC's first results showed that exploitable results could be achieved with a rather coarse but robust anonymisation approach.

⁸¹ Directive 2010/40/EU of the European Parliament and of the Council on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport [2010] OJ L 207/ 1

It follows from the foregoing that, as such, anonymisation and pseudonymisation techniques and their legal consequences are desirable concepts in the big data analytics lifecycle, including in the transport sector. However, a better alignment is needed between the legal and technical interpretations of those concepts, so that legal and technical professionals may share a common understanding on the consequences of the use of such techniques.

Additionally, the creation of codes of conduct and similar initiatives is indispensable to support stakeholders in assessing the risk of re-identification. Such initiatives should be further developed throughout the EU, including in the transport sector.

Finally, a wider and better uptake of anonymisation and pseudonymisation techniques should be encouraged, not only in the field of personal data protection, but also with respect to non-personal information requiring or meriting protection (e.g. trade secrets), in light of the advantages of those techniques discussed in this Chapter. To this end, investment in terms of both time and money should be made to further research, elaborate, and increase the robustness of such techniques, taking into consideration their possible concrete application to different types of data.



(Cyber-)security



In this fourth Chapter, we focus on some of the (cyber-)security aspects of big data processing. Where relevant, illustrations from the transport sector will be provided.

Given that cyber-threats and attacks may have devastating consequences, the issues related to cyber-security have never been more important. For instance, in the transport sector, cyber-attacks could have potentially serious consequences on the economy but also on individuals, resulting in certain cases in loss of lives.⁸²

It follows that any organisation, including notably actors in the big data value chain, is required to observe the legal obligations related to security and cyber-security, which derive not only from the GDPR, but also from other legislative instruments at both EU and national level.

The present Chapter will look into such security requirements under the GDPR, the Network and Information Security Directive (hereinafter the "**NIS Directive**")⁸³, and other European legislations and security standards.

Security requirements under the GDPR

The requirements relating to security under the GDPR apply whenever personal data is processed (see our second Chapter Privacy and Data Protection] for the definitions of "processing" and "personal data"). Considering that the use of big data technologies may entail massive personal data processing operations, the GDPR security requirements will have to be taken into account in such context.

The GDPR security requirements can be divided into, on the one hand, personal data governance

obligations and, on the other hand, obligations relating to the security of personal data processing.

As regards the personal data governance obligations laid down in the GDPR, a general obligation is imposed upon data controllers to adopt technical and organisational measures to ensure compliance with the GDPR and, importantly, to be able to demonstrate such compliance.⁸⁴ Operating a regular audit programme, implementing privacy by design and by default measures, running DPIAs, appointing a data protection officer, etc. are all measures considered to be in line with the data governance obligations, including the security-related requirements. Such measures must be reviewed and updated on a regular basis, taking into account the changing circumstances.

As for the obligations relating to the security of personal data processing, the GDPR requires data controllers and processors to implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk.⁸⁵ In the context of big data, this entails that both data controllers and processors should continuously evaluate, manage, and document the risks involved in their respective processing activities.⁸⁶ The GDPR does not detail the security measures that can or should be put in place. It nonetheless provides some, however limited, specific suggestions for what types of security

⁸² Joint statement of the European Commission, ENISA, EMSA, EASA and ERA of 23 January 2019, "Transport cybersecurity: Raising the bar by working together" in the context of the 1st Transport Cybersecurity Conference held in Lisbon.

⁸³ Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union [2016] OJ L 194/1

⁸⁴ GDPR, art 24

⁸⁵ Such measures shall take into account the following elements: (i) the state of the art; (ii) the costs of implementation; (iii) the nature, scope, context, and purposes of the processing; and (iv) the risk of varying likelihood and severity for the rights and freedoms of natural persons.

⁸⁶ Commission de la protection de la vie privée, 'Big Data Rapport' (CPVP 2017) 58 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 10 January 2019.

measures might be considered “appropriate to the risk”.⁸⁷

The GDPR moreover indicates that adherence to an approved code of conduct or certification mechanism may be used as an element to demonstrate compliance with data governance obligations⁸⁸ as well as with security requirements.⁸⁹ Currently, such codes of conduct or certification mechanisms are still being developed throughout the EU market.

Finally, it should be borne in mind that the GDPR imposes a high duty of care upon data controllers in the selection of personal data processing service providers, i.e. their processors.⁹⁰ In a data-rich environment, such as in the context of big data processing operations, the data controller should carefully impose security obligations in its respective agreements concluded with processors, including for instance cloud service providers. Also, it shall be contractually ensured that a processor relying on a sub-processor imposes security obligations on such sub-processor equivalent to those imposed by the controller on the processor.

Security requirements under the NIS Directive

The NIS Directive was adopted on 6 July 2016 to address the increasing challenges in relation to cybersecurity.⁹¹ This EU legislation aims to develop a common approach across Europe to address potential socio-economic damage caused by attacks on the network and information systems of Operators of Essential Services (“OESs”) and Digital Service Providers (“DSPs”).

Taking into account its nature as a directive, the NIS Directive had to be implemented by the EU Member States into their national laws by May

2018.⁹² It is therefore required to carefully consider the specific obligations flowing from the national implementing laws, which may be particularly relevant in a big data context, but also in the transport sector.

The Directive imposes (online) security obligations on providers of two different types of services discussed hereunder: essential and digital services.

- **Essential service:** Article 5 of the NIS Directive defines an essential service as “*a service essential for the maintenance of critical societal and/or economic activities depending on network & information systems, an incident to which would have significant disruptive effects on the service provision.*”

EU Member States had to identify the OESs established on their territory by 9 November 2018. Operators active in the following sectors may be included: energy, transport, banking, stock exchange, healthcare, utilities, and digital infrastructure.⁹³

⁸⁷ Pursuant to Article 32(1) GDPR, these are (i) the pseudonymisation and encryption of personal data; (ii) the ability to ensure the on-going confidentiality, integrity, availability and resilience of processing systems and services; (iii) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; and (iv) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

⁸⁸ GDPR, arts 24(3) and 28(5)

⁸⁹ GDPR, art 32(3)





⁹⁰ GDPR, art 28(1)

⁹¹ Cybercrime is indeed predicted to cost the world over \$ 6 trillion per year by 2021, see Mark Hue Williams and Jamie Monck-Mason, 'Guide to the NIS Directive for Transportation Companies' (Willis Towers Watson, 8 August 2017) <<https://www.willistowerswatson.com/en/insights/2017/08/guide-to-the-nis-directive-for-transportation-companies>> accessed 10 January 2019.

⁹² NIS Directive, art 25. EU Member States had 21 months to transpose the Directive into their national laws and 6 additional months to identify the providers of essential services subject to the Directive's requirements.

⁹³ NIS Directive, Annex II

Illustration in the transport sector: The transport sector may provide services the NIS Directive considers to be essential. All modes of transportation are concerned⁹⁴, provided both by public and private entities⁹⁵. The NIS Directive classifies in an Annex the following transportation modes:

Transport Mode	Sub-sector
Air 	Air carriers
	Airport managing bodies, airports, and entities operating ancillary installations contained within airports
	Traffic management control operators providing air traffic control services
Rail 	Infrastructure managers
	Railway undertakings, including operators of rail related service facilities
Water 	Inland, sea and coastal passenger and freight water transport companies (not including the individual vessels operated by those companies)
	Managing bodies of ports including their port facilities and entities operating works and equipment contained within ports
	Operators of vessel traffic services
Road 	Road authorities responsible for traffic management control
	Operators of Intelligent Transport Systems

It follows that, given their close ties with the global economy and ever-increasing reliance on technology, many operators active in the transport sector may be under the obligation to abide by the obligations set under the NIS Directive and the implementing national legislations.

⁹⁴ NIS Directive, Annex II

⁹⁵ NIS Directive, art 4(4)

- **Digital service:** a digital service is defined as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services" and which can be qualified as one of the following: (i) an online marketplace; (ii) an online search engine; or (iii) a cloud computing service.⁹⁶

In contrast with the OESs, which are identified by each EU Member State, online businesses must self-assess whether they are targeted by the rules of the NIS Directive, and in particular whether they fall within one of the three different types of digital services mentioned above.

The fact that cloud computing services are targeted by the NIS Directive is particularly relevant in a big data context, especially in light of its broad definition; i.e. a digital service that enables access to a scalable and elastic pool of shareable computing resources. This being said, other stakeholders in the (big) data value chain, taking an active role in the provision of services (such as in the transport sector), may also be concerned by different concepts of the NIS Directive. It seems likely that the big data value chain will include operators of online market places (generally described as operators of platforms that act as an intermediary between buyers and sellers), online sites that redirect users to other services to conclude contracts or facilitate trade between parties, and sites that sell directly to consumers.

Finally, it shall be noted that even if a particular actor of the data value chain would not be qualified as a DSP or an operator of essential services, the NIS Directive obligations may indirectly apply to suppliers of digital or essential service providers as a result of flow down obligations.



Illustration in the transport sector:

The application of the NIS Directive may lead to complex situations. An integrated urban mobility plan can illustrate the possible complexity, where for instance, the plan aims to meet the following three key objectives⁹⁷:

- Provide tools for urban traffic management, notably performed by increasing the number of data sources, such as traffic cameras, social media, and data sources on roadworks, as well

as integrating and analysing data.

- Provide tools to inform drivers and public transport users regarding traffic status and traffic disruptions.
- Improve urban logistics, notably achieved by providing tools to improve the access of goods delivery vehicles to parking places. For this purpose a reservation system for selected parking places for goods delivery may be deployed and piloted.

In order to fulfil the above objectives, many actors may come into play that could qualify as OESs and DSPs or that could be obliged to take into account the NIS Directive due to flow down obligations. Indeed, in such context, road authorities responsible for traffic management control and/or operators of ITS are likely to be involved. Similarly, cloud computing service providers will be relied on. Finally, 'online market places' could be involved and targeted by the NIS Directive rules in the context of the third objective.⁹⁸

Under the new rules of the NIS Directive and the national implementing legislations, the essential and digital service providers will have to (i) interact with new key actors; (ii) implement security measures; and (iii) notify security incidents.

With regard to the security measures, the NIS Directive includes generic obligations by requiring OESs and DSPs to take appropriate and proportionate technical and organisational measures to manage the risks posed to the networks and information systems which they use for the provision of their services, and to prevent and minimise the impact of incidents affecting the security of such network and information systems.⁹⁹ These security measures must take into account the state of the art to ensure a level of security of network and information systems adequate to the risk.

More particularly, when examining the security aspects of OESs and DSPs, it is worth considering the following:

⁹⁶ NIS Directive, art 4(5)

⁹⁷ Transforming Transport, 'Integrated Urban Mobility: Tampere Pilot' (TT, 2018) <<https://transformingtransport.eu/transport-domains/integrated-urban-mobility-tampere-pilot>> accessed 10 January 2019

⁹⁸ Online market places' are defined broadly as any digital service that allows consumers and/or traders to conclude online sales or service contracts with traders either on the online marketplace's website or on a trader's website that uses computing services provided by the online market place

⁹⁹ NIS Directive, arts 14 and 16

Security of OESs	Security of DSPs
<ul style="list-style-type: none"> • “Mapping of OES Security Requirements to Specific Sectors” published by ENISA (the European Union Agency for Network and Information Security) in January 2018¹⁰⁰: such report provides a substantial and comprehensive mapping of the security requirements for OES, as they have been agreed in the “NIS Cooperation Group”, to sector-specific information security standards. It therefore associates the security requirements for OES with information security standards applicable to the sectors referred to in Annex II of the NIS Directive. • “Reference document on security measures for Operators of Essential Services” published by the NIS Cooperation Group in February 2018¹⁰¹: this document does not aim to establish a new standard nor to duplicate existing ones (e.g. ISO) but to provide Member States with a clear and structured picture of Member States’ current and often common approaches to the security measures of OES.¹⁰² 	<ul style="list-style-type: none"> • The NIS Directive stipulates that DSPs must consider the following specific elements when determining appropriate security measures¹⁰³: <ul style="list-style-type: none"> – the security of systems and facilities; – incident handling; – business continuity management; – monitoring, auditing and testing; and – compliance with international standards.¹⁰⁴ • “Technical Guidelines for the implementation of minimum security measures for Digital Service Providers”¹⁰⁵ published by ENISA to assist Member States and DSPs and to provide a common approach regarding the security measures for DSPs. • Commission Implementing Regulation (EU) 2018/151, which specifies the elements to be taken into account by DSPs for managing the risks posed to the security of network and information systems and the parameters for determining whether an incident has a substantial impact.¹⁰⁶

¹⁰⁰ European Union Agency for Network and Information Security, 'Mapping of OES Security Requirements to Specific Sectors' (ENISA 2018) <<https://www.enisa.europa.eu/publications/mapping-of-oes-security-requirements-to-specific-sectors>> accessed 10 January 2019

¹⁰¹ NIS Cooperation Group, 'Reference Document on Security Measures for Operators of Essential Services' (European Commission 2018) <[https://circabc.europa.eu/sd/a/c5748d89-82a9-4a40-bd51-44292329ed99/reference_document_security_measures_OES\(O\).pdf](https://circabc.europa.eu/sd/a/c5748d89-82a9-4a40-bd51-44292329ed99/reference_document_security_measures_OES(O).pdf)> accessed 10 January 2019

¹⁰² Ibid 5

¹⁰³ No further security requirements shall be imposed on digital service providers, aside from requirements for the protection of essential State functions and for the preservation of law and order (NIS Directive, art 16(10) *juncto* art 1(6)).

¹⁰⁴ NIS Directive, art 16(1)

¹⁰⁵ European Union Agency for Network and Information Security, 'Technical Guidelines for the Implementation of Minimum Security Measures for Digital Service Providers' (ENISA 2016) <https://www.enisa.europa.eu/publications/minimum-security-measures-for-digital-service-providers/at_download/fullReport> accessed 10 January 2019

¹⁰⁶ Commission Implementing Regulation (EU) 2018/151 laying down rules for application of Directive (EU) 2016/1148 of the European Parliament and of the Council as regards further specification of the elements to be taken into account by digital service providers for managing the risks posed to the security of network and information systems and of the parameters for determining whether an incident has a substantial impact [2018] OJ L 26/48

Although the NIS Directive is a fundamental legal instrument laying down the core cyber-security obligations, clarification will be required at EU and national level in order to truly enhance cyber-security and resilience in the various concerned sectors. More particularly, as concluded in the context of the transport sector, but also applicable to others, *"non-regulatory actions are and should be pursued to address cyber threats already today: information exchange, capabilities building, awareness raising and development of cyber skills. The transport sector should work together to lay down the foundations for a "cybersecurity culture"*.¹⁰⁷ Furthermore, (better) cooperation between technical and operational levels will be needed, as well as between international partners and relevant international organisations.¹⁰⁸

Security requirements under other legislations

It is important to note that other legal instruments may impose security requirements as well. This is particularly true in the electronic communications sector where several EU Directives, transposed in the national laws of the (currently) 28 Member States, provide for security obligations – such as for instance:

- The ePrivacy Directive¹⁰⁹: this Directive requires providers of electronic communications services to take appropriate technical and organisational measures to safeguard the security of their services, where necessary in conjunction with the provider of the public communications network.
- The Framework Directive¹¹⁰: this complements the ePrivacy Directive by requiring providers of publicly available electronic communication networks and services to take appropriate measures to manage the risks posed to the

¹⁰⁷ Joint statement of the European Commission, ENISA, EMSA, EASA and ERA of 23 January 2019, "Transport cybersecurity: Raising the bar by working together" in the context of the 1st Transport Cybersecurity Conference held in Lisbon.

¹⁰⁸ Ibid

¹⁰⁹ Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector [2005] OJ L 201/37 (ePrivacy Directive). Please note that the e-Privacy legislation is currently being reviewed and that the European Commission has issued a Proposal for an ePrivacy Regulation.

¹¹⁰ Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services [2002] OJ L 108/33 (Framework Directive). Please note that this Directive will be repealed as from 21 December 2020, in accordance with the newly adopted Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code.

security of the networks and services. The Directive also requires the providers to guarantee the integrity of their networks and continuity of supply.

- The Radio Equipment Directive¹¹¹: pursuant to this Directive, radio equipment within certain categories or classes shall incorporate safeguards to ensure that the personal data and privacy of users and subscribers are protected.

Security standards

In addition to legal requirements on security, security standards indisputably have an important role to play in big data analytics, and are therefore also relevant to actors of the data value chain. Also, relying on standards and certification schemes facilitates demonstrating compliance with legal requirements, including security requirements.

By relying on existing schemes, such as for instance the ISO/IEC 27000 series issued by the International Standards Organisation ("ISO") and the International Electrotechnical Commission ("IEC"), big data services providers can demonstrate to the regulator and to their customers that their systems are adequate, or at least that security-related measures and processes have been implemented.

Furthermore, several standards development organisations have created and are currently developing big data-specific standards. It is essential for any big data service provider to follow up closely on the evolutions in this respect.

Security in practice: a complex reality

Despite the existence of guidance on the various security obligations and how to consider them practically, the implementation of security aspects remains difficult in reality and requires further and continuous research.

A good way to illustrate the complexities of applying appropriate security measures is through so-called "adversarial images". The concept of adversarial images consists in making minor changes to manipulate machine learning algorithms. To illustrate such specific security issue,

¹¹¹ Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC [2014] OJ L 153/62 (Radio Equipment Directive)

OpenAI relies on the work performed by Cornell University.¹¹² More concretely, "starting with an image of a panda, the attacker adds a small perturbation that has been calculated to make the image be recognized as a gibbon with high confidence".¹¹³

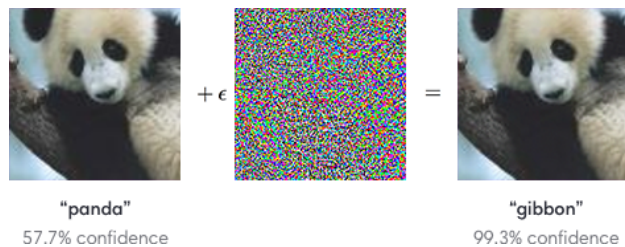


Illustration in the transport sector:

The concept of adversarial images can be particularly relevant in the transport sector. For instance, making changes to a street sign can make the algorithm think that the signs say something completely different than they actually do. The Institute of Electrical and Electronics Engineers published an article to illustrate how "Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms".¹¹⁴ Adversarial images can cause "signs that look like one thing to us to look like something completely different to the vision system of an autonomous car, which could be very dangerous for obvious reasons."¹¹⁵ For instance, in the image below, "the top row shows legitimate sample images, while the bottom row shows adversarial sample images, along with the output of a deep neural network classifier below each image."¹¹⁶



¹¹² Ian J. Goodfellow, Jonathon Shlens and Christian Szegedy, 'Explaining and Harnessing Adversarial Examples' (2015) <<https://arxiv.org/abs/1412.6572>> accessed 10 January 2019

¹¹³ Ian Goodfellow and others, 'Attacking Machine Learning with Adversarial Examples' (OpenAI, 24 February 2017) <<https://blog.openai.com/adversarial-example-research/>> accessed 10 January 2019

¹¹⁴ Evan Ackerman, 'Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms' (IEEE Spectrum, 4 August 2017) <<https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms>> accessed 10 January 2019

¹¹⁵ Ibid

¹¹⁶ Ibid

Conclusion

The requirement to put in place security measures is imposed in various legislations at EU and national level, including key instruments like the GDPR and the NIS Directive. Such legislations however remain rather general and vague as to which specific measures are deemed appropriate. It follows that organisations in the data value chain are required to:

- make a risk assessment (evaluate, manage and document the risks);
- carefully assess the available security measures on the market;
- adequately reflect the security aspects in the various contracts between stakeholders; and
- continuously assess the adequacy of the implemented measures in light of the evolving risks and the available measures.

In order to do so, organisations generally need to rely on security experts and take into account the evolving guidance documents published by authorities such as ENISA. Also, relying on certification mechanisms, seals, marks and codes of conduct will enable companies to comply with their legal obligations in terms of security and demonstrate their compliance.

Despite the enormity of the task still to be undertaken in order to improve cyber-security across the EU, the various stakeholders are aware of the need to move forward, notably through non-regulatory actions and improved cooperation. The EU institutions have also recently devised the appropriate means to tackle the cyber-security challenges, notably through the political agreement on 11 December 2018 by the European Parliament, the Council and the European Commission on the so-called "Cybersecurity Act" which aims to reinforce the mandate of ENISA and establish an EU framework for cybersecurity certification.¹¹⁷

¹¹⁷ European Commission press release of 11 December 2018 available at <https://ec.europa.eu/commission/news/cybersecurity-act-2018-dec-11_en> accessed 25 January 2019.



Breach-related obligations

In this fifth Chapter, we focus on some of the breach-related obligations in a big data context. Where relevant, illustrations from the transport sector will be provided.

In the present Chapter, we will look into the breach-notification obligations under the GDPR and the NIS Directive¹¹⁸. Subsequently, we will also look into breach notification obligations in the telecommunications sector.

Data breach notification obligation under the GDPR

The breach-related obligations under the GDPR apply whenever personal data is processed (see our second Chapter Privacy and Data Protection for the definitions of "processing" and "personal data"). Considering that big data analytics in particular may entail massive personal data processing operations, there is little doubt that these GDPR data breach notification obligations will apply to the processing of personal data in a big data context.

The GDPR requires the notification to the supervisory authority, without undue delay and in any case within 72 hours of “*a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed.*”¹¹⁹

It follows from such definition that many types of security incidents will be considered as data breaches within the meaning of the GDPR. It moreover goes without saying that the occurrence of breaches in the context of new technologies, including big data, is not hypothetical. This will require abiding by the strict obligations related to the notifications of such incidents to the appropriate data protection authorities across the EU (as well as potentially to other competent

authorities across the world in case of certain large breaches).

The table underneath provides an overview of the EU notification obligations imposed by the GDPR on the different actors involved:

¹¹⁸ Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union [2016] OJ L 194

¹¹⁹ GDPR, arts 4(12) and 33

Duty	Provision	Timing	Exemption
Data processor to notify data controller	Article 33(2) GDPR	Without undue delay after becoming aware of the data breach.	No exemptions mentioned in the GDPR, but the European Data Protection Board is tasked to issue guidelines on the particular circumstances in which a breach shall be notified.
Data controller to notify supervisory authority	Article 33(1) GDPR	Without undue delay and, where feasible, not later than 72 hours after having become aware of the data breach.	Notification is not required if the breach is unlikely to result in a risk for the rights and freedoms of individuals.
Data controller to notify affected individuals	Article 34 GDPR	Without undue delay.	Notification is not required if: the breach is unlikely to result in a high risk for the rights and freedoms of individuals; or appropriate technical and organisational protection measures were in place at the time of the incident (e.g. data encryption); or measures have been taken, subsequent to the incident, ensuring that the risk to the right and freedoms of individuals is unlikely to materialise; or it would trigger disproportionate efforts. However, in this case, a public communication or similar measure to inform the public is required.

It is therefore reminded that anonymisation techniques, as discussed in our third Chapter Anonymisation/ pseudonymisation, can serve as mechanisms to release data controllers from certain specific obligations related to personal data breach, i.e.:

- Notification of a personal data breach to the supervisory authority is not required when the data controller is able to demonstrate that the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons.¹²⁰ Although the GDPR is not explicit on this point, it could be reasonably advocated that a breach of anonymised or pseudonymised data is less likely, or even unlikely, to result in a risk to the rights and freedoms of natural persons.¹²¹

¹²⁰ GDPR, art 33(1) and Recital 85

¹²¹ Such reasoning is also supported by the Article 29 Working Party's Opinion on Personal Data Breach Notification and Guidelines on Personal data breach notification under the GDPR, pursuant to which appropriate measures, such as encryption with confidentiality of the key, may reduce the residual privacy risks on the data subject to a negligible level. In

- Communication of a personal data breach to the data subject shall not be required if the controller has implemented appropriate technical and organisational protection measures, which were applied to the personal data affected by the breach.¹²² The GDPR indeed mentions in particular "*those [measures] that render the personal data unintelligible to any person who is not authorised to access it, such as encryption.*"

addition, the Working Party recognises the utility of appropriately implemented pseudonymisation to reduce the likelihood of identification of individuals in case of a data breach, but stresses that pseudonymisation techniques as such are not sufficient to render data unintelligible (Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679' (2018) WP250rev.01, 25).

¹²² GDPR, art 34(3)(a)



Illustration in the transport sector:

In 2016, two individuals accessed user data stored on a third-party cloud-based service used by Uber. Although the incident did not breach its corporate systems or infrastructure, the hackers obtained over 600.000 U.S. driver's license numbers as well as data of approximately 57 million Uber users from around the world including names, email addresses and phone numbers.

As reported by the Financial Times: "Instead of disclosing the incident when it was discovered, senior executives decided to pay a ransom of \$100,000 to delete the stolen data." Hence, Uber had not notified the breach to any authority around the world. Its CEO only informed the world about the breach in November 2017. This has led Uber Technologies Inc. to pay in the U.S. \$148 million to settle claims related to this large-scale data breach.

In the EU, the Article 29 Working Party established a taskforce on the Uber data breach case. This taskforce, led by the Dutch DPA, is composed of representatives from the French, Italian, Spanish, Belgian and German DPAs as well as from the ICO.

Dutch DPA imposed on 27 November a fine of 600,000 Euros on Uber B.V. and Uber Technologies, Inc (UTI) for breaching the data leakage reporting obligation.

The abovementioned incident is just one example illustrating the considerable risk data breaches can pose for organisations, including big data service providers, as well as their potential impact and the consequences in case a personal data breach is not adequately notified.

Incident notification obligation under the NIS Directive

Under the NIS Directive (see also our previous Chapter (Cyber-)security), OES and DSPs must notify without undue delay to the National Competent Authority ("NCA") or the Computer Security Incident Response Team ("CSIRT") incidents having a significant impact on the continuity or provision of the services.¹²³

¹²³ NIS Directive, art. 14(3) and 16(3). Essential or digital service providers that do not comply with the security incident notifications laid down by the national provisions adopted pursuant to the NIS Directive may be subject to a penalty, which is to be determined by each EU Member State at national level.

On the basis of the NIS Directive, the factors to be considered when determining whether the impact of an incident is significant are the following:

OESs	DSPs
<ul style="list-style-type: none"> the number of users affected by the incident; the duration of the incident; and the geographical spread of the incident.¹²⁴ 	<ul style="list-style-type: none"> the number of users affected by the incident; the duration of the incident; the geographical spread of the incident; the extent of the disruption of the service; and the extent of the impact on economic and societal activities.¹²⁵

Given its nature as a directive, the NIS Directive is not directly applicable in the EU Member States but needs to be implemented in the legal order of each Member State. It can therefore be expected that there will be a difference in implementation of the security incident notification obligations between the different EU Member States, including on the concrete application of the above factors.

This being said, in addition to the above general rules included under the NIS Directive, the following clarification documents have been published at EU level:

- With respect to OESs:
 - "Reference document on Incident Notification for Operators of Essential Services – Circumstances of notification"¹²⁶, published by the NIS Cooperation Group in February 2018.¹²⁷ Such document details the incident notification scheme for OES but also the parameters used to measure the impact of incidents. It also

Pursuant to Article 21 of the NIS Directive, such penalty must be effective, proportionate and dissuasive.

¹²⁴ NIS Directive, art 14(4)

¹²⁵ NIS Directive, art 16(4)

¹²⁶ NIS Cooperation Group, 'Reference Document on Incident Notification for Operators of Essential Services. Circumstances of Notification' (European Commission 2018) <http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=53644> accessed 17 October 2018

¹²⁷ The NIS Cooperation Group is established by the NIS Directive and started its work in February 2017. It gathers national competent authorities responsible for cybersecurity and is composed of representatives of Member States, the European Commission, and ENISA. The NIS Cooperation Group facilitates the dialogue between different bodies responsible for cybersecurity in the EU. It represents a shared space where common cybersecurity challenges are discussed and coordinated policy measures are agreed upon.

examines the intricacies of cross-border situations and the interplay of the NIS Directive with notification requirements in other legislations (including the GDPR).

- “Reference document on Incident Notification for Operators of Essential Services – Formats and procedures”¹²⁸, published by the NIS Cooperation Group in May 2018.¹²⁹ Such document provides (non-binding) guidance to national competent authorities and CSIRTs with regard to formats and procedures for the notification of incidents by OES, to facilitate alignment in the implementation of the NIS Directive across the EU.

- With respect to DSPs:

- Commission Implementing Regulation (EU) 2018/151 of 30 January 2018 laying down rules for application of the [NIS Directive] as regards further specification of the elements to be taken into account by DSPs for managing the risks posed to the security of network and information systems and of the parameters for determining whether an incident has a substantial impact.¹³⁰ Such document notably clarifies four situations in which DSPs are required to notify the relevant NCA or CSIRT, notably: (i) if the digital service is unavailable for more than 5 million user-hours in the EU; (ii) if more than 100,000 users in the Union are impacted by a disruption; (iii) if the incident has created a risk to public safety, public security or of loss of life; (iv) if the incident has caused material damage of more than €1 million.

- “Guidelines on notification of Digital Service Providers incidents Formats and procedures”, published by the NIS Cooperation Group in June 2018. Such document provides non-binding technical guidance to national competent authorities and CSIRTs, with regard to formats and procedures regarding the notifications of incidents by DSPs, to facilitate alignment in the implementation of the NIS Directive across the EU.

- “Incident notification for DSPs in the context of the NIS Directive”¹³¹ report published by ENISA on 27 February 2017. Such report includes a comprehensive guideline on how to implement incident notification for DSPs.

Furthermore, some complex situations involving DSPs and OES may arise and require putting in place adequate (contractual) mechanisms. For instance, in case an operator of essential services depends on a DSP for the provision of such essential services, any significant impact on the continuity of those services due to an incident affecting the DSP must be notified by that operator.¹³² The NIS Directive remains however silent as to whether, in such circumstances, the DSP is obliged to notify such incident to the operator of essential services. It is therefore to be expected (and highly recommended) that the operator of essential services would require such notification by the DSP contractually.

Finally, it is worth noting that the notified NCA or CSIRT shall inform other Member States affected.¹³³ In such case, the NCA, the CSIRT and the single point of contact shall ensure that the service provider's security and commercial interests are safeguarded and that the information provided remains confidential. The NCA or CSIRT may also decide – after consultation of the notifying operator – to inform the public, where such public awareness would be necessary to prevent or manage an incident.¹³⁴

Essential or digital service providers that do not comply with the security incident notifications laid

¹²⁸ NIS Cooperation Group, 'Guidelines on Notification of Operators of Essential Services Incidents. Formats and Procedures' (European Commission 2018) <http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=53677> accessed 17 October 2018

¹²⁹ The NIS Cooperation Group is established by the NIS Directive and started its work in February 2017. It gathers national competent authorities responsible for cybersecurity and is composed of representatives of Member States, the European Commission, and ENISA. The NIS Cooperation Group facilitates the dialogue between different bodies responsible for cybersecurity in the EU. It represents a shared space where common cybersecurity challenges are discussed and coordinated policy measures are agreed upon.

¹³⁰ Commission Implementing Regulation (EU) 2018/151 laying down rules for application of Directive (EU) 2016/1148 of the European Parliament and of the Council as regards further specification of the elements to be taken into account by digital service providers for managing the risks posed to the security of network and information systems and of the parameters for determining whether an incident has a substantial impact [2018] OJ L 26/48

¹³¹ European Union Agency for Network and Information Security, 'Incident Notification for DSPs in the Context of the NIS Directive. A Comprehensive Guideline on How to Implement Incident Notification for Digital Service Providers, in the Context of the NIS Directive' (ENISA 2017) <<https://www.enisa.europa.eu/publications/incident-notification-for-dsps-in-the-context-of-the-nis-directive>> accessed 17 October 2018

¹³² NIS Directive, art 16(5)

¹³³ NIS Directive, arts 14(5) and 16(6)

¹³⁴ NIS Directive, arts 14(6) and 16(7)

down by the national provisions adopted pursuant to the NIS Directive may be subject to a penalty, which is to be determined by each EU Member State at national level. Pursuant to Article 21 of the NIS Directive, such penalty must be effective, proportionate and dissuasive.

Breach notification obligations in the telecommunications sector

The Directive concerning the processing of personal data and the protection of privacy in the electronic communications sector¹³⁵ (the “**e-Privacy Directive**”) was the first EU-wide legislative instrument to impose data breach notification obligations. Pursuant to the Directive, publicly available electronic communication service providers (hereinafter “**PECS providers**”) must, if they suffer a breach of security that leads to personal data being lost or stolen, inform the national authority and, in certain cases, the subscriber or individual.¹³⁶

Regulation 611/2013 on the measures applicable to the notification of personal data breaches (the “**Data Breach Notification Regulation**”) lays down the circumstances in which PECS providers must notify personal data breaches, the format of such notification and the procedure to follow.¹³⁷ Taking into account its nature as a Regulation, the Data Breach Notification Regulation has direct effect in all EU Member States, rendering any national implementation measures unnecessary.¹³⁸

The e-Privacy Directive is currently being reviewed in the framework of the EU Digital Single Market strategy. In this respect, the EU Commission held a public consultation, the report of which was made available in August 2016.¹³⁹ In its 'Opinion 03/2016 on the evaluation and review of the ePrivacy Directive', the Article 29 Working Party notably recommended to remove the provisions relating to

breach notification from the e-Privacy Directive given their “overlap” with the breach notification obligations under the GDPR (see below).¹⁴⁰ On 10 January 2017, the EU institutions adopted a draft e-Privacy Regulation, which would be directly applicable in all EU Member States.¹⁴¹ The latest version of the draft does not contain a data breach notification obligation as such, which is justified by the fact that the GDPR will apply to PECS providers.¹⁴²

Conclusion

In recent years the EU has made significant progress in terms of cybersecurity and related incident notification requirements. While it started with specific and scattered initiatives in certain sectors (e.g. telecommunications), the EU-related legal landscape has evolved, notably due to the Cyber Security Strategy and the NIS Directive.

It follows that organisations facing a security incident may need to notify such incident to one or more national competent authorities. The requirement to inform authorities will however depend on certain criteria laid down in the applicable legislations, as clarified by the guidance documents published at EU and national level. Accordingly, the various actors of the data value chain need to implement measures, procedures and policies in order to abide by the strict notification requirements and be prepared to provide the necessary information to the authorities, all within the imposed deadlines. Such requirements will also need to be adequately reflected in the various contracts between the stakeholders involved in the chain in order to adequately address any incident that may occur.

¹³⁵ Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] L201

¹³⁶ ePrivacy Directive, art 4(3)

¹³⁷ Commission Regulation (EU) 611/2013 on the measures applicable to the notification of personal data breaches under Directive 2002/58/EC of the European Parliament and of the Council on privacy and electronic communications [2013] OJ L 173/2

¹³⁸ Davinia Brennan, 'New Rules on Breach Notification by Telecoms and ISPs – Clarity at Last?' (2013) 14(1) P & DP 4.

¹³⁹ Summary report available online at <<https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-evaluation-and-review-eprivacy-directive>> accessed on 15 January 2019

¹⁴⁰ Article 29 Data Protection Working Party, 'Opinion 03/2016 on the evaluation and review of the ePrivacy Directive' (2016) WP 240, 19

¹⁴¹ Commission, 'Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC' (Regulation on Privacy and Electronic Communications), COM(2017) 10 final

¹⁴² Whereas GDPR focuses on general uses of personal data, the upcoming e-Privacy Regulation will supplement the GDPR with additional rules targeted at electronic communications services, the use of cookies, online behavioural advertising, direct marketing and machine-to-machine communications.



Supply of digital content

In this sixth Chapter, we look into the the possible provision of personal data by a consumer in order to receive digital content and how this practice interacts with the applicable data protection legislation. Where relevant, illustrations from the transport sector will be provided.

Digital content, in short, means data produced and supplied in a digital form. Forms of digital content may include computer programs, games, music, videos, applications, cloud storage and potentially social media.

Setting the stage

The fact that digital content can be provided "free of charge" is particularly popular with consumers, who have shown a strong appetite for such content. Indeed, only a small minority of consumers pays for digital content on a regular basis. Such "free" models allow companies to reach a large pool of consumers and thereby enable them to quickly test new ideas and innovative services. In this regard, digital companies foster the common perception that such digital content is indeed provided for free, while in reality it requires users to surrender valuable personal data in exchange and provides multiple future monetisation possibilities for companies.



Illustration in the transport sector:

In order to navigate the internet and to use "free" Wi-Fi services in airports or on public transport, users need to accept cookies and provide their email address. In essence, if a user wishes to make use of free internet, he or she must disclose to the supplier (who will often further share or sell such content to third parties) his or her email address, location data, history of the websites visited, etc.

Since the Cambridge Analytica data scandal, which came to light in March 2018, the provision of personal data as counter-performance for "free"

digital content has gained public visibility. The extent to which personal data can be monetised by companies has given rise to heated debates.

In terms of EU law, the EU adopted on 20 May 2019 a Directive setting new rules on sales contracts for goods and digital content (the "**Digital Content Directive**")¹⁴³, aimed at reconciling the EU legal framework on consumer and contract law with the economic reality.

The Directive shall apply to any contract where the trader supplies or undertakes to supply digital content or a digital service to the consumer and the consumer pays or undertakes to pay a price. Furthermore, it will also apply where the trader supplies or undertakes to supply digital content or a digital service to the consumer, and the consumer provides or undertakes to provide personal data to the trader, except where the personal data provided by the consumer is exclusively processed by the trader for the purpose of supplying the digital content or digital service in accordance with the Directive or for allowing the trader to comply with legal requirements to which the trader is subject, and the trader does not process that data for any other purpose.

It is interesting to note that the scope of the Directive in the initial Proposal was "*any contract where the supplier supplies digital content to the*

¹⁴³ Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L 136/1. EU Member States have time until 1 July 2021 to adopt and publish the measures necessary to comply with the Directive. The actual date of effect of the Directive will be 1 January 2022.1

consumer or undertakes to do so and, in exchange, a price is to be paid or the consumer actively provides **counter-performance** other than money in the form of personal data or any other data". However, in the adopted Directive, every reference to the concept of 'provision of personal data as a counter-performance' has been deleted. Recital 24 of the Directive now even explicitly mentions that, the protection of personal data being a fundamental right, "personal data cannot be considered as a commodity".

Indeed, some commentators, such as the European Data Protection Supervisor,¹⁴⁴ had expressed criticism vis-à-vis the introduction of the explicit possibility to use personal data as a counter-performance. They argued that personal data cannot be monetised and that the Digital Content Directive, covering the field of contract law, would not be the adequate instrument to regulate the use of personal data. In particular, protection is already granted by the existing legislation on personal data protection, and in particular the GDPR. Some stakeholders did not see the need to attach legal consequences to a practice which may be observed everywhere in the digital environment. It seems this interpretation has made it through when the final version of the Directive was being negotiated.

While the scope of the Directive did not change *in se* compared to the Proposal, the change in wording sends the signal that personal data should not be considered as merchandise, and that the supply of digital content by a trader, whereby a consumer undertakes to supply personal data, should always take place in full conformity with the applicable data protection legislation.

Quantifying personal data

Although the Digital Content Directive now explicitly states personal data is not to be considered a commodity, the question still remains whether it is economically speaking possible to quantify personal data. Unlike money, there exists no standardised value for personal data. Data is rather a dynamic product, characterised by fluidity and intangibility.¹⁴⁵

¹⁴⁴ European Data Protection Supervisor, 'Opinion 4/2017 on the Proposal for a Directive on Certain Aspects concerning Contracts for the Supply of Digital Content' (EDPS 2017) <https://edps.europa.eu/sites/edp/files/publication/17-03-14_opinion_digital_content_en.pdf> accessed 17 October 2018

¹⁴⁵ Rebecca Kelly and Gerald Swaby, 'Consumer Protection Rights and "Free Digital Content" (2017) 23(7) Computer and Telecommunications Law Review 165, 168

Attaching value to personal data is however not impossible. Proof of this can be found in the different existing initiatives allowing the monetisation of individuals' personal data. Indeed, there exist several ways to assess the value of personal data. In doing so, one should take into account the expressing value of personal data ("how to express monetary value"), the pricing factors ("which object is priced") as well as the pricing systems ("how to attach value to the object"):¹⁴⁶

- **Expressing value:** Given that personal data change over time and has therefore the potential to become outdated and lose some of its value, personal data cannot simply be expressed in a currency. For that reason, it seems logical to express the value of data in monthly terms, i.e. per month. Importantly, data are suitable for reuse. Contrary to tangible products, (personal) data can be sold several times. By giving his/her data, an individual is indeed not deprived of the possibility to give the same data again to another provider. It may therefore be accurate to further express the value of personal data per person.
- **Pricing factor:** Pricing personal data does not amount to pricing the value of each individual attribute in a personal record. These attributes are on an individual basis "valueless". It is the combination of the individual attributes (i.e. datasets) that creates value. In sum, the size, the completeness and the accuracy of the datasets are playing an important role in the determination of the monetary value of personal data.
- **Pricing system:** Various methodologies for determining the value of personal data have already been identified by the OECD.¹⁴⁷ Some of them are based on market evaluation whereas some are based on individual valuation (i.e. financial results per data record, market prices for data, cost of data breach, data prices in illegal markets, surveys and economic experiments, or data on willingness of users to protect their data).¹⁴⁸

¹⁴⁶ Gianclaudio Malgieri and Bart Custers, 'Pricing Privacy: the Right to Know the Value of your Personal Data' (2018) 34(2)Computer Law & Security Review 289

¹⁴⁷ OECD, 'Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value' (OECD Digital Economy Papers, No. 220, OECD Publishing 2013) <<https://www.oecd-ilibrary.org/docserver/5k486qtxldmq-en.pdf?expires=1539782608&id=id&accname=guest&checksum=9725A618211DF41C00207963B84C43Fo>> accessed 17 October 2018

¹⁴⁸ See OECD, 'Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value' (OECD Digital Economy Papers, No. 220, OECD Publishing 2013) <<https://read.oecd-ilibrary.org/science-and->

It is therefore possible to quantify the monetary value of personal data. Nevertheless, doing so can present practical and legal challenges, which, if not properly addressed, could amount to a setback for big data.

Practical issues

Regardless of any legal implications the monetisation of personal data can entail, the following practical implications should not be lost out of sight either:

- **Variety and specificity of data uses:** Companies do not always directly monetise data. The latter is often used for a wide range of commercial purposes involving indirect monetisation, such as security or improvement of customer experience. The idea of monetisation of data therefore designates a catch-all term and fails to address the variety and the specificity of data uses. The question remains when exactly a trader has supplied or has undertaken to supply digital content and the consumer provided or has undertaken to provide personal data.
- **Inconsistency:** Returning data to consumers in the event they exercise their right to terminate the contract also presents challenges for big data. In addition to data isolation, anonymisation and pseudonymisation could further make it impossible to return the data to the user without collecting more data than currently collected. Some speak of inconsistency in the principles of data retrieval and data anonymisation.¹⁴⁹ For a dedicated analysis of the impact of anonymisation and pseudonymisation in a big data context, see our third Chapter Anonymisation/pseudonymisation.
- **Inoperability:** The data provided or generated by the users accessing the digital content enable the product or service to function. Attention should therefore be drawn to the potential impact of data retrieval or return on the remaining users' experience. In some cases, this could go so far as to render certain current content and services inoperable.

[technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en#page1](https://www.edima.eu/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf)> accessed 22 January 2019

¹⁴⁹ Deloitte, 'Impact of the European Commission's Draft Directive on Contract Rules for the Supply of Digital Content. Final Report' (Deloitte 2016) <<http://edima.eu/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf>> accessed 4 February 2018



Illustration in the transport sector:

In order to comply with the data retrieval obligation, a carpooling service may have to delete reviews users have uploaded. Returning this data would negatively alter the experience of other users of the service by affecting the featuring and star ratings of drivers.

The abovementioned, non-exhaustive, practical concerns demonstrate that further clarifications are required in order to provide greater certainty for suppliers of digital content and the big data value chain in general. The subject calls for the establishment of adequate *ex ante* guidelines, or similar initiatives to assist the suppliers of digital content.

Legal challenges

By recognising that a trader can supply digital content to a consumer who provides his or her personal data, the Digital Content Directive intends to codify a social practice. The legal recognition of a common social practice is likely to have legal consequences for both parties to the contract.

Consequently, in addition to practical challenges, several difficulties from a legal perspective can be identified in the Digital Content Directive:¹⁵⁰

- Accepting the principle that parties can contractually agree for a trader to provide digital content and a consumer to provide his or her personal data, intensifies the rights and duties of both parties. For the consumer, the Digital Content Directive makes clear that the data subject providing his/her personal data to the supplier shall have the same rights as a consumer paying money to the supplier. However, the Digital Content Directive says nothing about the duties of the consumer and the rights of the supplier. Those will therefore be regulated by national law. In the same vein, while the Digital Content Directive provides detailed rules for termination of the contract by the consumer, the Digital Content Directive remains nearly silent on the termination rights of the supplier.
- The combination of EU law for the rights of one party (the consumer) and national law for the rights of the other party (the supplier) raises a number of fundamental challenges, especially in

¹⁵⁰ Axel Metzger, 'Data as Counter-Performance: What Rights and Duties do Parties Have?' (2017) 8(1) JIPITEC 2 <<http://www.jipitec.eu/issues/jipitec-8-1-2017/4528>> accessed 29 January 2019

light of the harmonisation attempt of the Digital Content Directive and the principle of effectiveness of EU law.

- The supplier should have the right to claim the provision of a consumer's personal data within the limits of data protection law. It follows that the consumer is under an obligation to submit his/her data in accordance with the terms and conditions, as well as the privacy policy, of the supplier. This however requires looking into the intricacies of the applicable privacy and personal data protection legislation, and in particular the GDPR (see also our second Chapter Privacy and Data Protection).
- Whether the Digital Content Directive will finally improve the legal situation of consumers on the digital market will also depend on the protection given to the supplier at national level. On the one hand, it will hardly be acceptable to give full protection to the consumer providing his/her personal data without looking at the same time at the suppliers' rights in such contractual settings. On the other hand, the rights of the supplier in application of national contract laws should not be able to undermine the legislative purpose of the Digital Content Directive.
- The Digital Content Directive does not harmonise the rules on the formation of contracts, nor on the validity of the contract for the supply of digital content. Hence, these issues will also remain in the realm of autonomous national contract law.

Conclusion

The recognition that digital content can be supplied by a trader whereby a consumer provides his or her personal data, for the first time indicates the desire of the EU legislature to take into account an underlying economic reality of transactions using personal data and to express, once again, its concern regarding the protection of individuals with regard to the processing of their personal data. Such acknowledgment is per se welcome as this concept will increase transparency, raise awareness of the economic value of personal data, and foster the rational behaviour of consumers (the so-called "educational" dimension).

However, the abovementioned difference in wording between the Proposal and the adopted Directive emphasises personal data cannot be considered a commodity and that therefore, the applicable data protection legislation will always have to be taken into account when contracts fall within the scope of the Digital Content Directive.

As demonstrated through this Chapter, legalising this economic reality generates practical and legal concerns. Accordingly, clarifications and guidelines are necessary to allow a greater degree of predictability for digital market actors and to ensure the usefulness of big data.



Free flow of data

In this seventh Chapter, we focus on the free flow of data in the context of big data processing. Where relevant, illustrations from the transport sector will be provided.

The “free flow of data” is typically mentioned in the debate on restrictions to cross-border data flows. In such context, free flow of data represents an ideal scenario in which no (legal) barriers to cross-border data flows remain. Efforts have been taken at EU level with the adoption on 14 November 2018 of the Regulation on the free flow of non-personal data (hereinafter the **"Free Flow Regulation"** or the **"Regulation"**).¹⁵¹ This adds to the GDPR (see also our second Chapter Privacy and Data Protection), which stipulates under Article 3(1) that *"the free movement of personal data within the Union shall neither be restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data"*.

The present Chapter briefly addresses the topic of cross-border data flows and looks into the issues and opportunities presented by the Free Flow Regulation and, where relevant, the difficult interaction with the GDPR.

Restrictions to the free flow of data and their impact

Historically, the free flow of data has been hindered by the existence of so-called 'data localisation requirements'. Data localisation requirements are a global phenomenon and come in many different shapes and forms. They can apply to personal data or to non-personal data, but could also apply indiscriminately to all types of data regardless of their qualification. In essence however, a data localisation requirement constitutes a restriction on the flow of data from one country to another. These localisation requirements can range from a Russian law requiring all processing of Russian citizens' personal data to be carried out using servers located in the Russian Federation to a French Ministerial Circular making it illegal to use a non-“sovereign”

cloud for data produced by a public (both national and local) administration.¹⁵²

Data localisation requirements have one feature in common: they raise the cost of conducting business across borders.¹⁵³ In the EU, over 60 of such restrictions were identified in 25 jurisdictions.¹⁵⁴ These restrictions are often prompted by legislators' or policy makers' perception that data are more secure when stored within a country's border. A perception that is often ill-conceived, as data security depends on the specific security measures used to store the data rather than on the location where the data is stored.¹⁵⁵ Security measures are just as strong or weak in a foreign country as they are domestically, or in other words: a secure server in Poland should not be different from a secure server in Belgium.

Cloud service providers are particularly affected by data localisation requirements. They argue that these restrictions undermine the cloud business model, either by preventing providers from accessing markets where they do not have a data center or by preventing users themselves from using cloud services provided from another EU Member State.¹⁵⁶

¹⁵² Martina F. Ferracane, 'Restrictions on Cross-Border Data Flows: A Taxonomy' (ECIPE Working Paper, No. 1/2017) 14; 23 <<https://ecipe.org/publications/restrictions-to-cross-border-data-flows-a-taxonomy>> accessed 14 February 2019

¹⁵³ Ibid 2

¹⁵⁴ See p.37 of Annex 5 to the Commission staff working document impact assessment, citing: LE Europe study (SMART 2015/0016) and TimeLex study (SMART 0054/2016) (Commission, 'Impact assessment accompanying the document Proposal for a Regulation of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union' (Staff Working Document) SWD(2017) 304 final)

¹⁵⁵ Daniel Castro, 'The False Promise of Data Nationalism' (ITIF 2013) <<http://www2.itif.org/2013-false-promise-data-nationalism.pdf>> accessed 14 February 2019

¹⁵⁶ European Commission, 'Annex to the Synopsis Report. Detailed Analysis of the Public Online Consultation Results on 'Building a European Data Economy'' (European Commission 2017)

¹⁵¹ Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ L 303

Data localisation requirements thus limit the access of businesses and public sector bodies to cheaper and more innovative services or force companies operating in multiple countries to contract excess data storage and processing capabilities. For start-ups and SMEs (including in the transport sector), this constitutes a serious obstacle to growth, to entering new markets, and to the development of new products and services.¹⁵⁷



Illustration in the transport sector:

In 2014, Brussels Airport launched the idea to start developing cloud-based logistics applications. This resulted in the creation of BRUcloud. It enables the different stakeholders in the air cargo supply chain to work in a more integrated manner and increasingly act as a network. BRUcloud's main priority is to make data sharing in a cloud environment possible. Data is stored only in a central location. Once a company is connected to the cloud, it can start using the different existing applications and can exchange data very easily with other stakeholders instead of maintaining system-to-system connections with all different partners individually. Applications create quick and easy efficiency gains for the parties involved. Several applications have already been created to improve the cargo handling process.¹⁵⁸ The increased competition in the EU's cloud services market that would result from eliminating data localisation requirements would engender the creation of more services such as BRUcloud across the EU, which would generate cost reductions and efficiency gains for all actors in the transport sector.

The Free Flow Regulation

Recognising the fact that growth of and innovation emanating from the European data economy may be slowed down or hindered by barriers to the free cross-border movement of data within the EU, the European Commission presented a proposal for a Regulation on the free flow of non-personal data in the EU. This Regulation was adopted on 14 November 2018 and has become applicable as of 28 May 2019.

The Free Flow Regulation applies to all processing of electronic data other than personal data within

the meaning of the GDPR.¹⁵⁹ The underlying rationale for this scope of application is to complement the GDPR, which already makes up the legal framework applicable to personal data.

The Free Flow Regulation includes the following key provisions:

- A general prohibition of data localisation requirements in the EU.¹⁶⁰ EU Member States are no longer allowed to restrict the location of data processing activities to a particular Member State's territory, nor are they able to achieve the same result by imposing restrictions on the processing of data in other Member States.¹⁶¹ Only in exceptional circumstances, where justified on grounds of public security and taking into account the principle of proportionality, could a data localisation requirement be accepted;
- A double obligation for Member States as regards any existing data localisation requirements. On the one hand, they must repeal any existing laws or regulations that are not compliant with the abovementioned rules and, on the other hand, they need to justify any instances where they consider a certain data localisation requirement permissible and therefore intend to retain such requirement;¹⁶²
- The availability of (non-personal) data for authorities in the performance of their duties, establishing the principle that an authority may not be refused access to data on the basis that it is processed outside that authority's Member State. If that is the case, and the authority cannot get access, it may request assistance from a competent authority in the relevant Member State through a procedure set out in the Regulation;¹⁶³
- On data porting and the switching of service providers, no hard and fast obligations are imposed. Instead, the Regulation states that the Commission shall encourage and facilitate the development of self-regulatory codes of conduct at EU level, which among others should offer guidance on best practices in assisting end-users that wish to switch providers.¹⁶⁴

<http://ec.europa.eu/information_society/newsroom/image/document/2017-36/annex_to_the_synopsis_report_-_data_economy_A45A375F-ADFF-3778-E8DD2021E5CC883B_46670.pdf> accessed 14 February 2019

¹⁵⁷ Commission, 'Building a European Data Economy' (Communication) COM(2017) 9 final, 6-7

¹⁵⁸ Nallian, 'Streamlining Cargo at Brussels Airport' (Nallian) <<https://www.nallian.com/communities/brucloud>> accessed 17 October 2018

¹⁵⁹ Free Flow Regulation, art 2(1)

¹⁶⁰ Free Flow Regulation, art 4

¹⁶¹ See the definition of 'data localisation requirement', Free Flow Regulation, art 3(5)

¹⁶² Free Flow Regulation, art 4(3)

¹⁶³ Free Flow Regulation, art 5 and 7

¹⁶⁴ Free Flow Regulation, art 6; self-regulatory codes of conduct on the porting of data and switching between cloud service providers (SWIPO) and cloud security certification (CSPCERT)

Challenges related to the Free Flow Regulation's scope of application

As mentioned above, the Free Flow Regulation applies to electronic data, with 'data' meaning all data other than personal data as defined in the GDPR in order not to affect the existing framework for personal data protection. On the contrary, the Regulation aims to complement the GDPR and the e-Privacy Directive (2002/58/EC) and thereby create a comprehensive and coherent EU framework for the free movement of all data in the digital single market.¹⁶⁵

Upon closer analysis however of the scope of both the Free Flow Regulation and the GDPR, concerns arise regarding the alleged comprehensiveness and coherence of this free movement of data framework.

It is no secret that the definition of personal data under the GDPR is far-reaching.¹⁶⁶ The possible extent of the term "personal data" was clarified by the CJEU in its judgment of 12 May 2016, commonly known as the *Breyer* case.¹⁶⁷ In essence, the Court clarified that a piece of information can be considered personal data whenever additional information can be sought from third parties to identify a data subject.

When applying the principles of *Breyer* in practice, it is not unlikely that many individual pieces of data which *prima facie* seem to constitute non-personal data, still end up falling within the ambit of the GDPR's definition of personal data. As examples of sources of non-personal data, the Free Flow Regulation mentions the Internet of Things, artificial intelligence and machine learning, for instance as used in automated industrial production processes, as well as a few very specific

examples.¹⁶⁸ While this clarifies the European Commission's intention to a certain extent; one can imagine situations of data (re-)combination and re-identification - particularly in a context of big data analytics - that would render even these types of data personal data.

This concern also arises when for instance a set of non-personal data is ported from one controller to another and the latter then merges the data with either non-personal or personal data to generate new information or single out individuals, which results in the entire dataset becoming personal data. This is not an unlikely scenario in the context of big data analytics applications. In such scenario, this dataset will fall entirely within the scope of the GDPR, and the Free Flow Regulation will no longer apply.¹⁶⁹

This gives rise to some uncertainty as to what data will actually fall within the scope of the Free Flow Regulation. As required by Article 8(3) of the Free Flow Regulation, the European Commission on 29 May 2019 adopted guidance to clarify the situation.¹⁷⁰ In such guidance, the European Commission indicates that non-personal data falling within the scope of the Free Flow Regulation can be categorised as (i) data which was never personal, i.e. data which originally did not relate to an identified or identifiable individual; or (ii) data which were initially personal data but were subsequently rendered anonymous.¹⁷¹ With respect to the latter, we reiterate the difficulties of proper and effective anonymisation, as spelled out in our third article on Anonymisation & Pseudonymisation. Indeed, the European Commission itself stresses that the assessment of whether data is properly anonymised may be a demanding one.¹⁷²

are currently being developed in the context of the Digital Single Market cloud stakeholder working groups.

¹⁶⁵ Commission, 'Proposal for a regulation of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union' COM(2017) 495 final, 3

¹⁶⁶ Personal data is defined as "*any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly*".

¹⁶⁷ Case C-582/12. The central question in *Breyer* was whether dynamic IP addresses constitute personal data in the hands of an online service provider, when the additional knowledge required to identify a data subject is held by a third party (such as an Internet service provider). It should be noted that, while the *Breyer* judgment concerns the interpretation of personal data under the Data Protection Directive (95/46/EC), this term remains unchanged under the GDPR and the CJEU's interpretation remains relevant in this respect.

¹⁶⁸ Free Flow Regulation, Recital 9

¹⁶⁹ European Digital Rights, 'Feedback on the Free Flow of Non-personal Data' (EDRi 2017) 1 <https://edri.org/files/freeflowdata_consultation_EDRi_20180122.pdf> accessed 14 February 2019

¹⁷⁰ Communication from the Commission to the European Parliament and the Council, 'Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union', COM(2019) 250 final

¹⁷¹ *Ibid*, 6

¹⁷² *Ibid*



Illustration in the transport sector:

In its Opinion on processing personal data in the context of C-ITS, the Article 29 Working Party (the predecessor of the European Data Protection Board) offered an interesting perspective as to how much is covered by the concept of personal data. Noting that the messages exchanged by vehicles in a C-ITS contain on the one hand authorisation certificates which are associated with the sender, and that on the other hand these messages contain heading, timestamp and location data, they must be considered personal data. Moreover, the Article 29 Working Party notes that messages may communicate information concerning “signal violation”, for instance when a driver ignores a red light at an intersection. Since this constitutes a traffic violation, the data could even become criminal data, which is a special category of personal data under the GDPR.¹⁷³ This shows that what initially may be considered non-personal data - generated from sensors built into impersonal machines - may still constitute personal data and consequently lead to application of the GDPR and non-applicability of the Free Flow Regulation.

Tying the Free Flow Regulation’s application entirely to the residual category of non-personal data leads to uncertainties for the various stakeholders active in the data ecosystem. Indeed, the applicability, and the possible exceptions, of the Free Flow Regulation and the GDPR are determined entirely based upon the nature of the data. In such context, it is worth noting that in the impact assessment that was conducted in preparation of the proposal for the Regulation, a different scope of application had been envisaged. The approach presented there was to determine the Free Flow Regulation’s scope in terms of the type of data localisation requirement concerned rather than in terms of the nature of the data. This was based on the idea that the GDPR itself already eliminates a number of data localisation requirements.¹⁷⁴ With the aim of creating a comprehensive and coherent framework for the free movement of data within the EU, the approach suggested was therefore to have the Free Flow Regulation apply to all data localisation

¹⁷³ Article 29 Data Protection Working Party, 'Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS)' (2017) WP252, 7

¹⁷⁴ As indicated in the above introduction, Article 1(3) of the GDPR prevents Member States from restricting the free movement of personal data in the EU “for reasons connected with the protection of natural persons with regard to the processing of personal data”.

requirements other than those enacted for data protection purposes. As a consequence, data localisation requirements imposed on personal data would also be covered by the Free Flow Regulation, as long as they were adopted for a different purpose than the actual protection of such personal data. If localisation requirements were adopted for purposes of personal data protection, such restrictions would already be addressed by GDPR and the Free Flow Regulation would not (need to) apply.

However, the approach that was adopted eventually in the Free Flow Regulation entails that in principle, Member States could still impose data localisation requirements on personal data for other reasons than those connected with personal data protection.

Challenges with mixed datasets

A further challenge involves mixed datasets of personal and non-personal data. As acknowledged by the European Commission in its guidance published on 29 May 2019, mixed datasets represent the majority of datasets used in the data economy.¹⁷⁵ Particularly in the context of big data, which may involve large amounts of unstructured data of various natures, this raises practical concerns. In theory, applying both pieces of legislation would lead to the GDPR being applicable to all personal data elements of a dataset and the Free Flow Regulation to all non-personal data elements. In the same vein, the exceptions adopted on the basis of the GDPR or the Free Flow Regulation would depend on the type of data.

The Free Flow Regulation confirms that, in the event of a dataset composed of both personal and non-personal data, it shall only apply to the non-personal data part of that dataset. It follows that the applicable provisions of the GDPR must be fully complied with in respect of the personal data part of the set. The Regulation moreover clarifies that, in case personal and non-personal data in a dataset are “*inextricably linked*”, it should not prejudice the application of the GDPR.¹⁷⁶ In such event, the GDPR fully applies to the entire mixed dataset. The European Commission has clarified in its guidance that this is the case even when personal data represent only a small part of the dataset.¹⁷⁷

¹⁷⁵ COM(2019) 250 final, 8

¹⁷⁶ Free Flow Regulation, art 2

¹⁷⁷ COM(2019) 250 final, 9

The notion of "inextricably linked" is not defined by the Free Flow Regulation. According to the European Commission's guidance, it refers to "a situation whereby a dataset contains personal data as well as non-personal data and separating the two would either be impossible or considered by the controller to be economically inefficient or not technically feasible."¹⁷⁸ It flows from this broad interpretation that personal and non-personal data will be inextricably linked in, and GDPR will thus apply without limitation to, the majority of mixed datasets.

The question arises how such situation will be resolved in practice. Will all GDPR requirements and obligations apply unabridged to the non-personal data component of a mixed dataset in the event of, for instance, a personal data breach? Will a controller be able to impose the strict obligations emanating from GDPR on its processors through data processing agreements, also with respect to the non-personal data included in a mixed dataset? Will supervisory authorities take into account the specific circumstances related to a mixed dataset when deciding on fines? Regrettably, the guidance remains silent on these points.

In addition, given that in practice it will often be impossible to determine which parts of a dataset contain personal data and which contain non-personal data, and therefore to apply each Regulation to the relevant part of the dataset, this could again create a loophole for Member States to still impose exceptions and re-instate data localisation requirements on other grounds than public security, simply by applying data localisation requirements to personal data for reasons that are not connected to the protection of such personal data.¹⁷⁹

Therefore, while the Regulation explicitly specifies that it does not "impose an obligation to store the different types of data separately"¹⁸⁰, one can only wonder whether this is the direction market actors should be taking from now on and for the time being, where practically and technically feasible, in order to avoid overburdening enforcement actions by the supervisory authorities.

¹⁷⁸ The European Commission adds in its guidance that "separating the dataset is also likely to decrease the value of the dataset significantly", COM(2019) 250 final, 10

¹⁷⁹ Cathal Flynn, 'Shortcomings of the EU Proposal for Free Flow of Data' (2018) 45(4) InterMEDIA 30, 34

¹⁸⁰ Free Flow Regulation, recital 10 and art 2

Other challenges

Another point of uncertainty relates to the cross-border access to non-personal data for competent authorities. The Free Flow Regulation does not foresee the situation in which such disclosure of data is prohibited by the Member State in which the data is located. It does however stipulate that access to data "may not be refused on the basis that the data are processed in another Member State".¹⁸¹ Service providers could thus be confronted with a situation in which on the one hand, they are under an obligation to provide access to an authority from another Member State, and on the other hand, doing so is prohibited under the laws of the Member State in which the data is located.

Finally, the Regulation does not foresee any safeguards surrounding such access by competent authorities, for instance to protect intellectual property rights of third parties or data protected by commercial confidentiality such as trade secrets.



¹⁸¹ Free Flow Regulation, art 5(1)

Conclusion

Despite some of the challenges mentioned above, the Free Flow Regulation remains an important step in the elimination of restrictions to cross-border data flows and their negative impact on business. Companies expect cost reductions to be the main benefit of eliminating data localisation requirements. This is deemed to be particularly significant for start-ups and SMEs, as it is expected that abolishing data localisation requirements will reduce the cost of starting a business in the EU. For start-ups contemplating an activity involving extensive data storage and processing, the need to organise data storage across different countries significantly increases costs and potentially even eliminates the benefits to be realised by innovative technologies such as (big) data analytics.¹⁸²

Furthermore, start-ups in the European transport sector and in the EU in general increasingly rely on competitive cloud services for their products or services. Prohibiting localisation restrictions would therefore increase competitiveness of the EU cloud services market. This in turn could allow start-ups to go to market quicker, to increase their pace of innovation and would also support scalability and achieve economies of scale.¹⁸³ .



¹⁸² European Commission, *'Annex to the Synopsis Report. Detailed Analysis of the Public Online Consultation Results on 'Building a European Data Economy''* (European Commission 2017) 7-8
<http://ec.europa.eu/information_society/newsroom/image/document/2017-36/annex_to_the_synopsis_report_-_data_economy_A45A375F-ADEF-3778-E8DD2021E5CC883B_46670.pdf> accessed 14 February 2019

¹⁸³ Ibid

Liability



In this eighth Chapter, we look into liability issues in the context of new technologies, including with respect to big data, applied in the transport sector.

Setting the scene

The term "liability" is to be understood rather broadly, as meaning the responsibility of one party (or several parties) for harm or damage caused to another party, which may be a cause for compensation, functionally or otherwise, by the former to the latter.¹⁸⁴

Liability has already been recognised as a legal issue to be carefully assessed and further examined by EU and national authorities. More particularly, some Member States have already adopted limited initiatives to permit – under strict conditions – highly or fully automated vehicles on their road infrastructures.¹⁸⁵ At EU level, there has been no regulatory intervention to date. However, both the European Parliament and the European Commission have been very active in relation to the identification of the liability issues surrounding new or disruptive technologies, notably through the following recent publications and initiatives:

- The European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics;
- The Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability; a study (SMART 2016/0030) prepared by Deloitte for the European Commission and published in 2017;
- The Workshop on liability in the area of autonomous and advanced robots and Internet of Things systems, organised by the European Commission and held in Brussels on 13 July 2017;

- The establishment by the European Commission of a Working Group on Liability and New Technologies, which includes two formations, i.e. the Product Liability Directive and the New Technologies; and
- The European Commission Staff Working Document on liability for emerging digital technologies accompanying the Communication from the Commission on Artificial intelligence for Europe, which was published on 24 April 2018.

It clearly follows from the foregoing that the European institutions recognise the need to potentially review the current rules on liability to take into account the rise of disruptive technologies. The above initiatives however specifically aim to assess and rethink the rules in light of AI, devices that are (fully) automated and able to take autonomous decisions, and robots. Undeniably, the output of such technologies is more far-reaching than big data analytics, even if they are not mature yet. Such technologies may however rely on big data in order to function properly. Accordingly, any initiatives in relation to more far-reaching technologies will also be relevant to big data.

This Chapter therefore aims to provide a general overview of the liability issues that may arise in relation to new technologies, focusing in particular on big data in the transport sector. It will also determine whether regulatory intervention is desirable in the long and the short term.

¹⁸⁴ See Commission, 'Liability for emerging digital technologies Accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial intelligence for Europe' (Staff working document) SWD(2018) 137 final, 2 footnote 1

¹⁸⁵ SWD(2018) 137 final, 9

Extra-contractual and statutory liability and safety regimes

The current extra-contractual, statutory and safety-related liability legal framework in the EU is rather complex. This is mainly due to the high number of legal instruments regulating parts of the issue, but also to the discrepancies that may exist between Member States.

An attempt to schematise the current system in a simplistic manner may look as follows¹⁸⁶:

Liability		Safety
Member States-driven	EU-driven	
Extra-contractual liability (including tort)	Statutory liability (including product liability)	Safety requirements
<p>Relates to the civil law responsibility for damage caused outside the context of a contract (i.e. damage is caused by a violation of a right or a legitimate interest protected by law).</p> <p>Extra-contractual liability can be imposed by general civil law rules, but also by specific legislation.</p> <p>Two main categories exist:</p> <p>Fault-based liability (applicable in most Member States): the fault of the author of the wrongful behaviour must be proven by the victim. In some cases, national law introduces variations to facilitate the burden of proof.</p> <p>Strict liability: it is not dependent on a fault. The victim must only demonstrate the damage and the causal link (e.g. the damage caused by the owner of a vehicle).</p>	<p>The EU product liability legislation (Directive 85/374) provides for a strict liability regime of producers of defective products that cause damage to natural persons or their property. The regime further includes a 'cascade' system in order to ensure that the injured person can bring his/her claim.</p>	<p>The EU safety legislation aims at ensuring that only safe products can be placed on the internal market of the EU.</p> <p>This includes various instruments such as for instance:</p> <p>Directive 2001/95 on general product safety</p> <p>Directive 2006/42 on machinery</p> <p>Directive 2014/53 on radio equipment</p> <p>Such system is further reinforced by harmonised standards, where such standards provide a presumption of conformity with the EU safety legislation.</p>

¹⁸⁶ See also Deloitte, 'Emerging Issues of Data Ownership, Interoperability, (re)Usability and Access to Data, and Liability: Liability in the Area of Autonomous Systems and Advanced Robots / IoT-systems' (Openforum Europe, 13 July 2017) <http://ec.europa.eu/information_society/newsroom/image/document/2017-30/hans_graux_-_the_study_emerging_issues_of_data_ownership_interoperability_reusability_and_access_to_data_and_liability_6213FA9A-FB14-08A4-31E51A564C60F2A7_46146.pdf> accessed 26 October 2018; Martina Barbero and others, 'Study on Emerging Issues of Data Ownership, Interoperability, (re-)Usability and Access to Data, and Liability' (European Commission 2017) <<https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and>> accessed 26 October 2018.

As clearly affirmed by the European Commission, the above regimes are not specifically applicable to damages caused by new or disruptive technologies but they "certainly constitute helpful precedents or points of reference to which one can turn to further a reflection about how to best address, from a normative standpoint, certain distinguishing elements of risks and damages created by the emerging digital technologies."¹⁸⁷ It goes without saying that it should be clearly assessed whether changes to the above legal systems are needed in order to ensure effective redress mechanisms for victims, but also legal certainty for the various actors involved in such technologies.

For instance, given that various products and services generate data, which is ultimately being processed, the availability and quality of data is considered essential. However, in case of faulty or corrupted data, or situations of supply of erroneous data or analyses, the allocation of liability is unclear under the current regimes, which leads to legal uncertainty. Such issue is of course of utmost importance to all actors in the (big) data value chain.

In the context of its Staff Working Document on liability for emerging digital technologies, the Commission provides two examples relevant to the transport sector. It however does not dig into the intricacies related to the highly complex data value chain and the number of actors involved in purely data-related services (collection, analysis, aggregation, etc.), which could – to a greater or lesser extent – cause damage.



Illustration in the transport sector:

It is already possible to rely on fully autonomous unmanned aircrafts, or "drones", for instance for the delivery of packages.¹⁸⁸ A parcel delivery drone flying autonomously from the seller's warehouse to the customer's residence may cause damage in various ways, e.g. it may suddenly fall to the ground, collide in-air with another flying vessel, or drop the package resulting in property damage or personal injury. Without prejudice to any national legislation covering liability for autonomous drones specifically, it could reasonably be argued that autonomous drones are "aircrafts" and therefore covered by national and international rules regarding liability for aircrafts. The following

claims from the victim could be imagined¹⁸⁹:

- The victim would have a strict liability claim against the operator of the drone (provided that the national law on liability for aircraft accidents is considered to cover drones). Indeed, aircrafts are typically subject to a strict liability regime. In the case of autonomous drones, the operator would be the person or entity controlling the drone's overall use. The victim only needs to prove that the drone caused the damage without having to demonstrate the cause of the drone falling down or dropping the package.
- The victim could have a claim against the operator under general national tort law rules which would require demonstrating a fault on the operator's part (e.g. operating the drone under dangerous weather conditions or lack of required maintenance). The operator could under certain conditions also be responsible if the accident was caused by malfunctioning of any third-party services (e.g. GPS mapping) he chose to rely on.
- The victim may also sue the manufacturer under the national law provisions implementing the Product Liability Directive. To this end, the victim would have to prove a defect in the drone and that the damage resulted from such defect.



Illustration in the transport sector:

Only few EU Member States have thus far adopted specific rules covering highly or fully automated vehicles.¹⁹⁰ The liability regime for automated vehicles therefore generally consists of the national civil liability rules applicable to motor vehicles. Nevertheless, the Motor Insurance Directive requires all EU Member States to ensure that civil liability for the use of vehicles is covered by insurance and that the victim of an accident can bring a direct claim against the insurer of the party that caused the accident. In the event a fully automated vehicle causes an accident, the following may be held liable for the damage:

- The driver/holder of the vehicle under civil liability rules; or
- The manufacturer of the automated vehicle under national laws implementing the Product Liability Directive, provided that the victim can identify and prove a defect in the vehicle as well as the causal link between the defect and the damage.

¹⁸⁷ SWD(2018) 137 final, 9

¹⁸⁸ SWD(2018) 137 final, 11-13

¹⁸⁹ Ibid

¹⁹⁰ Ibid

Contractual liability

While the previous section only looked into the extra-contractual liability aspects, one should not ignore the contractual liability issues, which are particularly relevant with respect to the relationship between the actors of the (big) data value chain, as well as the relationship with the end-user.

On the one hand, the customer wishes to be able to act against the big data analytics host or provider in case it suffers any damage related to the use of the service. On the other hand, the big data analytics host/provider is looking to limit as much as possible its liability in case of failure, such as service failures. Also, it will want to include provisions in order to cover the hypotheses where the customer from its side may be held responsible for types of use of the platform that are not allowed.

In this sub-section, we examine issues related to limitations and exclusions of liability, both in a business-to-consumer ("**B2C**") and business-to-business ("**B2B**") context.

As a matter of principle, limitations and exclusions of liability can be regulated contractually. However, although this is possible throughout the EU Member States, there still remain discrepancies between national systems and case law.

The general principle is that parties may freely agree on liability limitations or exclusions. However, in certain instances, mandatory statutory provisions prohibit, and thus invalidate, limitation or exclusion of liability. This is typically the case for fraud, wilful intent, physical damage, or death. The question can however be a bit more complex when a party wishes to limit liability for gross negligence. In some EU Member States, liability limitations for gross negligence are prohibited, whereas in other countries these are not. Moreover, under many laws, the exoneration clause may not have the effect of rendering the agreement devoid of any meaning or purpose.

In addition to the above-mentioned rules in relation to liability and the limitation or exclusion thereof, it is important to take into account additional rules such as those related to data protection or consumer protection.

Specifically in a B2C context, clauses limiting or excluding liability may rapidly be considered as creating an imbalance between the rights and obligations of the parties. Many of the restrictions

stem from European legislation, such as the Directive on unfair terms in consumer contracts.¹⁹¹

In a B2B context, the contractual freedom between parties is usually perceived to be without any limit. Nonetheless, in certain cases, clauses agreed between professional parties may be declared invalid in case the limitation of liability clauses could be considered unreasonable. The legal grounds for these considerations differ from country to country.

It follows from the foregoing that when looking into the liability aspects, it is also important to carefully (re-)consider the contractual liability rules as these may have an impact on the actors of the data value chain, but also end-users. However, the current status of these rules, which may differ across the EU, is likely to limit the uptake of new technologies, including big data.

Limitation of liability for intermediaries – Safe Harbour

The liability of intermediaries, those entities offering infrastructures on which massive abuses of third parties' rights can occur, has been brought to the attention and has given rise to a specific liability regime at EU level (the so-called secondary liability regime or "safe harbour"). Such regime was deemed necessary in light of the rise of technologies which had enabled the multiplication of massive abuses of third parties' rights due to the ease of sharing large amounts of information via networks and platforms.

More specifically, Directive 2000/31/EC on Electronic Commerce¹⁹² ("**the e-Commerce Directive**") aims at promoting electronic commerce and tries to ensure net neutrality. This Directive attempts to achieve those objectives by prohibiting the imposition of a general monitoring obligation, and by introducing three liability exemptions, according to specific activities, namely "mere conduit"¹⁹³, "caching"¹⁹⁴ and "hosting".¹⁹⁵

¹⁹¹ Council Directive 93/13/EEC on unfair terms in consumer contracts [1993] OJ L 95/29

¹⁹² Directive 2000/31/EC of the European Parliament and of the Council on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on Electronic Commerce') [2000] OJ L178/1

¹⁹³ Ibid art 12. Mere conduit consists of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network. The acts of transmission and of provision of access include the automatic, intermediate and transient storage of the information transmitted in so far as this takes place for the sole purpose of carrying out the transmission in the communication

In short, the core idea is to protect intermediaries who are not the authors of the infringing or damaging activity but who are involved in the transit or hosting of the infringing content. This allows 'protecting', to a certain extent, such intermediaries from the tempting idea of acting against those entities that are easily identifiable, known, and creditworthy. It shall be noted however that the EU Commission is currently examining the rules related to intermediaries, as part of its Digital Single Market strategy.

It goes without saying that the emergence of new technologies and the complexity of the data value chain put pressure on the current safe harbour regime, which was not created in view of new services such as AI, IoT and big data.



Illustration in the transport sector:

The difficult application of the safe harbour regime to new players on the market can be illustrated by referring to the recent Uber judgment by the CJEU. On 20 December 2017, the CJEU provided important guidance as to the scope of the term 'information society services', as used in the E-Commerce Directive (Directive 2000/31/EC). According to the CJEU, Uber's services must be regarded as forming an integral part of an overall service the main component of which is a transport service and, accordingly, must be classified not as 'an information society service' but as 'a service in the field of transport'. The CJEU specifically ruled as follows: "*an intermediation service such as that [provided by Uber], the purpose of which is to connect, by means of a smartphone application and for remuneration, non-professional drivers using their own vehicle with persons who wish to make urban journeys, must be regarded as being inherently linked to a transport service and, accordingly, must be classified as 'a service in the field of transport' within the meaning of EU law.*" Consequently, such a service must be excluded from the e-Commerce Directive, and thus from the safe harbour regime. That means that Member States are free to regulate the conditions under which such services are to be provided.

Through the Uber ruling, the CJEU made it clear that information society services that form an

network, and provided that the information is not stored for any period longer than is reasonably necessary for the transmission.

¹⁹⁴ Ibid art 13. Caching consists of the transmission in a communication network of information provided by a recipient of the service.

¹⁹⁵ Ibid art 14

integral part of an overall service the main component of which consists of a service that is not an information society service, cannot be qualified as an information society service. Other online service providers (such as online platforms) will need to determine whether their services form an integral part of an overall service without an information society service as the main component. If that is the case, their service might not be classified as an information society service.

Liability aspects of the Directive on the supply of digital content

On 20 May 2019, an EU Directive setting new rules on sales contracts for goods and digital content (the "**Digital Content Directive**") was adopted.¹⁹⁶ The Digital Content Directive notably aims to deal with the liability of suppliers of digital content towards the consumer (read also our sixth Chapter Supply of digital content). This section aims to demonstrate briefly the necessary evolution of liability regimes in the EU in order to tackle new technologies.

According to the Digital Content Directive, the supplier's liability is limited to any failure to supply the digital content and for any non-conformity existing at the time of the supply of the digital content or digital service. In a situation where the digital content is provided on a continuous basis, the liability of the supplier is extended over the time of said supply. In other words, the digital content supplier remains liable for defects existing at the time of supply without any time limit.

Because of the complexity characterising digital content, suppliers are in the best position to prove that defects existed at the time of their supply. It is indeed almost or even impossible for consumers to properly evaluate those technical products and identify the cause of their potential defects. In other words, digital content is not subject to the classic "wear and tear" governing more traditional goods. This is why the Digital Content Directive provides for a reversal of the burden of proof, i.e. the burden of proof will lie with the supplier.

Pursuant to Recital 44 of the Digital Content Directive the supplier's liability is an essential element. The Recital more precisely states that "*the consumer should be entitled to claim compensation*

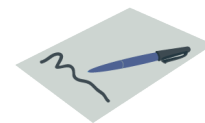
¹⁹⁶ Directive (EU) 2019/770. EU Member States have time until 1 July 2021 to adopt and publish the measures necessary to comply with the Directive. The actual date of effect of the Directive will be 1 January 2022.

for detriment caused by a lack of conformity with the contract or a failure to supply the digital content or digital services." The Digital Content Directive however leaves it up to the EU Member States to define the complete conditions for the exercise of this right to damages.

Conclusion

We welcome the EU institutions' ongoing work regarding extra-contractual and statutory liability. On such basis, it will be possible to determine whether regulatory intervention is required. In all likelihood, intervention should take place in two phases. In the short- and mid-term, non-regulatory intervention, such as the creation of model contract clauses or the identification of appropriate safety standards, should be pursued. In the long term, regulatory intervention should be considered in the form of sector-specific legislation on minimum liabilities to be borne by certain service providers in certain sectors, a general revision of liability law, and/or legislation on insurance-related obligations.

Nonetheless, this Chapter has shown that the current status of contractual liability rules, which may differ across the EU, is likely to limit the uptake of new technologies, including big data in the transport sector.





Intellectual property rights

In this ninth Chapter, we examine the aspects related to copyright, database rights and trade secrets. More particularly, we determine to what extent such protection mechanisms can apply to (big) data.

Intellectual property is defined by the Oxford English Dictionary as "intangible property that is the result of creativity". Intellectual property rights are the rights that adhere to such creations and that grant the holder(s) thereof a monopoly on the use of that creation for a specified period and subject to certain exceptions.¹⁹⁷ The underlying aim of granting such (temporary) monopoly, which – admittedly – entails a certain social cost, is to incentivise creators to share their creation with the public, and to achieve the social benefits of increased creative activity.¹⁹⁸

In light of these elements, it cannot be excluded that certain elements of the big data lifecycle, such as individual pieces of data or entire datasets, fall within the scope of protection of certain intellectual property rights. This Chapter examines those intellectual property rights that may be relevant in a big data context, and will look into the particular application in a big data environment of (i) copyright; (ii) database rights; and (iii) trade secrets.¹⁹⁹

Copyright



Copyright ensures protection of various types of works, awarding protection to individual data as long as they are original and can be expressed in a material, concrete form. The broad understanding of these protection requirements facilitates extending, in

principle, protection to different types of works, including to data.

It is however worth examining some of the most important characteristics of the EU copyright system in order to determine whether it may apply to (big) data.

Minimal EU harmonisation

Although the copyright rules applicable in the Member States are similar, the threshold of protection, the exceptions, the practical implementation, and the enforcement proceedings and remedies differ substantially. It is therefore of utmost importance to take into consideration the national legal traditions, examining both the applicable national legislation and its interpretation by national courts

The lack of full harmonisation of copyright protection at EU level is likely to have a chilling effect on EU-wide big data projects, since it requires a separate protection assessment for data originating from different Member States.

Originality

For a work to be protected by copyright, it must be original, meaning it is the author's own original creation and reflects his/her personality, where he/she has been able to express his/her creative freedom by making free and creative choices and thus stamping his/her personal touch onto the work. Generally speaking, the threshold for a work to be original is relatively low, especially in certain Member States.

This being said, although copyright protection has a broad scope, it nonetheless requires an intellectual human intervention and the consciousness of achieving a result. Therefore, raw data such as weather forecasts, stock quotations or sports scores

¹⁹⁷ R. S. Khemani and D. M. Shapiro, 'Glossary of Industrial Organisation Economics and Competition Law' (OECD 1993) <<http://www.oecd.org/regreform/sectors/2376087.pdf>> accessed 17 October 2018

¹⁹⁸ Ibid

¹⁹⁹ Computer programs, including those used to obtain, verify, store, present and analyse data, can also be protected by copyright as literary works, as set out *inter alia* in Directive 2009/24/EC on the legal protection of computer programs. The Directive also guarantees the right to create interoperable products, which is particularly important in the context of big data projects.

would in principle be excluded from copyright protection.

Unfortunately, there is no unequivocal answer as to what types of data fall under such protection, and thus, the eligibility for protection needs to be examined on a case-by-case basis and in light of the particular rules and case-law in each country.

In the context of big data projects, it is crucial to understand to what extent the data used can be copyright protected. In all likelihood, most of the data collected and processed in a big data analytics context will not be considered original and will therefore not benefit from copyright protection. Having said that, it cannot be excluded that the individual data can gain originality once they are connected with other information or presented in an original way (by means of different possible forms of expression).

Fixation

For a work to be protected, it must be fixed in some material (concrete) form. In this context, 'fixation', in a data context, would mean that the specific information needs to be saved in a tangible form. The form of saving the data can differ from handwritten notes (files), through photographic documentation (image) or recorded testimonies (sound) to digitised archives (digital files), as long as it remains concrete, can be easily identified and described. Results that have not yet been produced (future data), or results that cannot yet be described (e.g. because there are no means yet to express them) cannot benefit from copyright protection for as long as they have not materialised.

This can present some difficulties in a big data context, given that big data tends to involve dynamic datasets and notably relies on cloud computing services.

Absence of registration

The legal framework for copyright does not provide for a registration system. Accordingly, the eligibility for protection (and its scope) can only be confirmed *a posteriori* by a court, leading to a lack of legal certainty in the meantime.

Exclusive rights

The copyright holder is granted several exclusive economic rights that allow controlling the protected work's use and facilitate enforcement in case a third party uses the work without authorisation. The rights of reproduction, communication to the

public and distribution are indeed a useful toolkit which, balanced by the copyright exceptions, allows for an optimal protection of right holder's interests. Copyright law therefore provides for a wide scope of measures securing the rights of the author in case of dissemination of his work and the use of these works by third parties. The rules governing copyright protection aim at enabling further use of the works, securing at the same time the legitimate interests of the author.

In a data environment, the most important hindrance resulting from copyright protection is the necessity to obtain authorisation from the copyright holder of each individual data. In the context of big data projects, to the extent copyright applies, it would require identifying authors of hundreds (if not hundreds of thousands) of works. In many cases, it might be difficult to identify or find the right holder and/or understand whether he has given his authorisation for use of the work. In practice, this means that time-consuming analyses need to be performed before the data gathered can be used.

Furthermore, as regards the possibility to acquire copyright in data, the exclusivity of this type of right constitutes a hindrance, since it does not allow acquiring copyright in the same data "in parallel". The copyright protection foresees for the work to have one author or several co-authors (meaning respectively sole or joint ownership of rights), but excludes the possibility that different entities acquire the same right independently under a different title (e.g. if the data were collected independently or on the basis of different sources). The latter may however often be the case in a big data context, in particular where parties will be independently collecting the same or similar data, leading to the creation of convergent datasets.

Moral rights

In addition to the exclusive economic rights, authors are also granted so-called "moral rights", which are related to the idea that a work is not a mere staple commercial object, but also the expression of the personality of the author.

Moral rights are not harmonised across the EU but a common concept is included in the Berne Convention²⁰⁰, which provides for minimum standards in this respect: the author has the right, even after the transfer of the economic rights, to

²⁰⁰ The Berne Convention for the Protection of Literary and Artistic Works of 9 September 1886

claim authorship of the work and to object to derogatory actions (distortion, mutilation, or other modification) to the works which would be harmful to the author's honour or reputation. In some Member States, there is no possibility to validly assign moral rights, whereby additional measures need to be taken to guarantee that the acquirer of the economic rights is free to use and modify works protected by copyright.

Looking from a transactional angle, moral rights of authors can also be seen as a hindrance. Since at least in some Member States there is no possibility to validly assign moral rights, additional measures need to be taken to guarantee that the acquirer of the economic rights is free to use and modify data protected by copyright, to the extent necessary for big data projects.

Copyright reform

Finally, it is worth noting that on 14 September 2016, the Commission published several legislative proposals aiming to modernise the existing EU copyright rules.²⁰¹ One of the core pillars of the reform is the Directive on copyright in the Digital Single Market²⁰² (the “**DSM Directive**”). Political agreement was reached on 13 February 2019 by the European Parliament, the Council of the EU and the European Commission on the proposal for the DSM Directive. The DSM Directive does not aim to clarify the protection of data under copyright law nor provide for new rules relating to the development and increased use of digital tools such as big data and the Internet of Things. It however includes a new – yet limited – exception for text and data mining aimed at enabling universities and research organisations to use automated techniques to analyse large sets of data for scientific purposes, including in the context of public-private partnerships. The DSM Directive also introduces an additional exception into their national legislation for text and data mining for other users beyond the area of academic research. However, rightholders may expressly make reservations “*in an appropriate manner, such as machine readable*

means for the content made publicly available online”.²⁰³

Database rights



Apart from individual data, collections of data (databases) are another element important to consider when examining the protection of data, including in a big data context. When considering such protection, a distinction needs to be made between, on the one hand, the database's contents (individual data), and, on the other hand, its structure and the investment made in its creation. We examine the latter elements below.

EU specificities

While the general rules governing the protection of database are established at international level, EU law provides for a specific protection of databases which goes beyond other international legal instruments. In such respect, the EU institutions adopted the Database Directive²⁰⁴ with the objective of harmonising the protection of databases in all Member States.

Similarly to copyright, the level of protection ensured across the Member States, especially concerning the copyright on databases, is significantly different. This particularly hinders the possibility to manage pan-European projects, since it implies the necessity to examine multiple national legislations in order to have clearance on the possibility to use data, or secure the investment made in a database containing data originating from different territories.

Dual protection

The protection established by the Database Directive is dual, and supplements the possible protection granted to the data as such.

More specifically, databases, within the broad meaning of the Database Directive, are protected in the EU by (i) copyright, where such copyright protection echoes the one recognised in the international treaties; and (ii) a *sui generis* right. While copyright protects the (original) structure of the database, the *sui generis* right aims to cover the investment made in its creation. These two rights

²⁰¹ More information available at <<https://ec.europa.eu/digital-single-market/en/modernisation-eu-copyright-rules>>

²⁰² Commission, 'Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market' COM(2016) 593 final

²⁰³ Article 3a of the Proposal for a Directive on copyright in the Digital Single Market (version following the Political agreement reached on 13 February 2019)

²⁰⁴ Directive 96/9/EC of the European Parliament and of the Council on the legal protection of databases [1996] OJ L 077/20 (Database Directive)

are independent, and can be applied separately. They will however apply cumulatively if the conditions for both regimes are simultaneously met.

The term of the *sui generis* protection is much shorter than that of the copyright protection. It is limited to 15 years as from the first of January of the year following the date of completion of the database. However, such protection may in practice be much longer. According to the Database Directive, any substantial change to the contents of the database, that could be considered to be a new investment, will cause the term of protection to run anew.²⁰⁵ In practice, should such protection be applied in a big data context, this could result in providing an indefinite protection, given that databases are usually dynamic, hence, leading in all likelihood to "substantial changes to the contents of the database".

Copyright protection of databases

Copyright protection is granted to databases which, as such, by reason of the selection or arrangement of their contents, constitute the "author's own intellectual creation".²⁰⁶ A database structure may be protected under copyright even if the elements contained therein are in the public domain or are otherwise not protected by copyright.

It also follows from the previous considerations that the originality criterion might be more difficult to fulfil in case of automatically created electronic databases that contain data selected by software, without the actual involvement of an author. In such situations it seems more likely to award copyright protection to the underlying software (algorithm written in a way allowing for selection of specific data/types of data), than to the database itself.

This is particularly relevant in a big data context. Indeed, the development of technology has enabled data analytics of unstructured data. Accordingly, while protection of datasets is particularly relevant, the protection of the database structure has become

²⁰⁵ Article 10(3) of the Database Directive stipulates indeed that "any substantial change, evaluated qualitatively or quantitatively, to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations, which would result in the database being considered to be a substantial new investment, evaluated qualitatively or quantitatively, shall qualify the database resulting from that investment for its own term of protection".

²⁰⁶ No other criteria shall be applied to determine the eligibility of databases for that protection (Database Directive, art 3(1)).

less relevant and more difficult when confronted to new types of databases, unforeseen by the (over twenty-year-old) Database Directive.

Sui generis protection of databases

The second type of protection introduced by the Database Directive is the protection awarded on the basis of a *sui generis* right²⁰⁷, rewarding the substantial investment of the database maker in creating the database. It was developed in order to prevent free-riding on somebody else's investment in creating the database and exists in parallel to the copyright protection on the structure of the database.

In order for a database to be protected by the *sui generis* right, an investment must be made in the creation of the database. The jurisprudence of the CJEU has clarified that an investment in the creation of the data as such does not suffice to merit protection under the *sui generis* right.²⁰⁸ Such reasoning would entail that the *sui generis* right does not apply to machine-generated databases, as it could be argued that the data included in such databases are 'created' instead of 'obtained'. This could have a broader effect on the data economy, which relies on digitisation processes such as IoT devices, big data, and AI; as it becomes increasingly difficult to distinguish between the generation and the obtainment of data in the context of such processes.²⁰⁹

That being said, there is no automatic exclusion from *sui generis* protection when the database's creation is linked to the exercise of a principal activity in which the person creating the database is also the one creating the materials that are processed in the database. It is however always the responsibility of that person to demonstrate a substantial investment (qualitative and/or quantitative) in the obtaining, verification or presentation of the content, independent from the resources used to create the content.²¹⁰

²⁰⁷ The term "*sui generis* right" is a generic one and means "the right of its own kind".

²⁰⁸ Case C-46/02 *Fixtures Marketing Ltd v. Oy Veikkaus AB* [2004] ECLI:EU:C:2004:694; Case C-338/02 *Fixtures Marketing Ltd v. Svenska Spel AB* [2004] ECLI:EU:C:2004:696; Case C-444/02 *Fixtures Marketing Ltd v. Organismos Prognostikon Agonon Podosfairou* [2004] ECLI:EU:C:2004:697; Case C-203/02 *British Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 42

²⁰⁹ Commission, 'Evaluation of Directive 96/9/EC on the legal protection of databases' (Staff Working Document) SWD(2018) 146 final

²¹⁰ Case C-203/02 *Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 35

In any event, we foresee that it will become increasingly difficult to satisfy the *sui generis* right protection requirements in a data economy context, given that the processes of obtaining, verifying and/or presenting the data will happen more and more automatically, as they will be normally conducted using an algorithm. In many cases, it might be true that the investment in creating the raw material exceeds the investment made in segmenting and aligning that pre-existing raw material. In those cases, it might be more difficult to rely on the *sui generis* protection.

It is in our view regrettable that the Database Directive, which was drafted in the 90s, does not accommodate for the technical evolution and thus everything that is possible with data and databases today. For instance, it is unclear how techniques of enrichment, partitioning, harmonisation, homogenisation, etc. of data would fit within the criteria of obtaining, verification or presentation of the database contents. Moreover, the criterion of 'verification' may become less and less pertinent, especially in a big data context which allows analytics of unstructured data.



Illustration in the transport sector:

In 2010, the German Federal Court of Justice held in its *Autobahnmaut* decision²¹¹ that a highway company could claim a *sui generis* right in a database of machine-generated data about motorway use, i.e. toll data. The Court found that the company had made a substantial investment in the 'obtaining' of pre-existing data on cars using the motorway and in the processing of such data through software ('verifying' and 'presenting').

If the same reasoning is transposed to other databases in the transport sector, e.g. of data generated by sensors in cars, this could become problematic as certain companies (such as car maintenance services or secondary vehicle accessory providers) could be denied access to data vital to their services on the basis of a *sui generis* right.

Possibility to protect data under database rights

In view of the rules described above it seems that there is very limited to no possibility to secure individual data by means of database protection.

It is true that the *sui generis* protection forbids extraction of all or a substantial part of the database contents to another medium, preventing thus also the copying of the individual data collected in a database. However, once the database maker

renders the contents of its database accessible to the public, it cannot prevent third parties from consulting that database. The public is therefore aware of these data (information), and may use them without necessarily having to copy the database contents. Also, the current legal regime seems difficult to reconcile with developments in technologies such as big data or data mining that do not necessarily require data to be reproduced in order to perform analytics or mining processes.

In consequence, the ownership of rights to a database does not confer the rights to the individual data as such.²¹² In this context, database protection (both by copyright and the *sui generis* protection) should rather be seen as a complementary measure to protection granted to individual data under other titles such as traditional copyright or trade secret protection.

Having said that, it is important to observe that employing specific technical measures to block access to the database's content may ensure a *de facto* protection of individual data, preventing the possibility to subject them to data mining or other types of automatic filtering initiated by third parties.

²¹² Recital 45 of the Database Directive indeed states that "Whereas the existence of a right to prevent the unauthorized extraction and/or re-utilization of the whole or a substantial part of works, data or materials from a database should not give rise to the creation of a new right in the works, data or materials themselves."

²¹¹ BGH, 25 March 2010, I ZR 47/08

Trade secrets



While copyright and database rights provide measures enabling control over the diffusion and use of works (including data that fulfil the originality criterion) and databases, the objective of trade secret protection is to keep commercially valuable information confidential or secret. Protecting undisclosed know-how and business information enables its creator to transform the effort invested in generating this know-how and information into a competitive advantage.

In view of big data projects, trade secret protection may provide a safeguard as it allows for protection of individual pieces of information regardless of their originality. It also does not differentiate between the types of data that might be protected. Moreover, the protection is unlimited in time, as long as the information has not been disclosed.

EU legal framework



Similarly to databases, only general rules requiring protection of trade secrets have been embedded in international law. At EU level however, trade secret protection has been established by the adoption by the European Parliament and the Council of Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure ("**Trade Secrets Directive**").²¹³ The Directive aims to standardise the national laws of the Member States as regards the unlawful acquisition, disclosure and use of trade secrets.

The Directive harmonises the definition of trade secrets in accordance with existing internationally binding standards. It also defines the relevant forms of misappropriation and clarifies that reverse engineering and parallel innovation must be guaranteed (since trade secrets are not, strictly speaking, a form of exclusive intellectual property right).

Data protected as trade secrets

According to the definition provided in the Trade Secrets Directive, a 'trade secret' is a piece of information which meets all of the following requirements: (i) it is secret in the sense that it is

not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (ii) it has commercial value because it is secret; and (iii) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.²¹⁴

Trade secrets should be seen as complementary to intellectual property rights. They are heavily used in the creative process leading to innovation and the creation of intellectual property rights. Trade secrets are also used in relation to commercially valuable information for which there is no intellectual property rights protection, but for which investment and/or research are nevertheless required and which are important for innovation.²¹⁵ Moreover, some may prefer to opt for a trade secret protection rather than an intellectual property right, as this may allow them to have an everlasting protection (as long as the conditions for trade secret protection remain fulfilled).

In a big data context, the protection established for trade secrets will expand to every piece of information (data), as long as it fulfils the protection requirements mentioned above. Some requirements are however difficult to fulfil, such as the need for the data to remain secret. It seems that at least in some jurisdictions it is possible to rely on confidentiality agreements to ensure that the requirement of secrecy of the data under the Trade Secrets Directive is maintained even after the transfer of data has been exercised. This is however yet to be confirmed by the courts. Also, it may be difficult to demonstrate that an individual data has commercial value because it is secret. Many data will be considered valuable only if they are part of a bigger dataset.

Trade secrets rights

As such, a trade secret holder has no private or exclusive rights to its use. Trade secrets are thus different from intellectual property rights, which are safeguarded through an exclusive right that is legally enforceable. This is notably confirmed in Recital 16 of the Trade Secrets Directive which states that "*in the interest of innovation and to*

²¹³ Directive (EU) 2016/943 of the European Parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L 157/1

²¹⁴ Trade Secrets Directive, art 2

²¹⁵ European Commission, 'Trade Secrets' (European Commission, 2016) <https://ec.europa.eu/growth/industry/intellectual-property/trade-secrets_en> accessed 17 October 2018

*foster competition, the provisions of this Directive should not create any exclusive right to know-how or information protected as trade secrets". This entails that the independent discovery of the same know-how or information remains possible.*²¹⁶

In the event that one may rely on trade secret protection, the holder of a trade secret cannot prevent competitors from copying and using the same solutions – reverse engineering (i.e. the process of discovering the technological principles of a device, object or system through analysis of its structure, function and operation) is entirely lawful. Trade secrets are only legally protected in instances where someone has obtained the confidential information by illegitimate means (e.g. through spying, theft or bribery).²¹⁷

It follows that once the dataset is published, or disclosed in any other way, the protection can no longer be claimed. This is particularly relevant in a big data context, as data used for big data analytics, and made publicly available, will not qualify as trade secrets. Therefore, when considering to outsource big data analytics, any company should carefully assess whether its datasets comprise trade secrets that are valuable to the company and which cannot be disclosed for that reason.

Conclusion

It follows that it cannot be excluded that different actors in the big data analytics lifecycle will try to claim intellectual property rights or protection under trade secrets in (parts) of the datasets intended to be used. They may therefore try to exercise the exclusive rights linked to the intellectual property right concerned or keep the information secret. Any unreasonable exercise of rights may stifle data sharing and thus innovation through big data, including in the transport sector. This is however mainly due to the inherent nature and purpose of intellectual property rights and trade secrets protection, which may at the same time provide an incentive for stakeholders to engage in data sharing for big data purposes.



²¹⁶ Trade Secrets Directive, Recital 16

²¹⁷ European Commission, 'Frequently Asked Questions: Protection against the Unlawful Acquisition of Undisclosed Know-how and Business Information (Trade Secrets)' (European Commission, 2016) <https://ec.europa.eu/growth/industry/intellectual-property/trade-secrets/faq_en> accessed 17 October 2018

Open data



In this tenth Chapter, we address various legal issues and opportunities that one may encounter when using open data for big data technologies. As in our previous Chapters, illustrations from the transport sector will be provided where relevant.

The 'big data' required to feed big data analytics tools can emanate from a variety of sources. One such source is the public sector, which has been opening up certain of its datasets to the public.²¹⁸ Such public disclosure and use of these datasets is however subject to rules at both EU and national level, which will be discussed in this Chapter.

What is open data?

In the context of the Digital Single Market strategy of the European Commission, the concept of open data refers to "*the idea that certain data should be freely available for use and re-use*".²¹⁹ "Open data" moreover increasingly refers to so-called public sector information ("**PSI**"), i.e. material produced, collected, paid for and/or held by public sector bodies at national, regional and local level, such as ministries, agencies, municipalities, but also by organisations that are mainly funded by or under the control of a public authority.²²⁰

The EU institutions have taken both legislative and non-legislative measures to encourage the uptake of open data. On the non-legislative front, the European Commission has been very active in the field of open data, providing for soft measures facilitating access to data. Its involvement has included engaging with Member States through the Public Sector Information expert group (PSI

Group), funding the Legal Aspects of Public Sector Information (LAPSI) network and developing the EU Open Data Portal, which provides access to data from the EU institutions and bodies for re-use²²¹, to name a few.

The PSI Directive

On the legislative front, the EU adopted its first set of rules on the re-use of public sector information (the "**PSI Directive**") already in 2003.²²² The aim of that Directive was not so much to make public data more accessible and encourage its re-use, but to ensure that when public sector bodies decided to make data available, they did so in a fair and non-discriminatory manner.²²³ Consequently, while public authorities had to comply with these requirements when they decided to make data available, the making available of data as such had not been made mandatory. The initial version of the PSI Directive was even primarily aimed at paper documents, even though electronic data already fell within its scope of application.

In 2013, the PSI Directive was given a thorough makeover in order to keep pace with technological developments, which had led to the rise of the data economy, and to unlock the potential of big data held and accumulated by government authorities.²²⁴ In a significant departure from the first PSI Directive, an actual obligation was introduced for public sector bodies to make PSI

²¹⁸ This is for instance the case where a national ministry for transport makes available a dataset containing public transport data, following which that dataset can be used by private companies to develop commercial products and services.

²¹⁹ European Commission, 'Open Data' (*European Commission*, 8 June 2018) <<https://ec.europa.eu/digital-single-market/en/open-data>> accessed 18 October 2018

²²⁰ European Commission, 'European Legislation on Reuse of Public Sector Information' (*European Commission*, 25 April 2018) <<https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>> accessed 18 October 2018

²²¹ Accessible online at <<http://data.europa.eu/euodp/en/data>>

²²² Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, OJ L 345, 90

²²³ PSI Directive, Recital 8

²²⁴ Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, OJ L 175, 1

available.²²⁵ This effectively eliminated the possibility for public bodies to avoid application of the Directive by deciding not to make information available in the first place. Still, the amended Directive includes a number of exceptions to the principle of mandatory data provision.²²⁶ Other provisions introduced by the Directive stipulate, among others, that the information can be made available "as is" or subject to conditions, which can be imposed by way of a licence. Member States are moreover encouraged to develop standard licences that should be made available in digital format.²²⁷

Opportunities in the use of open data

PSI is a resource with great potential for a number of beneficiaries, ranging from other public sector bodies, to private businesses including start-ups, SMEs and multinationals, to academia and citizens themselves.²²⁸ Start-ups and SMEs typically do not have the same amount or type of resources as larger companies, and as a result may encounter difficulties when trying to gain access to certain data or may even fail to obtain access altogether. This competitive disadvantage can constitute a barrier for start-ups and SMEs to enter certain markets. The PSI Directive attempts to remove this disadvantage with respect to public sector information, among others through the non-discrimination principle. This principle ensures that start-ups and SMEs are able to use PSI for commercial purposes under the same conditions as would be imposed on any other company for a similar purpose.²²⁹

In the transport sector, open government data covers a wide variety of data categories. Departure and arrival times, timetables of public

transportation, fares, safety-related or other types of disruptions are only a few types of information that is typically held by public sector entities. As this data is opened up to the public in an open, standardised, machine-readable format, SMEs and start-ups may be enabled to enter markets they would have been prevented from entering if they were required to gather the relevant data in other ways. Similarly, the proliferation of tools to analyse this information, including tools for big data analytics, can pave the road for those companies to explore new business opportunities.



Illustration in the transport sector:

In maritime industries, a huge amount of data is created and collected through AIS. 'AIS' stands for Automatic Identification System and was created as a navigation and anti-collision tool. Hoping to foster innovation in the industry, the Danish Maritime Authority decided in 2016 to make historical AIS data available through an open data platform, in addition to the live AIS data feed that it was already offering.²³⁰ While AIS was originally designed to improve maritime safety conditions, many other uses can be envisaged. One application that could result from the accessibility of AIS data is being considered in the port of Rotterdam, where AIS data is used to analyse current and historical vessel dwell times. The dwell time of a ship in a port is the time during which it is docked. Long, avoidable dwell times are a big waste of time and resources for operators. The analysis of AIS data aims to forecast dwell times, which individual shippers would then be able to use to support transport decisions.

Challenges in the use of open data

Today, an EU-based company that wishes to rely on PSI for big data applications may still encounter several challenges, three of which we will touch upon in this section: (i) licensing; (ii) the interplay between the legal regimes on open and personal data; and (iii) the interplay between the PSI Directive and the Database Directive.²³¹

The PSI Directive allows public sector bodies to make the re-use of data subject to conditions, notably through the use of licences.²³² While

²²⁵ Consolidated PSI Directive, art 3 (1)

²²⁶ Public sector information that contains personal data or is covered by intellectual property rights for instance must not be made available. Exceptions also apply for certain institutions (e.g. museums, libraries, and archives) and for situations where the authority has to generate revenue to cover a substantial part of the costs relating to its public duties. (Consolidated PSI Directive, art 2)

²²⁷ Consolidated PSI Directive, art 8

²²⁸ Barbara Ubaldi, 'Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives' (OECD Working Papers on Public Governance, No. 22, OECD Publishing 2013) 11 <<https://www.oecd-ilibrary.org/docserver/5k46bj4f03s7-en.pdf?expires=1539851361&id=id&accname=guest&checksum=92B1E44F15BE9F52F8C3A2974C9F062D>> accessed 18 October 2018

²²⁹ Stefaan Verhulst and Robyn Caplan, 'Open Data: A Twenty-first-century Asset for Small and Medium-sized Enterprises' (The Governance Lab 2015) 11 <<http://images.thegovlab.org/wordpress/wp-content/uploads/2015/08/OpenData-and-SME-Final-Aug2015.pdf>> accessed 18 October 2018

²³⁰ MI News Network, 'Danish Maritime Authority Makes Historical AIS Data Available To Everybody' (*Marine Insight*, 28 December 2016) <<https://www.marineinsight.com/shipping-news/danish-maritime-authority-makes-historical-ais-data-available-everybody/>> accessed 18 October 2018

²³¹ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 20

²³² The only limitation in this respect is the fact that conditions may not "unnecessarily restrict possibilities for re-use and shall

Member States are required to have in place standard licences for the use of public sector information, public sector bodies are merely "encouraged" and thus not obliged to use them.²³³ Despite guidelines on recommended standard licences being adopted by the Commission in 2014²³⁴, little uniformity is seen as EU Member States have embraced very different licensing practices.²³⁵ As a consequence, any company that wishes to reuse PSI from different Member States with the aim of developing a product is obliged take into account as many (and perhaps even more) licences as the number of Member States in which it operates.

On the interplay between open and personal data it should be noted that, in theory, the relationship between the PSI Directive and GDPR evokes little question. The former clearly states that it is without prejudice to the rules on personal data protection (at the time still contained in Directive 95/46/EC) and that documents may be excluded from the Directive's scope on account of data protection rules.²³⁶ In the same vein, the GDPR clarifies that the PSI Directive in no way affects "*the level of protection of natural persons with regard to the processing of personal data*" and does not alter the rights and obligations set out in the GDPR. It does however allow the principle of access to PSI to be taken into account when applying the GDPR.²³⁷ While the abovementioned rules should not be understood as meaning that PSI containing personal data cannot in any case be disclosed, they nevertheless create a tension which typically leads to PSI remaining locked.

Still, what the above really implies is that a careful assessment should be made to determine the circumstances under which the personal data part of PSI could lawfully be disclosed. That assessment involves among others determining whether the relevant public sector dataset contains personal data and if that is the case, ensuring that following

not be used to restrict competition". (Consolidated PSI Directive, art 8(1))

²³³ Ibid art 8(2)

²³⁴ The Commission published Guidelines in July 2014 to help the Member States implement the revised rules and to indicate best practices regarding recommended standard licences, datasets, and charging for the re-use of public sector documents. See Commission Notice Guidelines on recommended standard licences, datasets and charging for the reuse of documents [2014] OJ C 240/1

²³⁵ In some Member States, notably Poland, public authorities do not promote any model licence agreements. In others, like France and the United Kingdom, standard licences are in force. In other Member States such as Belgium, a lack of unity even exists within the different levels of government.

²³⁶ Consolidated PSI Directive, arts 1(2)(cc) and 1(4)

²³⁷ GDPR, Recital 154

disclosure, the dataset is processed in accordance with data protection laws.²³⁸ This gives rise to a number of additional challenges, among others stemming from the broad definition of "personal data". Another example is the fact that making available PSI for re-use for all commercial and non-commercial purposes risks being at odds with the principle of purpose limitation enshrined in the GDPR. The same holds true for the principle of data minimisation. A potential means to avoid grave violations of the GDPR would be to conclude agreements with third parties to make arrangements for bilateral data sharing involving exclusivity, but these are principally forbidden by the PSI Directive as such practice would not create a level playing field.²³⁹ It is thus clear that data protection legislation presents a unique challenge to the opening up of public sector information, either because it risks preventing a large part of PSI datasets from being disclosed altogether or because it creates compliance issues when public sector bodies do decide to disclose PSI containing personal data.

Uncertainty also exists about the precise relationship between the PSI Directive and the Database Directive. The PSI Directive states that it is without prejudice to that Directive and excludes from its scope all documents "*for which third parties hold intellectual property rights*".²⁴⁰ It appears that this has been frequently relied upon by public bodies to exclude applicability of the PSI Directive to their information.²⁴¹ A concern exists among stakeholders that in this way, public bodies are able to circumvent the rules of the PSI Directive even where the data is perhaps not actually covered by any intellectual property right.²⁴²

Proposal for a revised PSI Directive

On 25 April 2018, the European Commission presented a proposal for revision of the PSI Directive (the "**Recast Proposal**"). Political agreement on the text was reached on 22 January 2019 by the negotiators of the European Parliament, the Council of the EU and the

²³⁸ Consolidated PSI Directive, art 1(2)(cc)

²³⁹ Consolidated PSI Directive, art 11(1)

²⁴⁰ Consolidated PSI Directive, art 1(2)(b)

²⁴¹ European Commission, 'Consultation on PSI Directive Review, Synopsis Report' (European Commission 2018) 3 <<https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-revision-directive-reuse-public-sector-information>> accessed 18 October 2018

²⁴² Ibid 6-7

European Commission.²⁴³ The most fundamental change with respect to the existing version of the PSI Directive relates to the Recast Proposal's material scope of application, which is extended to data held by public undertakings. The Recast Proposal clarifies that an undertaking is considered 'public' if public sector bodies may exercise "*a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it*", regardless of whether that is a direct or an indirect influence. The only relevant criterion is therefore whether public sector bodies are able to exercise control over an undertaking.

While not all public undertakings are covered by the Recast Proposal, it does extend (among others) to (i) those active in the areas defined in Directive 2014/25/EU, which includes transport services and ports and airports; (ii) those acting as public service operators pursuant to Regulation 1370/2007/EC, which covers public passenger transport services by rail and by road; (iii) those acting as air carriers fulfilling public service obligations pursuant to Regulation 1008/2008/EC; and (iv) those acting as EU ship owners fulfilling public service obligations pursuant to Regulation 3577/92/EEC (the Maritime Cabotage Regulation).²⁴⁴ The Recast Proposal is thus to a large degree targeted at public undertakings in the transport sector at large.

The Recast Proposal does limit its scope of application by excluding information held by public undertakings that is produced outside the scope of the provision of services in the general interest as defined by law or other binding rules in the Member State concerned.²⁴⁵ It will thus be important to consider whether or not a public undertaking has produced the requested information in the context of the provision of services of general interest. The scope has been limited further in the text on which political agreement was found, and now also excludes data that are related to activities for which public undertakings are directly exposed to competition and are therefore exempt from procurement rules.²⁴⁶

However, even where the revised Directive would be applicable and except where otherwise required under applicable law, the public undertaking in question could still decide whether or not to disclose the information as no mandatory information sharing obligation has been introduced (thus far). In this sense, the obligations imposed on public undertakings would be similar to those imposed on public entities under the regime of the initial version of the PSI Directive. The regime is optional, but as soon as a public undertaking decides to make information available, it will have to respect the rules laid down in the Directive. One can wonder what the consequences will be of introducing such regime that, admittedly, is optional but has been paired with strict modalities. Public undertakings may have concerns about the compliance burden that these strict modalities would entail and therefore choose not to disclose any data as a result. This has been mitigated to some extent, as certain (mainly procedural) requirements on the processing of re-use requests were not made applicable to public undertakings.²⁴⁷

Another novelty in the Recast Proposal is the introduction of the category of so-called "high-value datasets". These are datasets associated with important socio-economic benefits, the re-use of which should in principle be free of charge. The Annex of the Recast Proposal includes "mobility" as one of the thematic categories for high-value datasets. The datasets themselves are however not defined in the Recast Proposal itself, but would be adopted by the European Commission through a combination of Delegated Acts and Implementing Acts.²⁴⁸ Public undertakings fear that such future Delegated Acts could force them to make high-value datasets available for free and would thereby significantly affect their competitive position on the market, as they could be put in an inferior position compared to private undertakings operating on the same markets, upon which no such obligations would be imposed. This could hinder ongoing innovation in public service undertakings by increasing the risk of investing in own datasets and collaborating with start-ups and thus taking away the incentive for public undertakings to carry out such activities.²⁴⁹ This fear has been mitigated to

²⁴³ European Commission, 'Proposal for a Revision of the Public Sector Information (PSI) Directive' <<https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive>>, accessed 14 February 2019

²⁴⁴ Recast Proposal, art 1(1)(b)

²⁴⁵ Recast Proposal, art 2(1)(a)

²⁴⁶ The text refers to the exemption from procurement rules in accordance with Article 34 of Directive 2014/25/EU. See <<https://data.consilium.europa.eu/doc/document/ST-5635-2019-INIT/en/pdf>>, accessed 19 February 2019

²⁴⁷ Recast Proposal, art 4

²⁴⁸ Recast Proposal, art 13 and Council of the European Union, 'Interinstitutional file: 2018/0111(COD)' 55 <<https://data.consilium.europa.eu/doc/document/ST-5635-2019-INIT/en/pdf>> accessed 19 February 2019

²⁴⁹ Valeria Ronzitti, 'European Commission Proposal for a Review of the PSI Directive Risks Hindering Innovation and Investments in Public Services' (CEEP, 26 April 2018) <<https://www.ceep.eu/the-proposal-for-a-revised-psi-directive->

some extent in the text of 22 January 2019, which expressly excludes the requirement to make such high-value datasets available for free in case this would lead to a distortion of competition in the relevant market.²⁵⁰

The Recast Proposal further introduces various smaller changes. It contains provisions aimed at facilitating the re-use of dynamic data (e.g. real-time traffic information), such as the obligation to proactively make such data available via a suitable Application Programming Interface (API).²⁵¹ The text also clarifies that costs related to data anonymisation²⁵² may be included in the fees charged to re-users.²⁵³



Illustration in the transport sector:

In 2015, the German railway and infrastructure operator Deutsche Bahn, a public undertaking, organised its second Hackathon. Deutsche Bahn has an open data portal, and organised the contest under the motto “we provide the data, you innovate with it”. In 24 hours, the winning team managed to achieve very promising results through the evaluation of large amounts of data from infrastructure-related delays. More specifically, they enabled Deutsche Bahn to identify improvement potential for infrastructure by assessing whether problems are more often caused by concrete or by wooden sleepers and by indicating places with increased track position errors. Although Deutsche Bahn, as a public undertaking, was not (yet) under any obligation to make its data available, this is a clear example of the value that can be created by doing so.²⁵⁴

Limits to the desirability of opening up PSI: the case of essential services and critical infrastructure

The evolution of the PSI Directive since 2003 shows a continuous broadening of its scope. That trend is continued with the Recast Proposal which aims to

include public undertakings. Taking into account the potential benefits of opening up data, it seems that this broadened scope can only be applauded. There can however be some limits to the desirability of making available public sector data, which we will illustrate here through the example of essential services and critical infrastructure.

As explained in the fourth Chapter (Cyber-)security, the NIS Directive requires Member States to identify so-called operators of essential services (OESs). The latter are services that a Member State deems essential for the “*maintenance of critical societal and economic activities*”.²⁵⁵ Such operators must among others be identified for all major modes of transportation, notably air, rail, water, and road. Not unimportantly, the NIS Directive makes no distinction between public or private entities and thus impacts both public and private operators in the transport sector.

Furthermore, Directive 2008/114/EC²⁵⁶ (hereafter the “**Critical Infrastructure Directive**”) is concerned with the identification and designation of European critical infrastructures. These are assets, systems or parts thereof located in Member States that are essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would significantly impact the Member State concerned.²⁵⁷ Similarly to the NIS Directive, security requirements are introduced for such European critical infrastructures. Member States must among others ensure that operators and/or owners of such infrastructures develop security plans to ensure the infrastructure’s protection.

Many operators in the transport sector either provide essential services within the meaning of the NIS Directive or operate a critical infrastructure within the meaning of the Critical Infrastructure Directive. In the transport sector, many essential services operators are public undertakings. The essential services covered by the NIS Directive are moreover likely to constitute services provided in the general interest. This would mean that, under the Recast Proposal, the PSI regime would cover those services offered by essential services providers.

[risks-hindering-innovation-and-investments-in-public-services/>](#) accessed 18 October 2018

²⁵⁰ Council of the European Union, 'Interinstitutional file: 2018/0111(COD)' 56 <<https://data.consilium.europa.eu/doc/document/ST-5635-2019-INIT/en/pdf>> accessed 19 February 2019

²⁵¹ Recast Proposal, art 5(4)

²⁵² For more information on data anonymisation, we refer to our third Chapter Anonymisation/ pseudonymisation.

²⁵³ Recast Proposal, art 6(1)

²⁵⁴ Philipp Drieger, 'All aboard with Infrastructure 4.0 – Splunk wins Deutsche Bahn Internet of Things Hackathon' (*Splunk*) <<https://www.splunk.com/blog/2015/06/08/splunk-team-wins-db-infrastructure-data-challenge-in-24h-iot-hackathon.html#>> accessed 18 October 2018

²⁵⁵ NIS Directive, Recital 20

²⁵⁶ Council Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection [2008] OJ L 345/75

²⁵⁷ Critical Infrastructure Directive, art 2(a)

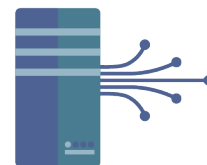
There is however an inherent tension between the Recast Proposal's aim to make public data more accessible and to encourage the re-use of this information, and the aim of the NIS Directive to ensure security and continuity of those services that are essential for the maintenance of critical societal and economic activities. A certain amount of data gathered and generated through the provision of essential services will necessarily be of a sensitive nature. Making this sensitive data accessible to the public would inherently entail risks for the security and continuity of the service. The same reasoning applies to operators of critical infrastructures under the Critical Infrastructure Directive. This clearly shows that, while open data policies are for the most part beneficial to society, these policies should not be pursued thoughtlessly and certain sensitivities should be taken into account in current and subsequent revisions of the PSI Directive.

Conclusion

The Open Data movement and governments around the world, including the EU, are committed to making data, and more particularly 'government data' or public sector information, publicly available and usable. The EU institutions have taken both legislative and non-legislative measures to encourage the uptake of open data, most notably through the PSI Directive which attempts to remove barriers to the re-use of PSI throughout the EU. Still, open data regimes also encounter a number of challenges – on a technical, economic and legal level – that cannot be ignored. The proposal for a recast of the PSI Directive aims to address some of these concerns. It introduces one major change by expanding the Directive's scope to include public undertakings. While information sharing has not been made mandatory for public undertakings (yet), the new regime still constitutes a significant development for the transport sector, where services are often provided by public undertakings.



Sharing obligations



In this eleventh Chapter, we will focus on legal instruments imposing data sharing obligations on private undertakings.

The previous (and next) Chapters offer a good overview of the most common legal challenges encountered by private companies trying to share data with or access and use data from other companies. Barriers to private sector data sharing are however not only of a legal nature. Many commercial and technical barriers also come into play. The EU legislators have therefore adopted instruments that impose data sharing and which may impact a company's control of, access to, or use of data. Such legislations are usually sector-focused and provide for an array of rights and obligations in relation to specific types of data in particular circumstances.²⁵⁸ While this Chapter in no way provides an exhaustive list, it attempts to offer a succinct examination of those pieces of legislation imposing data sharing obligations that are most relevant to the transport sector.

Intelligent Transport Systems

The advent of ITS has shown a proliferation of legislative instruments imposing data sharing obligations on private actors, among others for safety purposes and to provide transparent information to end-users. In 2010, a legal framework was adopted to foster the coordinated deployment of ITS in Europe. Directive 2010/40/EU aimed to establish interoperable and seamless ITS systems across the EU, while leaving it up to the Member States to decide which systems to invest in. The Directive moreover empowered the European Commission to lay down a range of specifications for ITS systems, in the form of delegated acts. Many of these contain data sharing obligations, as will be addressed briefly below.²⁵⁹

²⁵⁸ Commission, 'Impact assessment accompanying the document Proposal for a Regulation of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union' (Staff Working Document) SWD(2017) 2 final, 21

²⁵⁹ European Commission, 'Intelligent Transport Systems: Action Plan and Directive' (European Commission, 2018)

While the delegated regulations adopted pursuant to the ITS Directive focus on road transport, ITS are not limited to that mode of transport alone. We may therefore expect future regulation in this respect for rail, air, and maritime and inland waterways transportation as well. Another notable evolution is the increased adoption of technical specifications and standards for information sharing in the various modes of transport. Technical specifications have for instance been adopted for information exchange both in the domain of passenger rail services and in the domain of rail freight services.²⁶⁰ This is in part due to the fact that ITS entail pressing interoperability issues, which increase the need to adopt such technical specifications. We can therefore expect more technical specifications to be adopted in the future, which might in turn entail additional data sharing obligations.

Overview of data sharing obligations in the transport sector

This section offers a very succinct overview of the most relevant legislative instruments imposing data sharing obligations in the transport sector.

- Commission Delegated Regulation with regard to the provision of EU-wide multimodal travel information services²⁶¹: In order to achieve its goal of providing seamless Union-wide multimodal travel information services, the Delegated Regulation introduces a number of

<https://ec.europa.eu/transport/themes/its/road/action_plan_en> accessed 18 October 2018

²⁶⁰ European Union Agency for Railways, 'Telematic Applications for Freight (TAF), Telematic Applications for Passengers (TAP)' (European Union Agency for Railways 2017) <<https://www.transportstyrelsen.se/globalassets/global/jarnvag/branschradet/taftap/era-kresimir-raguz-stefan-jugelt2.pdf>> accessed 18 October 2018

²⁶¹ Commission Delegated Regulation (EU) 2017/1926 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services [2017] OJ L 272/1

obligations to facilitate the exchange and reuse of data. Notably, all transport operators, infrastructure managers and on-demand service providers – both private and public – will have to provide travel and traffic data about the relevant mode of transport to a centralised national access point for such data.²⁶² The data cannot simply be supplied *as is*, but certain conditions will have to be fulfilled.

- **e-Call Delegated Regulation²⁶³**: This lays down specifications for the location – operated either by a public authority or by a private organisation recognised by the Member State – where ITS systems emergency calls are first received, the so-called public safety answering point (PSAP). It is determined that this point must have access to an appropriate geographical information system, allowing it to identify position and heading of the vehicle. This information must in turn enable the PSAP operator to provide the location and certain other data to the appropriate emergency service or service partner.
- **e-Call Regulation²⁶⁴**: This instrument requires vehicle manufacturers to ensure that a vehicle's precise location, its identification, the time of incident and the direction of travel are transmitted to emergency services in case of a serious accident.²⁶⁵
- **Delegated Regulation on road safety-related minimum universal traffic information²⁶⁶**: This imposes on both public and private road operators and/or service providers an obligation to detect and identify events and conditions and to collect the relevant road safety-related traffic data. The latter must then be shared and exchanged through a national access point, where it will be accessible for reuse.
- **Commission Delegated Regulation with regard to the provision of information services for safe and**

secure parking places for trucks and commercial vehicles²⁶⁷: The objective of this Delegated Regulation is to optimise the use of parking places and to facilitate drivers' or transport companies' decisions about when and where to park through the deployment of information services. To this end, both static and dynamic data on safe and secure parking areas must be collected by all public and private parking operators and service providers and be supplied in standardised machine-readable formats to a national access point.

- **Delegated Regulation with regard to the provision of EU-wide real-time traffic information services²⁶⁸**: This instrument seeks to provide appropriate framework conditions enabling the co-operation of road authorities, road operators and any other ITS service providers involved in the traffic information value chain, and to support the interoperability, compatibility, and continuity of real-time traffic information services across Europe. Road authorities and road operators collecting certain road data must provide this in a standardised format, if available, or in any other machine-readable format to a national access point
- **Directive establishing an Infrastructure for Spatial Information in the European Community (the "INSPIRE Directive")²⁶⁹**: The INSPIRE Directive lays down rules to set up an infrastructure for spatial information, which is information directly or indirectly referencing a specific location or geographical area and includes information related to transport networks, for the purpose of EU environmental policies. While the Directive is mainly aimed at public authorities, it recognises that certain relevant spatial datasets and services are held and operated by third parties. Therefore, private parties should also have the possibility of contributing to the national infrastructures, but this is made subject to certain conditions.²⁷⁰

²⁶² Travel Information Services Delegated Regulation, arts 4-5

²⁶³ Commission Delegated Regulation (EU) No 305/2013 of 26 November 2012 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the harmonised provision for an interoperable EU-wide eCall [2013] OJ L 91

²⁶⁴ Regulation (EU) 2015/758 of the European Parliament and of the Council concerning type-approval requirements for the deployment of the eCall in-vehicle system based on the 112 service and amending Directive 2007/46/EC [2015] OJ L 123/77

²⁶⁵ Claudiu-Dan Bărcă, Rareș Ropot and Sorin Dumitrescu, 'eCall – Minimum Set Of Data (MSD)' (2009) *Journal of Information Systems & Operations Management* 428, 429

²⁶⁶ Commission Delegated Regulation (EU) No 886/2013 of 15 May 2013 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to data and procedures for the provision, where possible, of road safety-related minimum universal traffic information free of charge to users [2013] OJ L 247

²⁶⁷ Commission Delegated Regulation (EU) No 885/2013 of 15 May 2013 supplementing ITS Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of information services for safe and secure parking places for trucks and commercial vehicles [2013] OJ L 247

²⁶⁸ Commission Delegated Regulation (EU) 2015/962 of 18 December 2014 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide real-time traffic information services [2015] OJ L 157

²⁶⁹ Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) [2007] OJ L 108

²⁷⁰ INSPIRE Directive, Recital 18

- **Advance Passenger Information Directive²⁷¹**: Air carriers must communicate information concerning passengers, and thus "personal data" to certain authorities for the purpose of combating illegal immigration.²⁷² This legislation has little to no impact from a commercial perspective, as the data is not made publicly available and competitors thus have no access to the collected and transmitted data.
- **Regulation on rail passengers' rights and obligations²⁷³**: This is primarily an instrument of consumer protection. Pursuant to this Regulation, railway undertakings must provide passengers with specific information related to their journeys, including time schedules and conditions for the fastest trip as well as the lowest fares, information on accessibility and access conditions for bicycles and disabled persons and any activities that are expected to disrupt or delay the services. Ticket vendors offering transport contracts on behalf of railway undertakings are under the same obligation.²⁷⁴ Railway undertakings must additionally provide a limited amount of information during the journey.
- **Vehicle Emissions Regulation²⁷⁵**: This Regulation²⁷⁶ not only regulates vehicle emissions for small passenger and commercial motor vehicles, but also lays down rules on accessibility of vehicle repair and maintenance information ("**RMI**").²⁷⁷ It imposes an obligation on EU car manufacturers to provide unrestricted and standardised access to vehicle RMI. Access must be given through websites using a standardised format in a readily accessible and prompt manner. Manufacturers are not allowed to discriminate against independent operators involved in the repair and maintenance of motor

vehicles, which are often SMEs.²⁷⁸ Therefore, when a consumer buys a certain vehicle, the manufacturer cannot lock out independent repair workshops and make that person visit an approved workshop to get repair and maintenance. Notwithstanding the obligation to grant access to RMI, manufacturers are entitled to charge "reasonable fees" for this service.²⁷⁹

- **Car Labelling Directive²⁸⁰**: This aims to help consumers choose vehicles with low fuel consumption by requiring dealers in new passenger cars to provide potential buyers with useful information on these vehicles' fuel consumption and CO₂ emissions. This information must be displayed on the car's label, on posters and other promotion material, and in specific guides.
- **Vessel Traffic Monitoring Directive²⁸¹**: This Directive was adopted to help prevent accidents and pollution at sea and to increase the efficiency of maritime traffic. It introduces a number of information sharing obligations on certain categories of ships, which must, among others, be fitted with an automatic identification system ("**AIS**"). The Directive also requires any operator, agent or master of a ship bound for an EU port to inform the relevant port authority within a certain time scale of certain information items, including ship identification, port of destination, estimated time of arrival and total number of persons on board. Certain mandatory ship reporting systems are also addressed.

²⁷¹ Council Directive 2004/82/EC of 29 April 2004 on the obligation of carriers to communicate passenger data [2004] OJ L 261

²⁷² Advance Passenger Information Directive, art 1

²⁷³ Regulation (EC) No 1371/2007 of the European Parliament and of the Council on rail passengers' rights and obligations [2007] OJ L 315/14

²⁷⁴ Rail Passengers' Rights and Obligations Regulation, art 8

²⁷⁵ Regulation (EC) No 715/2007 of the European Parliament and of the Council of 20 June 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information [2007] OJ L 171

²⁷⁶ We note that similar requirements for heavy duty vehicles were laid down in Regulation 595/2009/EC. The information exchange requirements in that Regulation and in the Vehicle Emissions Regulation have now been consolidated and updated in Regulation 2018/858/EU on the approval and market surveillance of motor vehicles, which will apply as of 1 September 2020.

²⁷⁷ Vehicle RMI is information required for diagnosing, servicing and repairing a vehicle provided by a manufacturer to its authorised dealers and repair centres, including any amendments and supplements to such information.

²⁷⁸ Vehicle Emissions Regulation, art 6; Commission, 'Report from the Commission to the European Parliament and The Council on the operation of the system of access to vehicle repair and maintenance information established by Regulation (EC) No 715/2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information' COM(2016) 782 final, 5

²⁷⁹ Vehicle Emissions Regulation, art 7

²⁸⁰ Directive 1999/94/EC of the European Parliament and of the Council of 13 December 1999 relating to the availability of consumer information on fuel economy and CO₂ emissions in respect of the marketing of new passenger cars [2000] OJ L 12

²⁸¹ Directive 2002/59/EC of the European Parliament and of the Council of 27 June 2002 establishing a Community vessel traffic monitoring and information system and repealing Council Directive 93/75/EEC [2002] OJ L 208



Illustration in the transport sector:

The European Union Location Framework ("EULF") Transportation Pilot was designed to improve the dissemination of updated road safety information between road authorities and private sector map providers across borders. One of the pilot's aims was moreover to test the feasibility of reusing spatial data collected and disseminated on the basis of the INSPIRE Directive within the ITS community. To this end they created a pan-European platform and web service to provide up-to-date, authoritative, interoperable, cross-border, reference geo-information for use by EU public and private sectors and compliant with the INSPIRE Directive. It was found that the INSPIRE transport network data was an important source of data when national road databases are not available.²⁸²

General observations

The data sharing obligations that follow from the above legal instruments vary based on a number of factors, including reasons of public interest that have led to the adoption of the legislative instrument, such as for instance enhancing road safety or facilitating Union-wide interoperability for particular services. Furthermore, while creating increased consumer transparency is an objective of many of the examined data sharing obligations in the transport sector, some also include mechanisms to protect and limit the disclosure of certain types of data, such as commercially confidential information.

In terms of remuneration, a distinction can be observed between situations where data must be provided to public authorities only and those where the data is to be shared to a wider community including private stakeholders. When the legislation only imposes data sharing to authorities, it should usually be provided free of charge. Where such data sharing must however be extended to include private actors, undertakings are typically allowed to demand some kind of remuneration. A similar distinction applies depending on the nature of the purpose pursued. If an instrument mainly concerns data sharing for public safety purposes or other purposes of public interest, no remuneration for the mandatory data sharing is included. However, where data sharing obligations are imposed in

²⁸² Maria Teresa Borzacchiello, Raymond Boguslawski, Francesco Pignatelli, 'JRC Technical Reports: Improving Accuracy in Road Safety Data Exchange for Navigation Systems - EU Location Framework Transportation Pilot' (European Commission 2016) <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC104569/jrc104569_d%2021%20tp%20final%20report%20-%20v1.7%20pubsy.pdf> accessed 18 October 2018

order for innovative and competitive services to be developed on the basis thereof, the data provider may usually request at least a reasonable remuneration.

Interestingly, some of the more recent legislative instruments refer to the conditions for access and reuse imposed on public sector bodies in the PSI Directive.²⁸³ It would be useful to monitor future developments to know whether this is an approach that will be increasingly adopted with regard to private sector data sharing obligations. Another emerging trend is the requirement for information sharing to be done through a centralised access point.

Other data sharing obligations

Unfair Contract Terms and Unfair Commercial Practices

To a limited extent, data sharing obligations may also arise under the legislation relating to unfair contract terms²⁸⁴ and unfair commercial practices²⁸⁵ when a data-holding company is preventing access to data in a particularly unfair manner.

The Unfair Commercial Practices Directive protects consumers against misleading acts or omissions from a trader. The latter is for instance under an obligation to inform consumers if any data supplied by them to access the trader's service will be used for commercial purposes. Not providing such information may be considered a misleading omission of material information, prohibited under the directive.

The Unfair Contract Terms Directive seeks to protect consumers from unfair standard terms in consumer agreements by stipulating minimum rules in this respect. Its scope is broad enough to cover standard terms on the treatment and analysis of data. The Directive's main principle is that standard contract terms are considered unfair if, to the consumer's detriment and against good faith principles, they cause a significant imbalance in the

²⁸³ On the PSI Directive and the various legal issues and opportunities that are encountered when using open data for big data technologies, please read our tenth Chapter Open data.

²⁸⁴ Council Directive 93/13/EEC on unfair terms in consumer contracts [1993] OJ L 95/29

²⁸⁵ Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council [2005] OJ L 149/22 ("Unfair Commercial Practices Directive")

respective rights and obligations of the contracting parties. While this legislation is in principle applicable only to contracts in a B2C relationship, some Member States apply it (or its principles) to B2B relations as well.²⁸⁶ A drawback however is the fact that the indicative list of unfair contract terms annexed to the Directive does not reflect any of the challenges of a modern data economy.²⁸⁷

Platform-to-Business Transparency

On 26 April 2018, the European Commission published a proposal for a Regulation on promoting fairness and transparency for business users of online intermediation services (the "**Platform-to-Business Regulation**").²⁸⁸ Inter-institutional agreement was reached on the proposed Regulation on 13 February 2019. The Regulation aims to create a fair, transparent and predictable business environment for smaller businesses and traders when using online platforms.

The Regulation will apply to online platform intermediaries and online search engines providing services to businesses that are established in the EU and that offer goods or services to consumers located in the EU.²⁸⁹

Online platform intermediaries include:

- Third-party e-commerce market places (e.g. Amazon, eBay, etc.);
- App stores (Google Play, Microsoft Store, etc.);
- Social media for business (e.g. Facebook pages, etc.); and
- Price comparison tools (e.g. Skyscanner, etc.)

Online search engines in scope of the Regulation are those services that allow users to perform web searches on the basis of a query on a subject and return links corresponding with that search request.²⁹⁰

The Platform-to-Business Regulation may have an impact in respect of data sharing obligations, as it would *inter alia* require online intermediation services providers to:

- ensure that their terms and conditions aimed at professional users are both easily understandable and available;²⁹¹ and
- include in their terms and conditions a description of what data provided for or generated through their services can be accessed, by whom, and under which conditions.²⁹²

In addition, both online platform intermediaries and online search engines would be required to list the main parameters (such as characteristics of the goods and services, relevance of those characteristics for consumers, and website design characteristics) determining how goods and services are ranked in search results.²⁹³ The Regulation however provides that such obligation should not require online intermediation services or online search engines to disclose any of their trade secrets.

Competition law

When businesses wish to access and use a particular dataset generated and/or held by another economic operator, they usually attempt to enter into negotiations with the aim of concluding an agreement. Such negotiations will not always succeed however, particularly if the data-holding company does not see sufficient economic interest in granting the other party access. That party could then, under certain circumstances, invoke general competition law to gain wider access to the data. It should be stressed however that a refusal to grant access does not of itself sufficiently justify intervention through competition law. Refusal is not illegitimate where a company's exclusive control over and access to data provides it with a competitive advantage and thereby creates the necessary incentive to invest in data-driven business models.²⁹⁴

Striking the right balance between access to and legitimate control of data is thus a delicate task. The Court of Justice of the EU in its case law developed four conditions that must be fulfilled before an obligation to license the use of privately-held commercial information is imposed. These include the requirements that: (i) the data is absolutely necessary for the downstream product; (ii) there would be no actual competition between the upstream and the downstream product; (iii) refusal

²⁸⁶ SWD(2017) 2 final 21

²⁸⁷ Josef Drexler, 'Designing Competitive Markets for Industrial Data in Europe – Between Propertisation and Access' (2017) 8 JIPITEC 257 <<https://www.jipitec.eu/issues/jipitec-8-4-2017/4636>> accessed 18 October 2018

²⁸⁸ Commission, 'Proposal for a regulation of the European Parliament and of the Council on promoting fairness and transparency for business users of online intermediation services' COM(2018) 238 final

²⁸⁹ Proposed Platform-to-Business Regulation, art 1

²⁹⁰ Ibid art 2(5)

²⁹¹ Ibid art 3

²⁹² Ibid art 7

²⁹³ Ibid art 5

²⁹⁴ Josef Drexler, 'Designing Competitive Markets for Industrial Data in Europe – Between Propertisation and Access' (2017) 8 JIPITEC 257 <<https://www.jipitec.eu/issues/jipitec-8-4-2017/4636>> accessed 18 October 2018

would prevent the second product from being developed at all; and (iv) the refusal cannot be justified by objective reasons.²⁹⁵

It should moreover be noted that, while competition law allows enforcers to ban existing and identifiable anti-competitive conduct of data-rich businesses, they are not well equipped for regulating markets *ex ante*.²⁹⁶ It often takes years to achieve results from actions based on competition law. This is a major drawback for private companies seeking to gain access to datasets for their business today.

For a further analysis of the impact on competition rules on (big) data, we refer to the last Chapter Competition.

Public tendering

An entirely different way of imposing data sharing obligations is by including them as conditions in public tenders. This possibility was suggested by the SPICE (Support Procurements for Innovative transport and mobility solutions in City Environment) Project in the context of public authorities contemplating procurement of Mobility as a Service ("Maas") schemes. Recognising the fact that open data is essential to Maas development, they entertained the possibility of using public procurement to encourage open data (from private actors) by setting data sharing obligations in public tenders. The creation of an open interface (API) and open platform by the private company chosen for the tender could encourage start-ups and SMEs to develop innovative services.²⁹⁷

Conclusion

While private companies often generate huge amounts of data, they are not always prepared to voluntarily share this data outside the company. This is due to the large number of legal, commercial and technical challenges associated with private sector data sharing. In certain circumstances, private companies are therefore legally required to share their data.

Our analysis of the body of legislation specific to the transport sector shows that data sharing obligations are increasingly adopted in the context of ITS. In the framework of the ITS Directive, numerous data sharing obligations were established, mostly in the domain of road transportation. In general, data sharing obligations appear to vary based on a number of factors, including the reasons of public interest that have led to the adoption of the instrument, such as for instance enhancing road safety or facilitating Union-wide interoperability for particular service.

Overall, a clear increase can be observed in legislation imposing data sharing obligations, which can be linked to the development of ITS. In this respect, the European Commission should carefully consider whether the imposition of such general data sharing obligations is in each case equally necessary. An alternative that may be less burdensome but that could perhaps generate useful results could be to stimulate data sharing by including data sharing obligations in public tenders.

²⁹⁵ SWD(2017) 2 final, 22. Also check: Bertin Martens, 'JRC Technical Reports: An Economic Policy Perspective on Online Platforms' (Institute for Prospective Technological Studies Digital Economy Working Paper 2016/05, European Commission 2016) 41 <<https://ec.europa.eu/jrc/sites/jrcsh/files/JRC101501.pdf>> accessed 18 October 2018; Josef Drexl, 'Designing Competitive Markets for Industrial Data in Europe – Between Propertisation and Access' (2017) 8 JIPITEC 257 <<https://www.jipitec.eu/issues/jipitec-8-4-2017/4636>> accessed 18 October 2018

²⁹⁶ Josef Drexl, 'Designing Competitive Markets for Industrial Data in Europe – Between Propertisation and Access' (2017) 8 JIPITEC 257 <<https://www.jipitec.eu/issues/jipitec-8-4-2017/4636>> accessed 18 October 2018

²⁹⁷ Eva Buchinger and others, 'D3 SPICE Analysis and Recommendations. Version Final 29/08 2018' (SPICE 2018) 18 <<http://spice-project.eu/wp-content/uploads/sites/14/2018/08/SPICE-D3-Analysis-and-Recommendations-FINAL.pdf>> accessed 18 October 2018



Data ownership

In this Twelfth Chapter, we take a closer look at the data ownership issues surrounding the (big) data value chain and examine how such issues are addressed at EU and Member State level. Where relevant, illustrations from the transport sector will be provided.

The European Commission has voiced on multiple occasions the most important legal issues in a data environment. In its data-driven economy Communication of July 2014 already, but also in the context of its 2016 free flow of data initiative, it highlighted that *"barriers to the free flow of data are caused by the legal uncertainty surrounding the emerging issues on 'data ownership' or control, (re)usability and access to/transfer of data and liability arising from the use of data"*.²⁹⁸

Indeed, if they cannot rely on any of the other exclusive rights discussed in this publication (see for instance our Chapter Intellectual property rights), stakeholders in the (big) data analytics lifecycle increasingly try to claim "ownership" in (parts of) the datasets used in the analytics.

The "ownership" concept

There is often some kind of misunderstanding between legal practitioners and non-legal professionals on the meaning of the term "ownership".

Following the Oxford Dictionary of Law, the word "ownership" has the following meaning: *"it is the exclusive right to use, possess, and dispose of property, subject only to the rights of persons having a superior interest and to any restrictions on the owner's rights imposed by agreement with or by act of third parties, or by operation of law."*²⁹⁹ It is therefore something that implies

certain rights over a property such as being able to enjoy, use, sell, rent, give away, or even destroy an item of property. Ownership may be corporeal (i.e. title to a tangible/material (im)movable object) or incorporeal (i.e. title to an intangible object, such as intellectual property, or a right to recover debt).

However, for businesses, the meaning of "ownership" may be different, especially in a data environment. It is often used to assign responsibility and accountability for specific databases, whereby reference to the "data owner" is made.³⁰⁰ In such particular context, 'ownership' does not have a legal connotation but refers to other concepts such as assurance of data quality and security. There is thus no transfer of or licence over a property as such.

In this Chapter, the term "ownership" will be used in its legal meaning. This nevertheless includes certain difficulties due to the particularities of data. Indeed, data is not like any other tangible or intangible "thing". It has certain characteristics often put forth when discussing the data economy, such as the fact that data is limitless and non-rivalrous, that fit uneasily with the legal concept of "ownership".

Actors in the data value chain who could claim ownership in data

The issue of data ownership is even more complicated by the data value cycle which can be rather complex and involves numerous stakeholders. This increases the difficulties in determining who could or would be entitled to

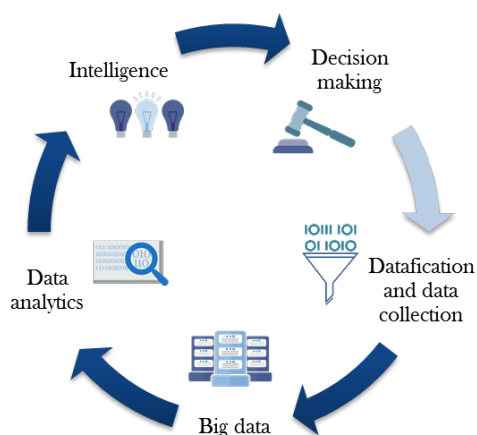
²⁹⁸ COM(2014) 442 final; European Commission, 'European Free Flow of Data Initiative within the Digital Single Market' (Inception impact assessment, European Commission 2016) <http://ec.europa.eu/smart-regulation/roadmaps/docs/2016_cnect_001_free_flow_data_en.pdf> accessed 21 February 2019

²⁹⁹ Jonathan Law and Elizabeth A. Martin, *A Dictionary of Law* (7th edition, Oxford University Press 2014) <<http://www.oxfordreference.com/view/10.1093/acref/978019>

[9551248.001.0001/acref-9780199551248-e-2745?rskey=2MFh2r&result=2900](http://www.oxfordreference.com/view/10.1093/acref/9780199551248-e-2745?rskey=2MFh2r&result=2900)> accessed 21 February 2019
³⁰⁰ OECD, *Data-driven Innovation: Big Data for Growth and Well-being* (OECD Publishing 2015) 195

claim ownership in data. Many of such stakeholders may attempt claiming ownership in data because, for instance, they create or generate data, or because they use, compile, select, structure, re-format, enrich, analyse purchase of, take a licence on, or add value to the data. Accordingly, in many instances, different stakeholders will have different powers depending on their specific role. Hence, no single data stakeholder will have exclusive rights.³⁰¹

The following Figure created by the Organisation for Economic Co-operation and Development (OECD) aims to depict the data value cycle.³⁰²



Looking at the data value cycle, one can distinguish various actors and determine their roles in the data economy, in particular in the "datafication" process, the analysis of data, and the decision-making phase. It should however be kept in mind that certain organisations may play multiple roles. Also, the data value cycle does not reflect the cross-border flow of data and the legal intricacies related thereto.

There is a multitude of actors on the market actively reaping the benefits of the data economy. The relationships between such actors are an essential element of the data value cycle. Some of the most important actors are depicted in the layered Figure below, whereby the underlying layers supply the upper layers with goods and services³⁰³:

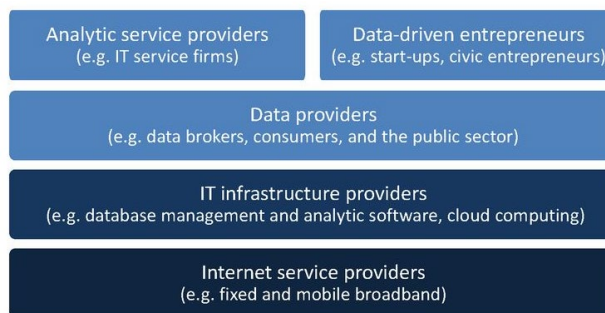


Illustration in the transport sector: The developments in relation to connected and autonomous vehicles have also raised questions with respect to data ownership.³⁰⁴ The on-board computing systems present in connected and autonomous vehicles will allow for the transfer of substantial amounts of information, including about the driver and its location. At the current stage, it is still unclear who will "own" this information among the many different actors involved; i.e. the driver who the personal data relates to, the owner of the vehicle (if different from the driver), the manufacturer of the vehicle, insurance companies, navigation service providers, the government, or any other third party. Any data ownership claim may have a far-reaching impact on the further implementation of the technology concerned. In any event, the personal data protection rules will need to be respected.

Legislation on data ownership

Our researches have not enabled us to identify any EU legislation that would specifically regulate the question of ownership in data. This being said, such absence of ownership-related legislation does not exclude the fact that there are numerous legislations that have an impact on data or that may confer some kind of protection to certain types of data or on datasets (i.e. copyright, database rights and trade secrets).

The same issues apply when looking at the situation at Member State level. There clearly is no specific data-related legislation that explicitly recognises ownership in data in the various Member States. Having said that, some countries have legislation in place allowing to control the flow of data. One example would be France, where the civil code sets

³⁰¹ Ibid

³⁰² Ibid 33

³⁰³ Ibid 72

³⁰⁴ Caitlin A. Surakitbanharn and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles (CATV)' (https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf) accessed 18 October 2018

out mechanisms (based on both civil and criminal law measures) enabling the holder of data to prevent or restrain the misuse of data.

Case law addressing the issues of "ownership" of data

Thus far, there has been no real EU or national jurisprudence satisfactorily dealing with the issues surrounding data ownership. Nevertheless, some decisions at EU and national level may give an indication on how these issues may be dealt with in the future:



- **EU:** According to some authors, the CJEU opened the door for a discussion on ownership in intangible assets in its *UsedSoft* judgment issued on 3 July 2012.³⁰⁵ In this ruling, the Court held that the commercial distribution of software via a download on the Internet is not only based on a licence, but on a sale of goods.³⁰⁶ Therefore, the owner of copyright in software cannot prevent a perpetual "licensee" from selling his software (understood as downloaded file). The decision implies that there is a specific ownership attributed to intangible goods like software downloaded via the Internet. Applicability of this model to other digital goods remains to be considered in future court decisions.



- **Germany:** In a case concerning the destruction of data, the Higher Regional Court of Karlsruhe considered that deletion of data stored on a data carrier may violate the ownership in the data carrier under the German Civil Law Code, extending the protection of the ownership in the data carrier to data stored on it.³⁰⁷ Later decisions of German courts opposed the possibility to hold ownership over data as such, since data lacks the necessary material character³⁰⁸ and since it is not considered a 'thing' under the German Civil Law Code.³⁰⁹ The Court of Appeal of Nuremberg³¹⁰ has built on the general principle adopted in Germany, according to which things that are neither rights nor goods may nevertheless be sold

within a sale contract (Section 453 of the German Civil Act). To decide whether former employees were allowed to delete the data stored on their company-owned laptops, the Nuremberg Court made reference to the theory of the so-called "Skripturakt". According to this theory, the person who generates the data gets the right to the data, even if the data afterwards are used for the business or for the sake of the employer. In consequence, under criminal law and in this particular case, the employees were allowed to delete the data.³¹¹



- **United Kingdom:** So far, the UK courts held that data is not property and therefore cannot be stolen³¹², that data are not eligible to be the subject of a common law lien³¹³, and that there is no proprietary right in the content of an email.³¹⁴



- **France:** The French Supreme Court ("*Cour de cassation*") rendered a ruling³¹⁵ in 2015 that could open a way to recognising the ownership of "data". The Court found that (remotely) downloading computer data without taking away their support may amount to the offence of theft, acknowledging therefore indirectly that such independent data may be owned.

³⁰⁵ Judgment of 3 July 2012, *UsedSoft*, C-128/11, ECLI:EU:C:2012:407

³⁰⁶ Thomas Hoeren, 'Big Data and the Ownership in Data: Recent Developments in Europe' (2014) 36(12) EIPR 751

³⁰⁷ OLG Karlsruhe, Urt. v. 07.11.1995 – 3 U 15/95 – *Haftung für Zerstörung von Computerdaten*

³⁰⁸ LG Konstanz, Urt. v. 10.05.1996 – 1 S 292/95 = NJW 1996,2662

³⁰⁹ OLG Dresden, Beschl. v. 05.09.2012 – 4 W 961/12 = ZD 2013,232

³¹⁰ OLG Nürnberg 1. Strafsenat decision of 23.01.2013, 1 Ws 445/12

³¹¹ One should bear in mind that the discussed case had a strong criminal law connotation; the employees who deleted the data without prior authorisation were accused of theft, with their employer asking for a conviction under Section 303(a) of the German Criminal Act (prohibiting unlawfully erasing, corrupting or altering computer data under penalty of imprisonment). It is unclear whether the same rule would be applied by German courts in a civil law matter; OLG Nürnberg 1. Strafsenat decision of 23.01.2013, 1 Ws 445/12, par. 14

³¹² *Oxford v Moss* [1979] 68 Cr App Rep 183

³¹³ *Your Response v Datateam Business Media* [2014] EWCA Civ 281

³¹⁴ *Fairstar Heavy Industries v Adkin*, [2013] EWCA Civ 886

³¹⁵ May 20, 2015 (No14-81336)

Commission Communications having an impact on the Data Ownership Debate

"Towards a Thriving Data-Driven Economy" (2014)

The 2014 Commission Communication entitled "Towards a thriving data-driven economy" expected the big data market to grow worldwide to USD 16.9 billion in 2015 at an annual rate of 40%. The Commission nonetheless also indicated that the EU had been slow in embracing this revolution and that the complexity of the legal environment and the insufficient access to large datasets created entry barriers to SMEs and stifled innovation.

The 2014 Communication addressed the various challenges by sketching the features of the European data-driven economy of the future and drawing some conclusions to support and speed up the transition towards it. It notably concluded that to be able to seize the opportunities related to a data-driven economy and to compete globally in such economy, the EU must "make sure that the relevant legal framework and the policies, such as on interoperability, data protection, security and IPR are data-friendly, leading to more regulatory certainty for business and creating consumer trust in data technologies".³¹⁶

In a section dedicated to the regulatory issues, the Communication further highlighted the issues related to personal data protection and consumer protection, data mining, and security. It also raised the concerns pertaining to the ownership and liability of data provision and data location requirements in various sectors that limit the flow of data.

"A Digital Single Market Strategy for Europe" (2015)

In its 2015 Staff Working Document related to the Digital Single Market, the Commission reiterated the legal issues by putting forth problem drivers related to the data economy: "currently, collecting, processing, accessing and protecting data is a major challenge. This includes issues such as ownership of data, treatment of personal and industrial data, availability, access and re-use, contractual terms and conditions, data security, quality of data (e.g. timely updates), authentication of users, cybercrime, acceptance of electronic documents,

liability for incorrect information, standardisation of languages and formats."³¹⁷

"Building a European Data Economy" (2017)

The EU Commission carefully examined the most topical issues related to data in its Communication on "Building a European Data Economy" and the associated Staff Working Document.³¹⁸

With respect to the particular issue of data access, we note in particular the EU Commission's conclusion according to which "*comprehensive policy frameworks do not currently exist at national or Union level in relation to raw machine-generated data which does not qualify as personal data, or to the conditions of their economic exploitation and tradability. The issue is largely left to contractual solutions.*"³¹⁹ In the same vein, the EU Commission also concludes that "*where the negotiation power of the different market participants is unequal, market-based solutions alone might not be sufficient to ensure fair and innovation-friendly results, facilitate easy access for new market entrants and avoid lock-in situations.*"³²⁰

Finally, the Communication suggests several non-exhaustive and not mutually exclusive possibilities³²¹, to be discussed with stakeholders, to move forward on the issue of access to machine-generated data. Some suggested measures are non-legislative and consist of (i) the creation of guidance on incentivising businesses to share data; (ii) fostering the development of technical solutions for reliable identification and exchange of data; and (iii) the creation of model contract terms. Other suggested measures are of a legislative nature and amount to (i) the creation of default contract rules; (ii) providing access to commercially-held data to public sector bodies for public interest and scientific purposes; (iii) granting a right to use and authorise the use of non-personal data to the "data producer"; and (iv) the creation of a legal framework governing access to data against remuneration.

³¹⁷ Commission, 'A Digital Single Market Strategy for Europe – Analysis and Evidence Accompanying the document Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A Digital Single Market Strategy for Europe' (Staff Working Document) SWD(2015) 100 final

³¹⁸ COM(2017) 9 final, 4

³¹⁹ Ibid 10. See also the summary of the findings in relation to the EU law regime applicable to processing data in SWD(2017) 2 final, 22.

³²⁰ Ibid

³²¹ Such possibilities are detailed in the Commission Staff Working Document SWD(2017) 2 final, 30 ff

³¹⁶ COM(2014) 442 final

"Towards a Common European Data Space" (2018)

In its Communication entitled "Towards a common European data space", the Commission proposes a package of measures as a key step towards a common data space in the EU.³²²

Such initiative was supported and driven by a stakeholder dialogue and replies to the Public Consultation on "Building the European Data Economy".³²³ As regards B2B data sharing, such stakeholder dialogue showed that stakeholders are not in favour of a new 'data ownership' type of right, on grounds that "*the crucial question in business-to-business sharing is not so much about ownership, but about how access is organised*".³²⁴

Legal doctrine related to data ownership

In line with the increasing coverage of data ownership by the Commission in its Communications, the problem of data ownership has been reported by numerous authors.

Some authors are generally in favour of the creation of an ownership right³²⁵, whereas others make the distinction between an exclusive and non-exclusive right to property in data. Thus, the Max Planck Institute for Innovation and Competition has stated, jointly with other authors, that it could see neither a justification nor a necessity to create exclusive rights in data.³²⁶ Other academics do not necessarily dismiss the idea of an exclusive right in

data, but claim its advent to be premature.³²⁷ The authors of this Chapter already expressed their preference for the creation of a non-exclusive ownership right paired with data sharing obligations in the context of the EU-funded H2020 project TOREADOR.³²⁸

Looking at the situation under Member States' laws, we observe a similar level of divergence.

The current lack of clarity as to the status of data under **UK** law was addressed for instance by Christopher Rees³²⁹, who believes that data could be classified as property (based on a simple definition of property as the right to use something and exclude others from its use).

Most of the German academics argue that **German** law does not know a right in data as such³³⁰, even if in some instances they recognised the need for creating such right. There are however voices opposing this line of thought, in view of the jurisprudence of the German Courts. In particular, Prof. Dr. T. Hoeren examined the issues of data ownership under the current German legal framework and jurisprudence³³¹, concluding that "*in general, the property in data is attributed to the originator, creator, or producer of these data. However, in the case of data made for hire (to use the US copyright term), the data belong to the employer*". Other scholars seem to suggest that one may rely on the current wording of Section 950 of the German Civil Code to claim some kind of property right in data. Such Section stipulates that "*A person who, by processing or transformation of one or more substances, creates a new movable thing acquires the ownership of the new thing, except where the value of the processing or the transformation is substantially less than the value of the substance. Processing also includes writing,*

³²² Commission, 'Towards a common European data space' (Communication) COM(2018) 232 final

³²³ European Commission, 'Public Consultation on Building the European Data Economy' (*European Commission*) <<https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy>> accessed 18 October 2018

³²⁴ COM(2018) 232 final 9

³²⁵ Herbert Zech, 'Information as Property' (2015) 6 JIPITEC 192 <<https://www.jipitec.eu/issues/jipitec-6-3-2015/4315>> accessed 18 October 2018

³²⁶ Josef Drexel and others, 'Data Ownership and Access to Data - Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate' (Max Planck Institute for Innovation and Competition Research Paper No. 16-10, 2016) <<http://dx.doi.org/10.2139/ssrn.2833165>>; Josef Drexel, 'Designing Competitive Markets for Industrial Data in Europe - Between Propertisation and Access' (2017) 8 JIPITEC 257 <<https://www.jipitec.eu/issues/jipitec-8-4-2017/4636>>

accessed 18 October 2018; Bernt Hugenholtz, 'Against Data Property' in Hanns Ullrich, Peter Drahos and Gustavo Ghidini (eds), *Kritika: Essays on Intellectual Property* (Volume 3, Edward Elgar Publishing Limited 2018); Wolfgang Kerber, 'A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis' (Joint Discussion Paper Series in Economics No. 37-2016) <https://www.uni-marburg.de/fbo2/makro/forschung/magkspapers/paper_2016/37-2016_kerber.pdf> accessed 18 October 2018

³²⁷ Andreas Wiebe, 'Protection of Industrial Data - A New Property Right for the Digital Economy?' (2017) 12(1) Journal of Intellectual Property Law & Practice 62

³²⁸ Benoit Van Asbroeck, Julien Debussche and Jasmien César, 'White Paper - Data Ownership in the Context of the European Data Economy: Proposal for a New Right' (Bird & Bird 2017) <<https://www.twobirds.com/en/news/articles/2017/global/data-ownership-in-the-context-of-the-european-data-economy>> accessed 21 February 2019; Benoit Van Asbroeck, Julien Debussche, Jasmien César, 'Supplementary Paper - Data Ownership: a new EU right in data' (Bird & Bird 2017) <<https://www.twobirds.com/en/news/articles/2017/global/data-ownership-a-new-eu-right-in-data>> accessed 21 February 2019.

³²⁹ Christopher Rees, 'Who Owns our Data?' (2014) 30(1) Computer Law & Security Review 75

³³⁰ See e.g.: Michael Dorner, 'Big Data und "Dateneigentum"' (2014) 9 CR 617, Malte Grützmacher, 'Dateneigentum - ein Flickenteppich' (2016) 8 CR 485

³³¹ Thomas Hoeren, 'Big Data and the Ownership in Data: Recent Developments in Europe' (2014) 36(12) EIPR 751

drawing, painting, printing, engraving or a similar processing of the surface." Despite the legal uncertainty surrounding such theory, and notably its particular application to intangible assets such as data, certain undertakings have already relied on it in their general terms and conditions. Having said that, the majority of German academics seems to agree that no right in data exists.

Commentators seem to be divided as to the ownership of data under **French** law. While some commentators indicate that data are not appropriable as such³³², others believe that in view of the abovementioned ruling of the French Supreme Court the ownership over data cannot be called into question.³³³ Having said that, most discussions on the recognition of ownership seem to focus on individuals' ownership over their personal data.³³⁴



Illustration in the transport sector:

In the course of 2017, the German Federal Ministry of Transport and Digital Infrastructure (Bundesministerium für Verkehr und digitale Infrastruktur – "**BMVI**") conducted a study, the results of which advocate the creation of an ownership right for (mobility) data.³³⁵ In said study, the BMVI highlights the opportunities of (big) data use in the transport sector. It however regrets the heterogeneity and fragmentation of data-related regulations, and therefore advocates the creation of a – potentially exclusive – property-like right in (mobility) data in order to encourage the development of new business models. The BMVI suggests assigning data to the one who has made a substantial investment in the creation thereof, as it feels this would be in line with the economic reality and would provide legal certainty. In order to implement the ownership right in practice, the BMVI considers two different

options. The first option entails the immediate creation of an entirely new "data law". The second option consists of different measures that would eventually lead to the development of a data law.

Conclusion

In a big data context, different third-party entities may try to claim ownership in (parts of) a dataset, which may hinder the production of, access to, linking and re-use of big data, including in the transport sector. This Chapter has however amply demonstrated that the current legal framework relating to data ownership is not satisfactory.

No specific ownership right subsists in data and the existing data-related rights do not respond sufficiently or adequately to the needs of the actors in the data value cycle. Up until today, the only imaginable solution is capturing the possible relationships between the various actors in contractual arrangements.

Nevertheless, filling the data ownership gap with contractual arrangements is far from ideal. It would be practically burdensome – and probably even impossible – to regulate with full legal certainty by means of contracts the ownership issues in large-scale data undertakings where there is a multitude of data sources, storages, analyses and thus a myriad of actors who would want to claim ownership in the data concerned. On top of all that, comes the issue where contracts are in principle nonbinding, and therefore unenforceable, vis-à-vis third parties. This issue is further examined in the next Chapter, which will address data sharing agreements in the context of big data, with illustrations drawn from the transport sector.

³³² Alexandra Mendoza-Caminade, 'La protection pénale des biens incorporels de l'entreprise: vers l'achèvement de la dématérialisation du délit' (2015) 7 *Recueil Dalloz* 415; Céline Castets-Renard, 'Les opportunités et risques pour les utilisateurs dans l'ouverture des données de santé: big data et open data' (2014) 108 *Revue Lamy Droit de l'immatériel* 38

³³³ Pierre Berlioz, 'Consécration du vol de données informatiques. Peut-on encore douter de la propriété de l'information?' (2015) 4 *Revue des contrats* 951

³³⁴ The particular issue of personal data ownership will be discussed in a separate Chapter; Alain Bensoussan, 'Propriété des données et protection des fichiers' (2010) 296 *Gazette du Palais* 2; Isabelle Beyneix, 'Le traitement des données personnelles par les entreprises: big data et vie privée, état des lieux' (2015) 46-47 *Semaine juridique* 2113

³³⁵ Bundesministerium für Verkehr und digitale Infrastruktur, 'Eigentumsordnung für Mobilitätsdaten? – Eine Studie aus technischer, ökonomischer und rechtlicher Perspektive' (BMVI) <<https://www.bmvi.de/SharedDocs/DE/Artikel/DG/studie-mobilitaetsdaten-fachkonsultation.html>> accessed 18 October 2018



Data sharing agreements

In this thirteenth Chapter, we offer a brief overview of what can be defined as a data sharing agreement, the rules that may apply to these agreements arising both from the law and from contractual obligations established by the parties, and of the guidance issued by the European Commission in this respect. This Chapter also provides a critical analysis of the common practice to use data sharing agreements to govern the access to and/or exchange of data between stakeholders in a big data analytics lifecycle.

It follows from our previous Chapter that there is a multitude of actors on the market actively reaping the benefits of the data economy. The relationship between these actors is at the heart of the data value cycle. It is however also apparent from the previous Chapter that the legal framework is unfortunately not satisfactory at this stage. In fact, it is clear that one of the factors limiting the availability, use, and exchange of data in commercial settings is the legal regime – or lack thereof – in place.

As things stand, the various commercial entities exchanging data in the context of the (big) data value cycle do so mainly on the basis of contractual agreements (i.e. data sharing agreements or "DSAs").³³⁶ It is therefore required to carefully assess the multiplicity of (often multi-layer) agreements governing the access and the exchange of data between the various actors, taking into consideration the type of data involved in the analytics processing.

³³⁶ European Commission, 'Synopsis Report: Consultation on the "Building a European Data Economy" Initiative' (European Commission 2017) 5 <http://ec.europa.eu/information_society/newsroom/image/document/2017-36/synopsis_report_-_data_economy_A0EFA8E0-AED3-1E29-C8DE049035581517_46646.pdf> accessed 23 October 2018; see also in the context of artificial intelligence: Hervé Jacquemin and Jean-Benoît Hubin, 'Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle' in Hervé Jacquemin and Alexandre De Streel (eds), *Intelligence artificielle et droit* (Larcier 2017)

Data sharing agreements: definition and applicable rules

A DSA can be defined as an agreement between two or more legal entities (or individuals) concerning the sharing of data or information of any kind between these legal entities (or individuals). The notion of 'data sharing agreement' is commonly used to refer to a broad typology of arrangements and documents between two or more organisations or different parts of an organisation. The present Chapter does not intend to cover any contractual relationships with natural persons in their capacity as consumers or data subjects.

Depending on the specific needs of the parties, the sharing of data may take different forms, such as for instance reciprocal exchange of data, one or more organisations providing data to one or more third parties, several organisations pooling information and making it available to each other or to third parties, one-off disclosures of data in unexpected or emergency situations, different parts of the same organisation making data available to each other, etc.

Finally, the types of data shared may be of a different nature, such as for instance³³⁷ data about identified or identifiable natural persons ("personal data" – see also our second Chapter Privacy and

³³⁷ Non-exhaustive and non-exclusive list

Data Protection), data protected by intellectual property rights or another kind of property-like right, data considered confidential (including trade secrets and know-how), financial data, etc.

The parties to a DSA are bound to comply with obligations at two levels:

- Mandatory rules arising from the applicable law(s); and
- Contractual terms and conditions specifically set forth and agreed upon by the parties.

The DSA shall first of all be in line and comply with the applicable (national) laws and regulations concerning the formation and execution of an agreement, notably relating to the activity of data sharing. Most of such rules derive from the contract law applicable to the DSA.

Such rules may concern, among others:

- **Formal requirements** (when applicable): e.g. the applicable law may require that certain types of DSAs – for instance the data processing agreement to be executed between a data controller and a data processor – are executed in writing; or the choice of the law applicable to the contract is valid and enforceable only if agreed in writing between the parties.³³⁸
- **Formation of the contract**: these rules are relevant to assess whether a DSA and its obligations are enforceable between the parties.
- **Termination**: the right of the parties to terminate the agreement.
- **Liability**: in case of breach of any contractual obligation, such as when one of the parties discloses the data received from the other party to another party not authorised to receive the data.
- **Capacity of signatories**: the legal capacity of the persons undersigning the agreement to act on behalf of an organisation (e.g. if a person who signs a DSA does not have the capacity or authority to sign it, the DSA will be ineffective).
- **Assignment**: the right of the parties to assign the DSA, or part of the rights and/or obligations under the DSA, to a third party (e.g. in most circumstances, and jurisdictions, the assignment or transfer of an agreement, especially if it is a DSA, requires the consent of the other party).

³³⁸ Regulation (EC) 593/2008 on the law applicable to contractual obligations (Rome I) [2008] OJ L 177/6, art 3(1)

In addition to compliance with the possible restrictions laid down by the applicable legislations and/or regulations on the sharing of data in general, parties shall bear in mind when drafting the DSA that the sharing of the data under the agreed terms and conditions may need to comply with all specific rules that the applicable legislation may have set for a particular type of data or information, such as for instance financial data or health data.

Finally, within the limits identified above, the parties to a DSA are free to agree on additional terms and conditions applicable to their sharing of data. For instance, the parties may agree on details related to specific obligations connected to the sharing of data, time of disclosure, warranties (or lack of warranties) on the accuracy and completeness of data, obligations of the receiving party to manage the data according to specific rules and to apply certain security measures to protect the data, right of or prohibition to the receiving party to transfer onward/disclose the data to a third party, ownership of the data and intellectual property rights, payment of any consideration for the sharing of data, confidentiality obligations, audit of the receiving party by the disclosing party or by the authorities, warranties on the power to disclose and receive data, duration of the agreement, governing law, and competent court.

Guidance from the European Commission

Following a broad stakeholder consultation and dialogue, the European Commission recently deemed it inappropriate to take horizontal legislative action with respect to private sector data sharing. Companies had urged the Commission to be prudent when considering taking action in order to make more data available for re-use. It was argued that data value chains and data-based business models are extremely diverging and that a one-size-fits-all solution would most likely prove inadequate. Instead, companies expressed their preference for agreements as the way to address most concerns. Stating that "*contracts build on trust*", the latter was considered an essential prerequisite for all private sector data sharing.³³⁹

³³⁹ European Commission, 'Synopsis Report: Consultation on the "Building a European Data Economy" Initiative' (European Commission 2017) 5
<http://ec.europa.eu/information_society/newsroom/image/document/2017-36/synopsis_report_-_data_economy_AoEFA8E0-AED3-1E29-C8DE049035581517_46646.pdf> accessed 23 October 2018

The European Commission then issued guidance on 'Sharing private sector data in the European data economy'.³⁴⁰ This was aimed at providing a practical toolbox for both data-holding and data-using businesses across industries regarding the legal, business, and technical aspects of data sharing. The guidance addresses data sharing among private companies (i.e. B2B), as well as the provision of data from a private company to the public sector (i.e. business-to-government or "**B2G**"). Taking account of the fact that data sharing usually takes place on the basis of an agreement, the Commission establishes five principles to govern B2B DSAs and six principles to govern B2G DSAs. These will be briefly addressed below.



Illustration in the transport sector:

On 19 October 2018, the European Commission published its Roadmap on Cooperative, Connected and Automated Mobility (CCAM) in light of its aim to publish a Recommendation on this subject during the first quarter of 2019. In addition, a Public Consultation was kicked off on 24 October 2018. One of the issues to be addressed by the Recommendation is access to in-vehicle data. The Commission indeed deems the centralisation of in-vehicle data (as it is currently practiced by some market players) insufficient to ensure fair and undistorted competition between service providers. The Commission Recommendation therefore aims to provide further guidance on a governance framework for access to and sharing of data generated by connected vehicles. The Roadmap and the Public Consultation were open for feedback on the Better Regulation platform until 16 November 2018 and 4 December 2018 respectively.³⁴¹ Any feedback will be taken into account for further development of the initiative.

B2B data sharing agreements

On a preliminary note, the Commission identifies five principles which should govern private data sharing in order to ensure "*fair markets for IoT objects and for products and services relying on data created by such objects*".³⁴² These principles are displayed in the table below:

³⁴⁰ Commission, 'Guidance on sharing private sector data in the European data economy Accompanying the document Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the Regions "Towards a common European data space"' (Staff Working Document) SWD(2018) 125 final, 18

³⁴¹ European Commission, 'Cooperative, Connected and Automated Mobility (CCAM)' (European Commission 2018) <https://ec.europa.eu/info/law/better-regulation/initiatives/ares-2018-5349236_en> accessed 22 February 2019

³⁴² SWD(2018) 125 final, 3

Principle	DSAs should therefore:
Transparency	Identify the persons or entities that will have access and use the data generated by the product or service in question and specify the purposes for such data use.
Shared value creation	Recognise that, where data is generated as a by-product of using a product or service, several parties have contributed to creating the data.
Respect for each other's commercial interests	Address the need to protect both the commercial interests and secrets of data holders and data users.
Ensure undistorted competition	Address the need to ensure undistorted competition when exchanging commercially sensitive data.
Minimised data lock-in	Allow and enable data portability as much as possible, particularly where companies offer products or services that generate data as a by-product.

The guidance then goes on to discuss some of the legal aspects related to B2B data sharing through DSAs (i.e. data usage or licensing agreements). It recognises that data monetisation agreements are not necessarily bilateral and may be concluded between multiple parties. Emphasis is also put on the fact that these contracts do not exist in a legal vacuum and attention should therefore be given to ensure compliance with existing legislation, particularly legislation that would prevent data sharing or make it subject to specific conditions. This includes for instance the GDPR whenever personal data are involved, but may also cover sector-specific obligations. The Commission also voices its plans to collect best practices, existing model contract terms and checklists through a Support Centre for data sharing which is expected to become operational in the course of 2019.³⁴³

The section on DSAs also contains a list of considerations to help companies in the preparation and/or negotiation of DSAs. It covers topics such as what data should be made available, who can access and (re-)use that data, what can that (re-)user do with the data, the definition of technical means of data access and/or exchange, what data should be protected and how, liability questions and audit rights for both parties.³⁴⁴ We briefly address the most important considerations below.

Companies are advised to describe the data in the most concrete and precise manner possible. This ideally includes the levels of updates to be expected in the future. Another important question concerns the quality of the data. The Commission states that

good quality data is accurate, reliable and where necessary up-to-date and that a dataset ideally does not have missing, duplicate or unstructured data. It should in any case be ensured that the rights of third parties are respected, including intellectual and industrial property rights.³⁴⁵

The contract should determine in a clear and transparent manner who has a right to access, a right to (re-)use, and a right to distribute the data. According to the Commission, rights to access and re-use do not need to be unlimited and may be subject to conditions, which should be clearly defined in the DSA. The contract may limit e.g. the right to access to members of a certain group, or affiliates of a certain company, or limit the right to re-use to certain specific purposes. Companies should moreover consider if and how data may be licensed for re-use and include the necessary specifications in this regard. Sub-licensing may also be considered in the sense that it should either be expressly excluded, or the conditions under which it is allowed should be clearly stipulated.³⁴⁶

The parties gaining access to the data should be as open and clear as possible about how the data will be used, including by other parties downstream. This ensures transparency and increases trust of the data supplier. The contract can address this by specifying the exact usage that can be made of the data, including rights on derivatives of such data (e.g. analytics). Non-disclosure rules regarding downstream parties and others may be helpful in this respect.³⁴⁷

³⁴³ Ibid 6
³⁴⁴ Ibid 6-8

³⁴⁵ Ibid 6
³⁴⁶ Ibid 7
³⁴⁷ Ibid

The DSA should moreover determine the technical means and modalities for data access and/or exchange. This includes among others the frequency of data access, maximum loads, IT security requirements and service levels for support.³⁴⁸

Considerations regarding the protection of data should be made at two levels. On the one hand, a company should require appropriate measures to be put in place for protecting its data. The measures ought to apply to data sharing transactions as well as data storage, taking account of the fact that data can be subject to theft or misuse by both organised groups and individual hackers. On the other hand, organisations should consider the protection of trade secrets, sensitive commercial information, licences, patents and other intellectual property rights. Neither party should aim at retrieving sensitive information from the other side as a result of the exchange of data.³⁴⁹

It is recommended to include liability provisions to cover situations of supply of erroneous data, disruptions in data transmission, low quality interpretative work if shared with datasets, or the destruction/loss or alteration of data (if unlawful or accidental) that may potentially cause damages. Companies are also advised to define a right for each party to perform audits regarding the respect of the mutual obligations. The duration of the contract and possibilities for termination should of course be carefully considered, as well as the applicable law and dispute settlement options.³⁵⁰

In addition to the legal (contractual) aspects, the Commission considers the technical aspects of B2B data sharing in its guidance. It notably distinguishes three types of technical data sharing mechanisms: (i) one-to-many via unilateral mechanisms, such as an application programming interface (API) or an industrial data platform; (ii) data monetisation via a many-to-many data marketplace; and (iii) data sharing via a technical enabler.



Illustration in the transport sector:

Nallian³⁵¹ has created a cloud-based customisable platform that facilitates real-time data sharing in a controlled, flexible and agile environment and supports process synchronisation.³⁵² The users of Nallian's platform are currently logistics hubs and companies, vertical supply chains and multimodal transport networks.³⁵³ Data suppliers can define rules for sharing and terms of use for the different members of the community through a rights-granting mechanism embedded in the platform (which could be qualified as a DSA).³⁵⁴ Examples of communities relying on the Nallian platform relevant to the transport sector are BRUcloud³⁵⁵, CargoStream³⁵⁶, NxtPort³⁵⁷, and Heathrow CargoCloud³⁵⁸.

B2G data sharing agreements

The Commission also identifies six principles to govern data sharing by private companies with public sector bodies (B2G data sharing), under preferential conditions for re-use. Said principles are listed in the table below.

³⁴⁸ Ibid
³⁴⁹ Ibid
³⁵⁰ Ibid 7-8

³⁵¹ See <<https://www.nallian.com>>

³⁵² Everis Benelux, 'Study on Data Sharing between Companies in Europe' (European Commission 2018) <<https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en>> accessed 23 October 2018

³⁵³ Ibid

³⁵⁴ SWD(2018) 125 final, 11

³⁵⁵ See <<https://brucloud.com>>

³⁵⁶ See <<https://www.cargostream.net>>

³⁵⁷ See <<https://www.nxtport.eu>>

³⁵⁸ See

<<https://www.heathrow.com/company/cargo/cargocloud>>

Principle	DSAs should therefore:
Proportionality in the use of private sector data	Justify any requests for supply of private sector data under preferential conditions by clear and demonstrable public interest. Requests should be proportionate and the associated costs and efforts for the undertaking concerned should be reasonable compared with the expected public benefits.
Purpose limitation	Specify one or more purposes for the re-use of data by the public body, which may also include a limited duration for use of the data. Additionally, specific assurances should be offered by the public body that the data will not be used for unrelated administrative or judicial procedures.
‘Do no harm’	Include safeguards to ensure that legitimate interests of the private company, notably the protection of its trade secrets and other commercially sensitive information, are respected, so as not to impede the company from being able to monetise the insights derived from the data in question with respect to other interested parties.
Conditions for data re-use	Seek to be mutually beneficial while acknowledging the public interest goal by giving the public sector body preferential treatment over other customers, particularly in terms of the agreed level of compensation. Ensure that the same public authorities performing the same functions are treated in a non-discriminatory way.
Mitigate limitations of private sector data	Ensure that companies supplying the data offer reasonable and proportionate support to help assess the quality of the data for the stated purposes, including through the possibility to audit or otherwise verify the data wherever appropriate. Companies should however not be required to improve the quality of the data in question.
Transparency and societal participation	Be transparent about the parties to the agreement and their objectives, and ensure that public bodies’ insights and best practices of B2G collaboration will be disclosed to the public as long as those do not compromise the confidentiality of the data.

Similarly to the part on B2B data sharing, the guidance then lists a number of considerations to help public bodies and private companies in the preparation and/or negotiation of DSAs. These will not be examined in detail in this Chapter but include topics such as (i) identification of a public interest purpose and of the private data concerned; (ii) identification of internal challenges and constraints related to the sharing of data; (iii) definition of technical means and modalities of data access and/or exchange; (iv) conditions for implementation; (v) common guiding principles for monitoring implementation of the contract; (vi) liability concerns; and (vii) dissemination by the public body of the results and/or insights of the

collaboration without compromising the confidentiality of the data involved.³⁵⁹

The Commission also outlines the following technical means to achieve B2G data sharing: (i) data platforms; (ii) algorithm-to-the-data; and (iii) privacy-preserving computation.

Data sharing agreements: a critical analysis

As already mentioned in other Chapters, big data analytics involves a multitude of complex data flows, data sources, algorithms, analyses, etc. Also, it entails the participation of many different actors and many different activities that can be performed on the data. To this end, access to and/or exchange

³⁵⁹ SWD(2018) 125 final, 14-16

of data must be enabled and facilitated. It is apparent from our research that, at least from a legal perspective, this can currently only be achieved through the conclusion of DSAs. In view of the aforementioned complexity and multitude of actors, data sources, data flows, algorithms, etc., an intricate chain of DSAs should be put in place in order for the big data analytics to (legally) function in practice.

However, the authors of this Chapter are reticent to settle for DSAs as the one and final solution forevermore, given the inherent limitations of agreements in a big data context. Some of these limitations are briefly discussed below.

First, contractual agreements in principle only generate rights and obligations for the parties to such agreements. They can therefore not be enforced vis-à-vis third parties. In practice, this would entail that there is no recourse available against third parties that obtain unjustified access to and/or misuse the data.

Second, contractual agreements require a clear and precise definition of the concepts they intend to regulate. It proves however extremely difficult to clearly define the concept of "data" as no strict legal definition of this concept exists. In practice, this leads to a myriad of possible interpretations of "data" in different agreements without any harmonised view on the legal meaning of "data". In the same vein, similar difficulties arise when stakeholders active in the big data analytics lifecycle attempt to contractualise data ownership through the terms of the DSA, given that the concept of "data ownership" is not legally defined. Such stakeholders can therefore try to define the concepts of "data" and "ownership" as broadly as possible, thereby creating a far-reaching entitlement to any element included in the big data analytics process, which would practically impede the implementation of the big data analytics as a whole.

Third, aside from a broad definition of "data ownership", the specific terms of a DSA covering the permitted actions to be performed on or with the data may be phrased in a highly restrictive manner, thereby prohibiting actions such as reverse engineering, merging, enriching, sharing, decompiling, translating, adapting, arranging, preparing, structuring, cleansing, altering, displaying, reproducing, visualising, communicating, loading, running, transmitting, storing, observing, studying, testing, etc. In essence,

this would render the whole data sharing exercise, and therefore the big data analytics, unworkable as the recipient(s) would be unable to do anything with the data.

Fourth, any restrictions on the downstream use of the data (such as e.g. those that may be imposed by a holder of intellectual property rights) and any warranties regarding the upstream source of the data (such as e.g. personal data collected directly from the data subject with the latter's consent) should be covered by complex back-to-back warranty clauses in the multiple DSAs in order to ensure the proper legal functioning of the big data analytics. In absence of such clauses, the further use of data may be prohibited or restricted, which would allow blocking the whole big data analytics chain.

Conclusion

This Chapter examined the common practice of using contracts, i.e. DSAs, to govern the access to and/or exchange of data between stakeholders in a big data analytics lifecycle.

It is unclear, however, whether such practice enables covering all possible situations with the necessary and satisfactory legal certainty. Indeed, DSAs entail numerous limitations in the absence of a comprehensive legal framework regulating numerous rights (e.g. ownership, access or exploitation rights) attached to data, the way in which such rights can be exercised, and by whom.

Against a background where the EU strives towards a data-driven environment in which both citizens and companies can reap the benefits of novel data technologies, but also against a background where the current legal framework does not sufficiently tackle all the issues related to data and where actors involved in the data value chain have no certainty as to the ownership of the data they have gathered, created, analysed, enriched or otherwise processed; a more solid and legally secure solution would be desirable.³⁶⁰

³⁶⁰ Benoit Van Asbroeck, Julien Debussche and Jasmien César, 'White Paper – Data Ownership in the Context of the European Data Economy: Proposal for a New Right' (Bird & Bird 2017) <<https://www.twobirds.com/en/news/articles/2017/global/data-ownership-in-the-context-of-the-european-data-economy>> accessed 22 February 2019; Benoit Van Asbroeck, Julien Debussche, Jasmien César, 'Supplementary Paper – Data Ownership: a new EU right in data' (Bird & Bird 2017) <<https://www.twobirds.com/en/news/articles/2017/global/data-ownership-a-new-eu-right-in-data>> accessed 22 February 2019

Competition



In this fourteenth Chapter, we will focus on the impact of big data on different aspects of EU competition law and seek to create more clarity on when and how the ownership or (mis)use of (big) data can give rise to competition law issues. Specific illustrations in the transport sector will be provided.

Big data has been a hot topic in competition law for several years now. It has been on the radar at national level for quite some time, and given significant attention by the European Commission more recently in the context of shaping competition policy in the era of digitisation.

As such, big data aggregation in the transport sector can give rise to a variety of competition law issues that suggest that certain aspects of competition law may not be fit for purpose. Abuse of dominance, merger control, and anticompetitive behaviour have all seen challenges in the application of competition law and will be addressed in this Chapter as well as in the context of the transport sector.

Abuse of dominance

The challenge in the context of big data and abuse of dominance lies in measuring market power and subsequently the potential for abuse of dominance. The simple fact that a company has access to large amounts of data does not automatically provide it with a dominant market position. Important factors that need to be taken into account to determine the existence of dominance include:

- Do other competitors have access to the same data?
- Is there data which can substitute the data collected by the company?
- Does the company have the ability to analyse and monetise the collected data?
- Is the data held by the company raw data or fully analysed data?

The trend in current competition analyses seems to focus primarily on the amount of data, with limited attention being given to the aspects listed above. These aspects may lead to the conclusion that, in a given case, even access to a very large amount of data does not provide a company with market power.

The main criteria to determine whether access to certain data gives market power include:

- **Quantity:** Once a certain volume of data has been gathered, the collection of additional data will not necessarily lead to any significant additional findings or benefits for the collecting company (so-called diminishing returns theory). The level above which the returns decrease will obviously differ between companies and industry sectors;
- **Quality:** Not all collected data has the same value. Raw data which cannot be processed and thus cannot be immediately monetised has a lower value than data which is ready for use;
- **Availability:** As mentioned above certain data is readily available to multiple companies since consumers typically use their personal data in different manners for different purposes (multi-homing).

The joint study published by the French and German competition authorities³⁶¹ suggests that future cases could be based on the logic that abuse of dominance can arise from a firm's ability to

³⁶¹ 'The French Autorité de la Concurrence and the German Bundeskartellamt launch a joint project on algorithms and their implications on competition' (Bundeskartellamt) <https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2018/19_06_2018_Algorithmen.htm> accessed 18 October 2018

derive market power from big data that a competitor is unable to match. Particularly, they propose two questions to be examined in such cases:

- Whether there is a scarcity of data and whether competitors are able to easily obtain/replicate this data.
- Whether the scope and scale of the relevant data matter for the assessment of market power.³⁶²

The important, and constantly evolving role of big data in the digitalised transport sector in general, and in transport companies in particular, naturally also has an impact on the competition law issues faced by companies active in that sector. An example given is companies selling their data to third parties who can then make use of it in an exploitative manner. For instance, a transport company which tracks, collects and aggregates users' location and specific routes, can sell this data to insurance companies which then justifiably raise their customers' car insurance premiums if they perceive them to regularly drive above the speed limit, take more dangerous routes or use their vehicle more frequently than the average user.

Transport companies that enjoy a dominant position on a specific market and who have in their possession large amounts of data on their customers, could very easily exploit such data with the view to cementing their dominant position in that market and to excluding rivals.



Illustration in the transport sector:

The District Court for the Northern District of California was recently called to rule upon a dispute between urban transportation apps Uber and Lyft.

Uber was using an app called "Hell", which used big data and an appropriate algorithm in order to identify drivers who used both the Uber app and the competing Lyft app. After identifying the drivers using both apps, Uber would subsequently subtly provide certain benefits to these drivers so as to ensure that they use Uber over Lyft.¹

The possible exclusionary effect of this behaviour is clear: by offering so-called targeted "reverse rebates", Uber's main competitor was unable to

³⁶² Europe Economics, 'Big Data: What Does it really Mean for Competition Policy? A Look into the Emergence of Big Data, its Fundamental Importance to Businesses and the Wider Economy, and the Critical Role of Competition Authorities in Ensuring Big Data is not Exploited' (Europe Economics 2017) <www.europe-economics.com/publications/mar-big_data.pdf> accessed 18 October 2018

compete on the market as Lyft was losing its attractiveness to current as well as potential drivers.

This exclusionary effect could be said to contrary to the competition rules, if it could be established that Uber held a dominant position and thus, by engaging in the above behaviour, it abused such position.

Merger control

Mergers between a large undertaking and small emerging companies can have a huge effect on data-related markets resulting in an increase in concentration or differentiated access if the newcomer possesses data or large access to data in a different market. Here, again, there are suggestions that merger control rules fall short on the basis that they are often based on financial thresholds and market shares which may not be triggered leaving the acquisition outside the scope of merger control altogether.

The essential issue that can be observed here is that the current competition tools available to the Commission may not be sufficient to properly analyse the effects of a given merger or possession of data on future competition, following the principle of causality where the Commission has to conduct a predictive assessment of the future market with and without the proposed merger. Therefore, a merger may be cleared only to prove anticompetitive later on down the line, which could not have been properly assessed under the current legislation

The first EU in-depth probe to consider the power of data came about with the European Commission's investigation of Apple's proposed acquisition of Shazam Entertainment (Apple/Shazam case).³⁶³ Shazam is a popular app used to identify a song. The use of the app is often brief and many of its users are anonymous. The Commission was concerned that Apple, by combining its data with Shazam, might obtain an unassailable competitive advantage over rivals. It also had concerns that Apple could gain access to commercially sensitive data on the customers of rival streaming services. After a five-month probe, the Commission concluded that Shazam's app was not unique and that rival streaming services would still have the opportunity to access and use similar

³⁶³ *Apple/ Shazam (Case M.8788) Commission Decision of 06/09/2018 [2018] (OJEU summary not yet available)*

databases.³⁶⁴ The clear message from this case is what matters is the kind of data you are acquiring, how unique it is, whether it can be easily replicated and whether you can shut out rivals.

Anticompetitive agreements

This area of competition law has seen particular challenges and forecasted issues in the transport sector and big data.

When it comes to big data and possible price fixing in an online environment, critical questions are now being asked as to whether price setting by algorithm amounts to an "agreement" or "concerted practice". If algorithms -which need big data- are purposefully programmed to exchange pricing information or other data between competitors or enforce collusion, this will clearly be seen as an agreement or concerted practice between human representatives of the colluding competitors. The more difficult question is to where to draw the line between actions that can be attributed to humans and those that may arise through machines using algorithms employing AI technology such as deep learning.

As pricing algorithms become more widespread amongst firms across all industries, the question arises whether algorithms then mean the end for cartels or, rather, whether they create new and more difficult-to-detect ones. The main concern in this area is with cartels and price collusion between competitors which cannot be proven following the traditional definitions of collusion despite the definition of 'agreements' itself being rather broadly construed under EU law.

The UK's Competition & Market Authority ("CMA") in a report to the OECD³⁶⁵ noted that alongside substantive legal challenges, certain features of algorithms may also make it more difficult as a practical matter to detect and investigate unlawful collusive, abusive or harmful conduct, or to distinguish such unlawful conduct from lawful independent commercial actions. These include the complexity of algorithms and the challenge of understanding their exact operation and effects can make it more difficult for consumers and enforcement agencies to detect algorithmic

abuses and gather relevant evidence. In addition, such challenges of detection may be heightened by the ability of algorithms to rapidly evolve, whether through constant refinement by developers or because self-learning is built into them. Or indeed by the fact that – in a world where most businesses have instant access to pricing data and where market transparency is high – unlawful collusion and “mere” conduct parallelism may look very similar.



Illustration in the transport sector:

Airlines have in their possession large amounts of data on their customers including whether or not a customer prefers to compare prices prior to booking a ticket, or whether he/she books their tickets through a travel agency or an app.

Upon this basis, it has been suggested³⁶⁶ that it is not inconceivable that airlines could take advantage of big data analytics and machine-learning mechanisms in order to engage in price setting through "parallel-pricing" or tacit collusion amongst them. Such behaviours could be found to be contrary to the competition rules as anti-competitive agreements or concerted practices between competitors.

Indeed, airlines are able to decide how to price their airfares upon the basis of different sets of data, such as the expected behaviour of the customer; the price competition; and objective operational factors (such as the aircraft capacity, remaining seats, etc.).

In light of the above, holding crucial information on customers' preferences can be key in setting the airfare price. The possibility of analysing and using this mass amount of information through computer algorithms and other machine-learning mechanisms could lead the airlines to de facto align on price (through the use of the algorithms, which would be in a position to automatically set the price at an optimal level for each type of customer), as the airlines would realise that they do not need to compete to attract customers who are already willing to pay the specific prices set by the algorithm, irrespective of the airline.³⁶⁷ Competition authorities could be faced with

³⁶⁴ See Commission press release IP/18/5662 'Mergers: Commission Clears Apple's Acquisition of Shazam' (6 September 2018) accessed on 19 October 2018

³⁶⁵ OECD, 'Algorithms and Collusion – Note from the United Kingdom' (DAF/COMP/WD(2017)19, OECD 2017). See 4.2 written contribution from the United Kingdom, 127th OECD Competition Committee on 21-23 June 2017.

³⁶⁶ Scott Millwood, 'Big Airlines with Big Data: The European Competition Law Issues Associated with Price-setting in the Airlines Industry Using Big Data Analytics and Machine-learning and the Case for 'Competition-by-design'' (2018) 43(3) *Air & Space Law* 287)

³⁶⁷ Ariel Ezrachi and Maurice E. Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-driven Economy* (Harvard UP 2016) 72

substantial evidentiary obstacles to prove a competition law infringement in the absence of neither human contacts nor human agreement between airlines but rather a tacit collusion between machines.

Conclusion

Assessing the market conduct of companies with access to large volumes of data raises complex issues under competition law. The difficulty of the exercise is compounded by the fact that the analysis also needs to take into account data privacy and consumer protection issues that are intimately linked to the questions under competition law.

Both the European Commission and various national competition authorities are continuing to invest significant time and effort into the competition law analysis of big data, and there is extensive and increasing legal literature on the topic. The recent public consultation on shaping competition policy in the age of digitisation has yielded some interesting insight on how to mould competition law to address these topical issues. However, many issues remain unexplored and new issues will arise as a result of on-going technological developments. An effective response to these developments will require close cooperation in particular between the European competition and data protection authorities and the use of thorough economic analysis to avoid an over-enforcement that could stifle innovation and the emergence of new services and business models.



Trust, Surveillance and Free Will

In the final three chapters in this publication, we examine certain ethical and social aspects of the use of big data. This fifteenth chapter notably looks into the concepts of trust, surveillance and free will in relation to big data. Where appropriate, illustrations from the transport sector are provided.

When big data technologies were first being developed, they were mainly deployed in the sphere of marketing. As such, their use was both lucrative and low-risk. However, as soon as big data technology moved beyond increasing the odds of making a sale to being used in higher-stakes decisions like medical diagnosis, loan approvals, hiring and crime prevention, more social and ethical implications arose.

Trust



The Oxford English Dictionary defines trust as the "*firm belief in the reliability, truth, or ability of someone or something*". Trust as such is not recognised as a fundamental right in the EU Charter of Fundamental Rights ("**EU Charter**"). However, together with the concept of surveillance, it may be discerned in the right to liberty and security acknowledged by Article 6 of the EU Charter.

As explained in the first Chapter General Overview, one of the main dimensions of big data, describing consistency and trustworthiness, is veracity.³⁶⁸ It is however quite difficult, if not impossible, to outline one general shared understanding of trust in relation to big data since the trustworthiness of information can change depending on whom we

speak to, where the data is gathered, or how it is presented.³⁶⁹

Nevertheless, an important aspect of "trust" in relation to big data is trust as the result of the belief in the honesty of stakeholders in the process of collecting, processing, and analysing the big data. In this respect, veracity is in principle a moral requirement according to which big data users (collectors, analysts, brokers, etc.) should respect the individual citizen as a data provider, for instance, by facilitating his or her informed consent. The right not to be subject to automated decision-making or the right to information for individual citizens (as enshrined in the GDPR) seem to be under pressure when veracity, honesty and consequently trust are not upheld.³⁷⁰

Big data for trust & trust in big data

The idea behind 'big data for trust' is to address questions of how to use big data for trust assessment (i.e. to measure trust). While the huge variety of big data may bring opportunities, it may also present challenges in relation to its quality (e.g. heterogeneous and unstructured data). One way to use 'big data for trust' is through so-called reputation systems. The general process of reputation systems can be separated into the

³⁶⁹ Shannon Vallor, *Technology and the virtues: A philosophical guide to a future worth wanting* (Oxford University Press 2016)

³⁷⁰ Bart Custers and others, 'Deliverable 2.2 Lists of Ethical, Legal, Societal and Economic Issues of Big Data Technologies. Ethical and Societal Implications of Data Sciences' (e-SIDES, 2017) 33-34 <<https://e-sides.eu/resources/deliverable-22-lists-of-ethical-legal-societal-and-economic-issues-of-big-data-technologies>> accessed 22 August 2018

³⁶⁸ Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) International Journal of Computer Applications 35

following three steps: (i) collection and preparation; (ii) computation; and (iii) storage and communication. In the first step, data or information about the past behaviour of a trustee are gathered and prepared for subsequent computing. Today's vast number of web applications such as e-commerce platforms, online social networks or content communities has led to huge amounts of reputation data being created from big data. Among others, the need to integrate both explicit and implicit information has been highlighted.³⁷¹ To extract implicit reputation information from big data, most of data that is semi- or unstructured according to the variety property of big data, is handled to get the implicit information through natural language processing and machine learning. In the second step, both the explicit and implicit reputation information are used in the computation phase to calculate a reputation value as its output. This phase consists of filtering, weighting and aggregation processes and is concerned with the question of how relevant the information used is for the specific situation. Finally, the reputation values are combined to generate one or several scores. The final storage and communication step stores the predicted reputation scores and provides them with extra information to support the end users in understanding the meaning of a score-value. In this regard, we may encounter challenges about the reusability of reputation information and the transparency of communication.

'Trust in big data' is about measuring the trustworthiness and accuracy of big data to create high values of data which are coming in large volume and different formats from a wide variety of applications/interfaces. In this regard, analysing this valuable information is not as easy as it seems. There are tools available to extract and process this valuable information from disparate sources, but the real challenge is to know whether the data processed are trustworthy, accurate and meaningful.³⁷² With respect to the trust in big data, there are several trust issues as follows: (i) trust in data quality; (ii) measuring trust in big data; and (iii) trust in information sharing.³⁷³ As mentioned

above, verifying their trustworthiness and especially evaluation of the quality of the input data is essential due to the higher volume of data sources than ever before. In order to ensure the quality of input data, detecting manipulations of the data should be conducted before processing it.³⁷⁴ To ensure data quality, several data mining approaches such as feature selection and unsupervised learning methods for sparsity, error-aware data mining for uncertainty, and data imputation methods for incompleteness have been studied³⁷⁵ as well as the authentic synthetic data benchmarking of different big data solutions has been generated.³⁷⁶

Challenges and opportunities for big data and trust

To derive business value from non-traditional and unstructured data, organisations need to adopt the right technology and infrastructure to analyse the data to get new insights and business intelligence analysis. It can be feasible with the completeness, trustworthiness, consistency and accuracy of big data.

Dynamic technological developments where individuals do not have enough time to adapt, may lead to significant trust issues.³⁷⁷ When it comes to mobility in the field of public transportation, safety and security are linked to trust: people need to feel safe when using transportation means which are new (e.g. self-driving cars) or which could be perceived as threatening (e.g. car-sharing with unknown drivers). As an example, gender or age perspectives could help in designing mobility-as-a-service, considering different needs in terms of easiness and sense of safety in public spaces.

³⁷¹ Johannes Sanger and others, 'Trust and Big Data: A Roadmap for Research' in Morvan F, Wagner R R and Tjoa A M (eds) *2014 25th International Workshop on Database and Expert Systems Applications* (IEEE, 2014) 278-282, DOI: [10.1109/DEXA.2014.63](https://doi.org/10.1109/DEXA.2014.63)

³⁷² Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) *International Journal of Computer Applications* 35

³⁷³ Johannes Sanger and others, 'Trust and Big Data: A Roadmap for Research' in Morvan F, Wagner R R and Tjoa A M (eds) *2014 25th International Workshop on Database and Expert Systems*

Applications (IEEE, 2014) 278-282, DOI: [10.1109/DEXA.2014.63](https://doi.org/10.1109/DEXA.2014.63)

³⁷⁴ Cloud Security Alliance Big Data Working Group, *Expanded top ten big data security and privacy challenges* (White Paper, 2013)

³⁷⁵ Xindong Wu and others, 'Data Mining with Big Data' (2014) 26(1) *IEEE Transactions on Knowledge and Data Engineering* 97

³⁷⁶ Zijian Ming and others, 'BDGS: A Scalable Big Data Generator Suite in Big Data Benchmarking' in Tilmann Rabl and others (eds) *Advancing Big Data Benchmarks* (WBDB 2013, Lecture Notes in Computer Science, vol 8585, Springer, Cham)

³⁷⁷ Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) *International Journal of Computer Applications* 35



Illustration in the transport sector:

Intelligent transportation systems are arguably the most anticipated smart city services.³⁷⁸ Ferdowsi et al. proposed an edge analytics architecture for ITS in which data is processed at the vehicle or roadside smart sensor level to overcome the ITS reliability challenges.³⁷⁹ The architecture is illustrated below.

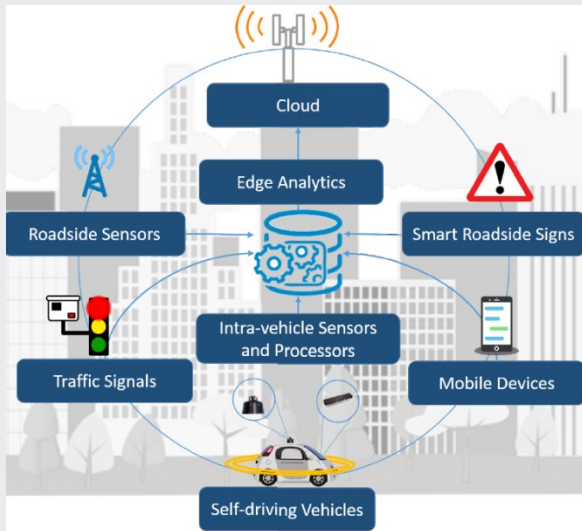


Figure: ITS edge analytics architecture and components

The proposed ITS edge analytics architecture exploits deep learning techniques running at the level of passengers' mobile devices and intra-vehicle processors to process large datasets and enable a truly smart transportation system operation. This architecture improves the performance of ITS in terms of reliability and latency. This example also presents a transparent way of big data collection and the immediate use of processed data in traffic management.

Surveillance



The Oxford English Dictionary describes surveillance as "close observation, especially of a suspected spy or criminal". In the context of this publication, however, the term surveillance will be used to describe the close observation of all humans in general, irrespective of their criminal tendencies. In light of the EU Charter of Fundamental Rights, surveillance can be linked to Article 6 on the right to liberty and security, Article 7 on the respect for

private and family life, and Article 8 on the protection of personal data.

In a big data context, surveillance is said to have the following six key characteristics³⁸⁰:

- 1 Tracking is "populational": big data has as a result that tracking relates to a group of people rather than being targeted at specific individuals.
- 2 Correlation and predictability are no longer needed: when the necessary conditions are fulfilled, big data analytics can provide a reliable and veracious outcome, thus rendering the drawing of assumptions on the basis of correlation and predictions redundant.
- 3 Monitoring is pre-emptive: in an analysis of the simulation of surveillance it was noted that the goal of predictive analytics is not simply to predict outcomes, but to devise ways of altering them. In policing terms, the goal of predicting the likelihood of criminal behaviour is to deter it.³⁸¹
- 4 Tracking is interventionist: in the future, we can expect that predictive analytics will become more sophisticated and will be deployed across a broad range of social life to shape and sort consumer behaviour and opportunities.
- 5 All information is relevant: because predictive analytics is, as it were, model-agnostic, it does not rule out in advance the relevance of any kind of information.
- 6 "Privacy" is irrelevant: any attempt to build a protective bulwark against big data surveillance on the foundation of privacy must confront the fact that much of the tracking is anonymous.

Challenges and opportunities for big data and surveillance

Two main issues arise in relation to surveillance, notably (i) risks of asymmetries in the control over information; and (ii) privacy.

The first issue points to the availability of big data giving a competitive advantage to those who hold the data in terms of capability to predict new economic, social, and political trends. The

³⁷⁸ Junping Zhang and others, 'Data-driven Intelligent Transportation Systems: A Survey' (2011) 12(4) IEEE Transactions on Intelligent Transportation Systems 1624

³⁷⁹ Aidin Ferdowsi, Ursula Challita and Walid Saad, 'Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems' (2017) abs/1712.04135 CoRR [arXiv:1712.04135](https://arxiv.org/abs/1712.04135)

³⁸⁰ Mark Andrejevic, 'Surveillance in the Big Data Era' in Kenneth D. Pimble (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014) 55

³⁸¹ William Bogard, *The Simulation of Surveillance: Hyper-control in Telematic Societies* (Cambridge University Press 1996)

information and knowledge deriving from big data is not accessible to everyone, as it is based on the availability of large datasets, expensive technologies and specific human skills to develop sophisticated systems of analyses and interpretation. For these reasons, governments and big businesses are in the best position to take advantage of big data: they have large amounts of information on citizens and consumers and enough human and computing resources to manage it.³⁸²

When it comes to the privacy issues, the shift from targeted to 'populational' monitoring is facilitated by the advent of interactive, networked forms of digital communication that generate easily collectible and storable meta-data.³⁸³ However, the logic is self-stimulating and recursive: once the switch to an inductive, data-driven form of monitoring takes place, the incentive exists to develop the technology to collect more and more information and to "cover" as much of everyday life as possible.

Privacy-wise we also note that the complexity of data processes and the power of modern analytics drastically limit the awareness of individuals, their capability to evaluate the various consequences of their choices, and the expression of a real free and informed consent.³⁸⁴ This lack of awareness is usually not avoided by giving adequate information to the individuals or by privacy policies, due to the fact that these notices are read only by a very limited number of users who, in many cases, are not able to understand part of the legal terms usually used in these notices, nor the consequences of consenting.³⁸⁵

An opportunity triggered by big data in relation to surveillance is the fact that comprehensiveness replaces comprehension. In other words, big data replaces detection with collection and lets the algorithm do the work. The whole population is monitored by allowing computers to detect anomalies and other patterns that correlate with

suspicious activity. In this respect, it is important to remember that the purpose of monitoring is not eavesdropping on everyone.

Surveillance in big data furthermore contributes to preventive policing. Rather than starting with a suspect and then monitoring him or her, the goal is to start from generalised surveillance and then generate suspects.³⁸⁶ This form of monitoring is not just for purposes of determent (for instance the placement of a surveillance camera in a notorious crime spot) but constitutes an actual strategy for intervention in the future through the use of modelling techniques.³⁸⁷

Applying this to the transport sector, ITS developments have given rise to car and in-car surveillance. This generates trails that are closely associated with individuals and which are available to various organisations. Crash cameras in cars, for example, could be imposed as a condition of purchase, insurance, or rental. Like so many other data trails, the data can be used for other purposes than originally intended (accident investigation), and with or without informed, freely given, and granular consent. In some countries, automatic number plate recognition (ANPR) has exceeded its nominal purpose of traffic management to provide vast mass transport surveillance databases.³⁸⁸



Illustration in the transport sector:

More and more traffic control tools are becoming smart digital devices able to collect and process a high amount of (personal) data.³⁸⁹ Some of them, such as red-light cameras or speed detectors, are used to enforce the law and to legally punish those who commit violations. However, the increasing number of traffic control tools allow for a much more substantial collection of personal data (e.g. parking meters, smart parking applications, automatic tolling systems, etc.). The processing and use of such personal data, for other purposes, would allow tracking individuals. Those data combined with personal data from other sources might give a very accurate image of individuals' social habits who might not have given their consent to those processing

³⁸² Alessandro Mantelero and Giuseppe Vacigo, 'The "Dark Side" of Big Data: Private and Public Interaction in Social Surveillance' (2013) 14(6) *Computer Law Review International* 161

³⁸³ Andrejevic, 'Surveillance in the Big Data Era' in Kenneth D. Pimple (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014) 55

³⁸⁴ Laura Brandimarte, Alessandro Acquisti and George Loewenstein, 'Misplaced Confidences: Privacy and the Control Paradox' (2013) 4(3) *Social Psychological and Personality Science* 340

³⁸⁵ Joseph Turow and others, 'The Federal Trade Commission and Consumer Privacy in the Coming Decade' (2007) 3 *ISJLP* 723

³⁸⁶ *Ibid*

³⁸⁷ William Bogard, *The Simulation of Surveillance: Hyper-control in Telematic Societies* (Cambridge University Press 1996)

³⁸⁸ Marcus R. Wigan, and Roger Clarke, 'Big Data's Big Unintended Consequences' (2013) 46(6) *Computer* 46

³⁸⁹ Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) *Transportation Planning and Technology* 115

activities nor even be aware of them.

Free will



Free will' is defined by the Oxford English Dictionary as "*the power of acting without the constraint of necessity or fate*" or also "*the ability to act at one's own discretion*". It is an underlying principle to most, if not all, rights and freedoms enumerated in the EU Charter of Fundamental Rights. Of course, free will is no absolute given, just like the recitals of the EU Charter state that enjoyment of the rights enshrined in the EU Charter "*entails responsibilities and duties with regard to other persons, to the human community and to future generations*".

Traditional deontological and utilitarian ethics place a strong emphasis on moral responsibility of the individual, often also called 'moral agency'. This idea very much stems from assumptions about individualism and free will.³⁹⁰ We note that these assumptions experience challenges in the era of big data, through the advancement of modern technology. In other words, big data as moral agency is being challenged on certain fundamental premises that most of the advancements in computer ethics took and still take for granted, namely free will and individualism.³⁹¹



Illustration in the transport sector:

Self-driving cars, by definition, monitor vehicles completely autonomously. This raises the ethical question of decision-making, especially in case of unavoidable impact.³⁹² The "Trolley Problem"³⁹³ is often mentioned in this context: if a group of people is in the middle of the road and the self-driving car cannot stop because it is too fast, the car would have to choose between (i) driving into the group of people; (ii) driving into the pedestrian crossing the other lane (the pedestrian being for example an old lady or a young child); or (iii) ploughing into a wall and injuring or killing the driver and/or the

passengers.³⁹⁴ Such decisions can only be made by humans, and the decision-making results will differ between different persons. Creators will need to define algorithms to deal with these kinds of situations.³⁹⁵ To address such types of moral dilemmas, the MIT Media Lab has developed a "Moral Machine"³⁹⁶ to gather citizens' opinions about particular scenarios and share the results with car manufacturers and engineers who develop algorithms. This poll has reached millions of people who took part in the experiment. Interestingly, though not entirely unexpected, the results of the study show that moral choices may differ depending on geographic location and/or culture of the respondents.³⁹⁷

Challenges and opportunities for big data and free will

With the current hyper-connected era of big data, the concept of power, which is so crucial for ethics and moral responsibility, is changing into a more networked concept. Big data stakeholders such as collectors, users, and generators of big data have *relational power* in the sense of a network.³⁹⁸ In this regard, retaining the individual's agency (i.e. knowledge and ability to act) is one of the main and complex challenges for the governance of socio-technical epistemic systems.³⁹⁹

Big data is the effect of individual actions, sensor data, and other real-world measurements creating a digital image of our reality, so-called "datafication".⁴⁰⁰ The absence of knowledge about what data are collected or what they are used for might put the 'data generators' (e.g. online consumers and people owning handheld devices) at an ethical disadvantage in terms of free will. Many

³⁹⁰ Alasdair MacIntyre, *A Short History of Ethics: A History of Moral Philosophy from the Homeric Age to the 20th Century* (Routledge 2003)

³⁹¹ Andrej Zwitter, 'Big Data Ethics' (2014) 1(2) Big Data & Society 1

³⁹² EDPS, Opinion 4/2015

³⁹³ Molly Crockett, 'The Trolley Problem: Would you Kill one Person to Save many Others?' *The Guardian* (12 December 2016) <<https://www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others>> accessed 23 August 2018

³⁹⁴ Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR [arXiv:1802.04103](https://arxiv.org/abs/1802.04103)

³⁹⁵ Caitlin A. Surakitbanharn and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles' <https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf> accessed 23 August 2018

³⁹⁶ <http://moralmachine.mit.edu/>

³⁹⁷ Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon & Iyad Rahwan, 'The Moral Machine experiment', *Nature* volume 563, 59–64 (24 October 2018), <<https://www.nature.com/articles/s41586-018-0637-6>> accessed 12 April 2019

³⁹⁸ Robert A. Hanneman and Mark Riddle, *Introduction to Social Network Methods* (University of California 2005)

³⁹⁹ Judith Simon, 'Distributed Epistemic Responsibility in a Hyperconnected Era' in Luciano Floridi (ed) *The Onlife Manifesto* (Springer, Cham, 2015)

⁴⁰⁰ Kenneth Cukier, 'Big Data is Better Data' (TED 2014) <<https://www.youtube.com/watch?v=8pHzROP1D-w>> accessed 23 August 2018

researchers⁴⁰¹ believe that big data causes a loss of free will and autonomy of humans by applying deterministic knowledge to human behaviour. Even the collection of anonymised data about individuals can lead to illegal behaviours in terms of free will of humans.⁴⁰² Indeed, aggregated and anonymised data could also be used to target individuals established on predictive models.⁴⁰³

With respect to supporting free will of humans, increasing accessibility and personalisation for passengers can provide benefits to people in the form of more personalised or affordable services. Organisations use certain data like journey data to ensure a better understanding and serving of people's needs.⁴⁰⁴

Furthermore, a huge part of what we know about the world, particularly about social and political phenomena, stems from analysis of data. This kind of insight can be extended into new domains by big data, which achieves greater accuracy in pinpointing individual behaviour, and the capability of generating this knowledge can be undertaken by new actors and more powerful tools.⁴⁰⁵ Although a growing body of information being generated from big data provides a level that is imperceptible to individuals⁴⁰⁶, various fields of IT such as information retrieval⁴⁰⁷, user modelling and recommender system⁴⁰⁸ have been studied to provide proper options for people.

With the changing role of data in transport, from data-poor to data-rich, big data in the field of transport is now accessible in new ways and at new

scales. Companies are collecting higher volumes of this data, more frequently, and in real time – as technology makes this more feasible and viable – and are using such data to innovate. Customers expect more personalisation and communication as well as more real-time data being shared. Transport data can be used and shared to benefit businesses, people and public services, potentially in ways that meet the needs of all three groups.

⁴⁰¹ Chris Snijders, Uwe Matzat and Ulf-Dietrich Reips, "'Big Data': Big Gaps of Knowledge in the Field of Internet Science' (2012) 7(1) International Journal of Internet Science 1

⁴⁰² Sciencewise Expert Resource Centre 'Big Data: Public Views on the Collection, Sharing and Use of Personal Data by Government and Companies' (Sciencewise 2014)

⁴⁰³ Charles Duhigg, 'How Companies Learn your Secrets' *The New York Times* (New York, 19 February 2012) 30

⁴⁰⁴ Libby Young and others, 'Personal Data in Transport: Exploring a Framework for the Future' (Open Data Institute 2018) <<https://theodi.org/wp-content/uploads/2018/06/OPEN-Personal-data-in-transport-.pdf>> accessed 23 August 2018

⁴⁰⁵ Larisa Giber and Nikolai Kazantsev, 'The Ethics of Big Data: Analytical Survey' (2015) 2(3) *Cloud of science* 400

⁴⁰⁶ Ralph Schroeder and Josh Cows, 'Big Data, Ethics, and the Social Implications of Knowledge Production' (Data Ethics Workshop, KDD@Bloomberg, New York, 2014) <<https://pdfs.semanticscholar.org/5010/f3927ca8133a432ac1d12a8e57ac11cb3688.pdf>> accessed 23 August 2018

⁴⁰⁷ Beth Plale, 'Big Data Opportunities and Challenges for IR, Text Mining and NLP' in Xiaozhong Liu and others (eds) *Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing* (ACM, 2013) DOI: [10.1145/2513549.2514739](https://doi.org/10.1145/2513549.2514739)

⁴⁰⁸ Fatima EL Jamiy and others, 'The Potential and Challenges of Big Data - Recommendation Systems Next Level Application' (2015) [arXiv:1501.03424v1](https://arxiv.org/abs/1501.03424v1) accessed 23 August 2018



Discrimination



This sixteenth Chapter will delve into a particular social and ethical issue that may materialise in a big data context, namely (data-driven) discrimination. Where appropriate, illustrations from the transport sector are provided.

According to the Oxford English Dictionary, the term 'discrimination' is defined as “*treating a person or particular group of people differently, especially in a worse way from the way in which you treat other people, because of their skin colour, sex, sexuality, etc.*” or in more general terms: “*the unjust or prejudicial treatment of different categories of people, especially on the grounds of race, age, or sex.*”

The principles of non-discrimination and equality are to a great extent covered in Title III of the EU Charter. Thus, the EU Charter recognises the following fundamental rights, freedoms and principles in relation to discrimination: (i) equality before the law; (ii) non-discrimination; (iii) cultural, religious and linguistic diversity; (iv) equality between women and men; (v) the rights of the child; (vi) the rights of the elderly; and (vii) the integration of persons with disabilities.⁴⁰⁹

Elements on which discriminatory treatments can be based are, as mentioned above, skin colour, race, sex, but also for example income or education level, gender, residential area, and others. Using big data analytics to improve business processes or to provide personalised services may lead to discrimination of certain groups of people. At any step of the big data analytics pipeline⁴¹⁰, unintended data biases may be created due to wrong statistical treatment or poor data quality. Big data poses certain challenges requiring expert

knowledge to estimate the accuracy of conclusions drawn from it.⁴¹¹



There is considerable interest in personalised services, individually targeting advertisements, and customised services and product offers. Personalising services means nothing else than to exclude people from or include them into certain target groups on the basis of personal data such as gender, income, education, consumption preferences, etc. Big data analytics relies on the categorisation of information and the conclusions that can be drawn from such categorisation. In that sense, the definition of discrimination in contrast to personalisation does not seem to be straightforward and discrimination might therefore be an inherent part of the analytics process.⁴¹²

Another important aspect of data-driven discrimination concerns the access to and knowledge of technology needed to use digital services or gather valuable information from online platforms or applications. The social differences in the corresponding access to technology and education or skills to use it, are often referred to as the “Digital Divide”.⁴¹³

⁴⁰⁹ EU Charter of Fundamental Rights, art 20-26

⁴¹⁰ Kim Hee and others, 'Big Data Methodologies, Tools and Infrastructures' (LeMO 2018) <https://static1.squarespace.com/static/59f9cdc2692ebdbde4c43010/t/5b6d4674032be489a442fa8b/1533888127770/20180716_D1.3_Big+data+methodologies%2C+tools+and+infrastructures_LeMO.pdf> accessed 24 August 2018

⁴¹¹ D. R. Cox, Christiana Kartsonaki and Ruth H. Keogh, 'Big Data: Some Statistical Issues' (2018) 136 *Statistics & Probability Letters* 111

⁴¹² Rena Coen and others, 'A User Centered Perspective on Algorithmic Personalization' (UC Berkeley 2016)

⁴¹³ Massimo Ragnedda and Glenn Muschert, *The Digital Divide: The Internet and Social Inequality in International Perspective* (Routledge 2013)

Challenges for big data and discrimination

The challenges related to data-driven social discrimination and equity discussed in the framework of this publication are (i) unintended data bias; (ii) intended data bias, i.e. personalised services, offers and advertisements; and (iii) the “Digital Divide”.

Unintended data bias

Biases in datasets or in statements or predictions based on the analysis of datasets can originate from various errors, shortcomings or misinterpretations along the analytics pipeline. The data collection process might be biased by design because of a biased formulation of a survey, a biased selection of data sources, an insufficient length of the surveyed time-period, or the neglect of relevant parameters or circumstances. Along the analytics process, correct statistical treatment and accuracy estimation require expert knowledge. The procedure is therefore prone to methodical and technical errors.

Some of the main causes of unintended data bias are discussed hereafter.

The sample size of a dataset directly influences the validity of a statement or the conclusions drawn from the data sample. The accuracy of a statistical analysis depends on the nature of the sample and its estimation. In the context of big data, the available data often consist of many subsets, demanding careful statistical treatment to normalise the estimating procedure to the single subsets in order to avoid overfitting or wrong conclusions. This heterogeneity calls for clean and careful aggregation of data from different data sources corresponding to different subsets where some unique features are not shared by all sets.

Dealing with a huge amount of data generated from a large amount of individuals or sensors makes analysis prone to errors resulting from bad data quality. Data derived from device measurements or automated studies must be carefully checked for various types of errors that may arise during the collection process. Since the cleaning and checking procedures are usually automated processes themselves, even more attention is required. In some sectors, well-established quality control and assurance procedures exist and should be

standardised in order to ensure reliable conclusions and predictions.⁴¹⁴

Due to the usually high dimensionality, the analysis of big data requires the estimation of simultaneously different variables. Each estimation relates to a corresponding error leading to accumulated errors if a conclusion or algorithm-based prediction is based on many variables. This effect is referred to as noise accumulation and can make it difficult to refer to the original signal. Statistical techniques dealing with this issue require special expertise. Parameter selection and reduction of dimensionality is also crucial to overcome noise accumulation in classification and prediction analytics.

Spurious correlation and incidental endogeneity are two other effects that may lead to wrong conclusions and predictions. Variables or instances might "spuriously" correlate if the correlation is caused by an unseen third variable or event and not by the original variables. High dimensionality makes this effect more likely to occur. It may also be that variables are actually correlated but without any meaning or cause. Incidental endogeneity occurs as a result of selection biases, measurement errors and omitted variables. These phenomena arise frequently in the analysis of big data. The possibility of collecting many different parameters with available measurement techniques increases the risk to create incidental correlation. Big data aggregated from multiple sources with potentially different data generation procedures increases the risk of selection bias and measurement errors causing potential incidental endogeneity.⁴¹⁵

Learning algorithms are often highly complex. This complexity combined with a lack of transparency or comprehensibility for a broader community increases the probability of uncovered errors. Often algorithms are black boxes within a company with limited reproducibility. Open communication, in particular about accuracy levels, uncertainties within the algorithms, or implicit assumptions may often be insufficient.

⁴¹⁴ Jules J. Berman, '15 - Big Data Failures and How to Avoid (Some of) Them' in Jules J. Berman (ed) *Principles and Practice of Big Data* (Second edition, Academic Press 2018); D. R. Cox, Christiana Kartsonaki and Ruth H. Keogh, 'Big Data: Some Statistical Issues' (2018) 136 *Statistics & Probability Letters* 111; Pierre-André G. Maugis, 'Big Data Uncertainties' (2018) 57 *Journal of Forensic and Legal Medicine* 7

⁴¹⁵ Ian L. Dryden and David J. Hodge, 'Journeys in Big Data Statistics' (2018) 136 *Statistics & Probability Letters* 121; David D. Dunson, 'Statistics in the Big Data Era: Failures of the Machine' (2018) 136 *Statistics and Probability Letters* 4

The causes for data bias discussed above are all relevant in the transport sector. They may differ in importance in a specific domain, e.g. for freight and passenger transport. In route optimisation using big data, a huge amount of various sensor data of freight transport-related items might be aggregated with data from other sources (e.g. weather data), which calls for an accurate data merging and cleaning process to ensure good data quality.

Intended data bias

Increasing knowledge on customer or user behaviour and access to personal data creates – besides new business opportunities and the possibility of growth – power, in the sense that personal data of individuals or groups of individuals such as their gender, race, income, residential area and even patterns of their behaviour (e.g. movement profiles) can be aggregated to detailed profiles. Power inherited from such profiles may be unintentionally or intentionally used to discriminate people. The distinction between value-added personalisation and segmentation on the one hand and discrimination on the other hand is not well-defined and therefore depends largely on the experience and perception of the affected individuals.

Some personalised services or advertisements might be discriminatory because they exclude certain groups or are only offered to those people who communicated their personal data. This also includes the selective visibility of a service due to personalised online search results: different groups are not provided with the same information or are offered the same product or service with different pricing or availability options.⁴¹⁶

Personalisation may also lead to discriminatory treatment if it is based on statistical analysis assuming wrong segmentation criteria that are not really representing the target groups or are addressing it in a prejudicial way. Given that the underlying algorithms are typically not accessible to the target groups themselves, their ability to object is limited and it may lead to the manifestation of existing prejudice. In other words, data-based predictions or conclusions are more likely perceived to be objectively true since they rely on “objective”

data. This might lead to even worse discrimination of a social group since prejudicial data can serve as evidence for the confirmation of prejudice.

By way of example, personalised job offers may limit the possibility for individuals of exploring new opportunities if algorithms based on educational backgrounds, professional experience, and other underlying factors do not make them aware of possibilities not fitting their profiles.

Lange, Coen and Berkeley confirmed in their study⁴¹⁷ that users negatively perceive personalisation based on race or household income level. Their study surveyed the opinion of 748 participants. Information on the income level, residence area, and gender were considered as very private information, and negative responses to the use of it for individualised services were recorded. The use of race as a parameter for personalisation was also seen as unfair across all researched domains, i.e. targeted advertising, filtered search results, and differential pricing.

One might consider that such forms of information offering and service platforms are often operated by corporations. Accordingly, the online communication environment is to a large degree dictated by commercial actors who aim to maximise profits. Discrimination might emerge from the fact that people with e.g. lower income or other traits that do not correlate to the business models of those corporations are of less interest.⁴¹⁸

Applied to the domain of passenger transport, this could mean that a segregation of services based on specific characteristics of individuals, such as income or residential area and implicitly race or gender, might take place. The possibility to create new mobility offers according to individualised needs, e.g. private shuttle services combining different modes and optimising routes, might lead to a graduated system of offers dedicated to different social groups with low permeability.

Digital Divide

Discrimination based on social factors and the “Digital Divide” are interconnected: different levels of access and skill in technology are influenced by individuals' social positions, which include

⁴¹⁶ Michael Schrage, 'Big Data's Dangerous New Era of Discrimination' *Harvard Business Review* (2014) <<https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination>> accessed 24 August 2018; Rena Coen and others, 'A User Centered Perspective on Algorithmic Personalization' (UC Berkeley 2016)

⁴¹⁷ Rena Coen and others, 'A User Centered Perspective on Algorithmic Personalization' (UC Berkeley 2016)

⁴¹⁸ Wendy Arianne Günther and others, 'Debating Big Data: A Literature Review on Realizing Value from Big Data' (2017) 26(3) *Journal of Strategic Information Systems* 191

characteristics like age, gender, race, income, and level of education amongst others.

The term “Digital Divide” was first referred to as the diffusion of Internet access throughout the population but is nowadays extended to a “second-level Digital Divide”, which includes the different degrees of skill, time, knowledge, and usage possibilities. It turns out that social status directly influences the online usage behaviour, as higher education for example correlates with a higher online user experience in the fields of information retrieval and transactional purposes. Certain user groups are more likely to become more disconnected from the benefits of Internet usage, which might lead to reinforcement of existing social inequities.⁴¹⁹

In countries with high diffusion rates of Internet access (see comparison for Europe⁴²⁰), the ability and skill to use online services or platforms becomes a substantial part of social life and individuals depend on it in various fields of their professional and private life.⁴²¹ In the transportation sector, this is for example the case in route planning. Route planning is increasingly managed by applications or navigation programmes ensuring, among others, the online availability of public transport schedules, the purchase of tickets, and access to real-time information about the route. This however requires a certain level of skills, access to technology in the form of appropriate devices and some financial contributions.

Opportunities for big data and discrimination

Personalisation and segmentation for customised services or targeting may resolve biases. Big data analytics might indeed also be utilised to decrease social inequity and to improve existing discriminatory situations or services. Discriminatory situations can be made visible using big data analysis, which is the first step to resolve biases. Personalised or individualised services could in a second stage offer the possibility to people with special needs, who are not fitting the majority, to improve their inclusion into society. This could be seen as “positive discrimination”.⁴²²

Several ongoing projects aim to improve existing discrimination situations in the transport sector.

Mobility services to rural and periphery areas are a big challenge. This coincides with the rapidly changing age structure in these areas, where people are getting much older on average. The MobiDig project in the region of Northern Bavaria in Germany aims to tackle these issues by improving mobility services in rural areas in order to increase social inclusion. The project led by five partner institutions (including the Technical University of Munich and the Fraunhofer Group for Supply Chain Services) intends to evaluate and promote new mobility concepts in order to provide efficient and sufficient transport services.⁴²³

Another issue in the transport sector seems to be gender equality. The systematic analysis of the situation based on big data allows identifying discriminatory practices and the reasons therefor. This is what several EU projects are aiming to do. They seek to make recommendations in order to improve the situation, such as implementing – as a starting point – information about the gender of workers in the transport sector in existing databases.⁴²⁴

⁴¹⁹ Massimo Ragnedda and Glenn Muschert, *The Digital Divide: The Internet and Social Inequality in International Perspective* (Routledge 2013); Monica Răileanu Szeles, 'New Insights from a Multilevel Approach to the Regional Digital Divide in the European Union' (2018) 42(6) Telecommunications Policy 452

⁴²⁰ Eurostat, 'Internet Access of Households, 2017' (Eurostat) <http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals> accessed 24 August 2018

⁴²¹ Petter Bae Brandtzaeg, Jan Heim and Amela Karahasanović, 'Understanding the New Digital Divide—A Typology of Internet Users in Europe' (2011) 69(3) International Journal of Human-Computer Studies 123; Petya Chipeva and others, 'Digital Divide at Individual Level: Evidence for Eastern and Western European Countries' (2018) 35(3) Government Information Quarterly 460

⁴²² 'Positive discrimination' is defined by the Oxford English Dictionary as “the practice or policy of favouring individuals belonging to groups known to have been discriminated against previously”.

⁴²³ Bundesministerium für Verkehr und digitale Infrastruktur, 'Mobilität digital Hochfranken – MobiDig' (BMVI) <<https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/mobilitaet-digital-hochfranken-mobidig.html?nn=326002>> accessed 24 August 2018

⁴²⁴ WISE 'Project Wise - Project Report: Women Employment in Urban Public Transport Sector' (WISE) <http://www.wise-project.net/download/final_wise_project_report.pdf> accessed 23 August 2018; Peter Turnbull, Julia Lear and Huw Thomas, 'Women in the Transport Sector - Promoting Employment by Preventing Violence against Women Transport Workers' (International Labour Organization 2013)



Illustration in the transport sector:

Uber, the ride-sharing company, has allowed making discrimination visible thanks to its online platform technologies. Several forms of discrimination have been observed in the Uber environment. The Uber rating system used by passengers to give feedback about drivers at the end of a ride has allowed highlighting discrimination against drivers from racial minority groups. This is problematic as the data collected via the tool are used to evaluate drivers, and eventually dismiss them if their ratings do not meet Uber's expected standards. Another form of discrimination concerns passengers. It has been observed that drivers are sometimes less keen to offer their services to riders willing to go to poorer neighbourhoods. Besides highlighting those discriminatory situations, the Uber platform could also be used to deter or prevent discrimination by for example configuring the level of passenger information available to drivers in order to decrease discrimination against them.⁴²⁵

Conclusion

Big data analytics can be a tool to make existing discriminatory decisions visible, hence this social issue may be resolved by personalised services (as “positive discrimination”) based on big data analytics. In spite of this opportunity, there are still biases because of big data's characteristics (e.g., heterogeneity, data size and quality, noise, etc.). Furthermore, also personalised services may cause discriminatory treatment by excluding certain groups. Finally, big data creates new visibilities and makes it possible to discern between people on a whole range of behaviour-related and other personal aspects. This also provides fertile ground for ‘new discriminations’.

These issues are of course highly relevant for the use of big data in the transport sector, for instance, for the planning of different routes on the basis of quality data or technologies used. Therefore, it is essential and important to reduce the likelihood of discrimination in the processing of big data and its analytics. In the same vein, “diversity, non-discrimination and fairness” has recently been listed by the High-Level Expert Group (“AI HLEG”) on Artificial Intelligence in their “Ethics Guidelines for Trustworthy AI” as one of the seven key requirements for realising trustworthy AI, to be implemented and evaluated throughout the AI system's lifecycle.⁴²⁶ Such Guidelines notably provide a self-assessment checklist in order to ensure that unfair bias creation or reinforcement is avoided and that accessibility, universal design principles, and stakeholder participation are considered.⁴²⁷

http://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---sector/documents/briefingnote/wcms_234882.pdf accessed 24 August 2018; Anne Loehr, 'Big Data for HR: Can Predictive Analytics Help Decrease Discrimination in the Workplace' (The Blog Huffpost 2015) <https://www.huffingtonpost.com/anne-loehr/big-data-for-hr-can-predi_b_6905754.html> accessed 24 August 2018

⁴²⁵ Alex Rosenblat and others, 'Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination' (2017) 9(3) Policy and Internet 256; Brishen Rogers, 'The Social Costs of Uber' (2015) (University of Chicago Law Review Dialogue, Forthcoming, Temple University Legal Studies Research Paper No. 2015-28) DOI: 10.2139/ssrn.2608017; Yanbo Ge and others, 'Racial and Gender Discrimination in Transportation Network Companies' (2016) NBER Working Paper No. w22776 <<http://www.nber.org/papers/w22776.pdf>> accessed 24 August 2018

⁴²⁶ High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy AI" (European Commission, 8 April 2019) <<https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>> accessed 15 April 2019

⁴²⁷ Ibid



Transparency, Consent, Control and Personal Data Ownership

In this seventeenth Chapter, we look into the social and ethical aspects of privacy, with a particular focus on transparency, consent and control, and personal data ownership in a big data context. This Chapter further elaborates, from an ethical perspective, the second and twelfth Chapters of this publication. Where relevant, illustrations from the transport sector will be provided.

Privacy is probably the most recurrent topic in the debate on ethical issues surrounding big data, which is not illogical given that the concepts of big data and privacy are *prima facie* mutually inconsistent.⁴²⁸ Indeed, the analysis of extremely large datasets may include personal data, and the more personal information included in the analytics, the more it might interfere with the privacy of the individuals concerned.⁴²⁹ In this context, the question of ownership over personal data is also raised, as individuals tend to have a sense of ownership over their personal data.

These aspects are also discussed in the recently published Ethics Guidelines to achieve trustworthy AI, issued by the Independent High-Level Group on Artificial Intelligence set up by the European Commission. The Guidelines list seven key requirements, including privacy and data governance and transparency, and suggest technical and non-technical methods to implement them. The Guidelines also provide an assessment checklist to ensure AI takes into account the ethical requirements.⁴³⁰

Setting the scene on privacy from an ethical perspective

The EU Charter of Fundamental Rights codifies the concept of privacy as a fundamental right in Article 7, according to which: "Everyone has the right to respect for his or her private and family life, home and communications."

Article 8 of the EU Charter provides specific fundamental rights and principles in relation to the protection of one's personal data in the following terms:

⁴²⁸ European Data Protection Supervisor, 'Opinion 4/2015 Towards a New Digital Ethics' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf> accessed 23 August 2018; European Data Protection Supervisor, 'Opinion 7/2015 Meeting the Challenges of Big Data: A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 23 August 2018; European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <<https://www.enisa.europa.eu/publications/big-data-protection>> accessed 23 August 2018; Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <<https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf>> accessed 23 August 2018

⁴²⁹ Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR [arXiv:1802.04103](https://arxiv.org/abs/1802.04103)

⁴³⁰ High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy AI" (European Commission, 8 April 2019) <<https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>> accessed 15 April 2019

- 1 *Everyone has the right to the protection of personal data concerning him or her.*
- 2 *Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.*
- 3 *Compliance with these rules shall be subject to control by an independent authority.*

The first Recital of the GDPR, which entered into force in May 2018, further elaborates Article 8 of the EU Charter. Nevertheless, Recital 4 of the GDPR clearly favours a balanced approach by stating that *"the right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality"*.

Ethical challenges and opportunities for privacy

After years of wilful abuse or unintentional ignorance in respect of people's personal data, the entry into force of the GDPR has increased the protection of individuals' personal data by obliging companies to abide by a strict set of rules. This Regulation addresses several ethical issues, including transparency, consent and control. The GDPR notably provides for the following:

- a strengthened principle of transparency in relation to personal data processing, ensuring better information to individuals about the processing of their personal data⁴³¹
- the requirement that any processing should be lawful, i.e. based on a legal ground⁴³²
- extended and strengthened rules on consent⁴³³
- new and reinforced rights for individuals aiming at giving individuals more control over their personal data, i.e. the rights of access, rectification, erasure, restriction of processing, data portability, objection and the right not to be subject to automated individual decision-making⁴³⁴

⁴³¹ GDPR, arts 5.1(a) and 12

⁴³² GDPR, art 6

⁴³³ GDPR, art 7

⁴³⁴ GDPR, Chapter III, art 12-22

The GDPR has raised the public's awareness in relation to privacy and data protection, which should improve end-users' trust in the use of personal data by private and public organisations. This may encourage them to communicate their personal data, and therefore improve big data analytics.⁴³⁵ This development can be seen as an opportunity by companies to guarantee high data protection standards and distinguish themselves from their competitors, particularly in a big data context where considerable amounts of data may be processed.

Although this can be qualified mostly as a positive evolution, it has also had some undesirable side effects, mainly due to incorrect reports on the GDPR's exact content, creating confusion both among data subjects and organisations. This is for example the case for the data subject rights, which are often considered as being absolute whereas in some conditions data subjects will not be able to exercise those rights. Both industry and government should take up responsibility to eliminate the existing misconceptions and educate data subjects about privacy and big data analytics in order to encourage the use of big data.

Furthermore, even though the GDPR is now applicable throughout the EU as one single set of rules, the expectations regarding privacy may vary between individuals or situations.⁴³⁶ It will therefore be difficult for companies and developers to adopt a one-size-fits-all approach, with the risk of opting for the strongest protection and therefore limiting big data analytics using personal data.

Transparency



The concept of transparency is indirectly included in Article 8 of the EU Charter, which states that *"Everyone has the right of access to data which has been collected concerning him or her"*. This entails that individuals have the right to be informed about any processing activities of their personal data.⁴³⁷ In a big data context, this also refers to the transparency

⁴³⁵ Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <<https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf>> accessed 23 August 2018; Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

⁴³⁶ World Economic Forum in collaboration with Bain & Company, Inc., 'Personal Data – The Emergence of a New Asset Class' (World Economic Forum 2011) <http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf> accessed 23 August 2018

⁴³⁷ EDPS, Opinion 7/2015

of the big data analytics, i.e. the entire ecosystem of big data analytics, the algorithms used to make predictions about individuals, and the decision-making process.⁴³⁸

Transparency regarding personal data processing activities and big data analytics may increase individuals' trust in the processing activities and the technology used. Moreover, it also ensures safer tools as transparency allows individuals to verify the conclusions drawn and correct mistakes.⁴³⁹

Today individuals' trust is however negatively affected by a lack of transparency, particularly in a big data environment.⁴⁴⁰ Individuals are indeed not always aware of the exact nature of the processing activities and of the logic of algorithms and the decision-making process behind big data analytics.⁴⁴¹ This challenge is even more important considering citizens' limited knowledge about big data analytics⁴⁴², particularly the possibility to combine individuals' personal data with other accessible data, allowing to make more accurate and broader decisions or predictions.⁴⁴³

From the perspective of organisations, transparency is also a challenge in the sense that some of them are reluctant to be transparent, invoking business confidentiality or trade secrets protection. In this respect, it is worth noting that other means of protection of information exist, such as intellectual property rights (see the ninth Chapter Intellectual property rights).⁴⁴⁴

⁴³⁸ EDPS, Opinion 7/2015; Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR [arXiv:1802.04103](https://arxiv.org/abs/1802.04103)

⁴³⁹ EDPS, Opinion 7/2015

⁴⁴⁰ Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

⁴⁴¹ European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <<https://www.enisa.europa.eu/publications/big-data-protection>> accessed 23 August 2018; EDPS, Opinion 7/2015; Article 29 Data Protection Working Party, 'Opinion 3/2013 on purpose limitation' (WP29 2013) <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> WP203 accessed 23 August 2018

⁴⁴² Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <<https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf>> accessed 23 August 2018

⁴⁴³ EDPS, Opinion 7/2015

⁴⁴⁴ Ibid

Consent



The concept of consent has been foreseen in Article 8 of the EU Charter stating that "Such [personal] data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law." This means that any processing of personal data should be based on individuals' consent or on another legitimate ground.

The GDPR defines consent as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."⁴⁴⁵

Collecting individuals' consent does not mean that the organisations processing the data are free to process the data as they wish. They are still accountable and have to meet the privacy standards (ethical, legal, etc.).⁴⁴⁶ It is also worth noting that if individuals have given their consent for a particular personal data processing activity, they also have the right to withdraw their consent.

As explained in the second Chapter Privacy and Data Protection, the GDPR requires all data processing activities to be lawful, i.e. based on a legal ground⁴⁴⁷, which means that, from a legal perspective, consent is not always needed and other legal grounds might be applied.⁴⁴⁸ This is another misconception of the GDPR, highlighting the lack of awareness and transparency observed among individuals.⁴⁴⁹ By informing individuals, notably through transparent notices, about the grounds for processing and the possible impacts on their privacy, they will indeed be more inclined to participate in big data analytics.⁴⁵⁰

⁴⁴⁵ GDPR, art 4(11)

⁴⁴⁶ EDPS, Opinion 4/2015

⁴⁴⁷ GDPR, art 5(1)(a)

⁴⁴⁸ GDPR, art 6. Article 9 of the GDPR also provides legal grounds specific to the processing of special categories of personal data (i.e., data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data, health data, data concerning a natural person's sex life or sexual orientation).

⁴⁴⁹ Elizabeth Denham, 'Blog: Consent is not the "Silver Bullet" for GDPR Compliance' (Information Commissioner's Office 2017) <<https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/08/blog-consent-is-not-the-silver-bullet-for-gdpr-compliance/>> accessed 24 August 2018

⁴⁵⁰ Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <<https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf>> accessed 23 August 2018

It is worth noting that relying on consent in the context of big data analytics may be risky, given that when an individual decides to withdraw consent, as foreseen in the GDPR⁴⁵¹, the big data analytics process may be completely jeopardised.



Illustration in the transport sector:

Self-driving cars will collect a high amount of personal data about the users but also about the environment of the car (neighbourhood, other drivers, etc.), and those data may be shared with many stakeholders. Users might be reluctant to give their consent to such massive processing of their personal data. However, without such processing activities, self-driving cars would not work properly and safely. Indeed, a high amount of partners will be involved in such ecosystem to make it function. It might be that individuals will have no other choice than to accept such processing.⁴⁵² From a legal perspective, the GDPR introduces different lawful bases for processing (see the second Chapter Privacy and Data Protection).

Control



The concept of control is implied in Article 8 of the EU Charter, particularly when referring to the "*consent of the person concerned*", "*the right of access to data which has been collected*", and "*the right to have it rectified*". Several aspects of the GDPR, such as transparency, consent, and data subjects' rights, also allow individuals to retain control over their personal data, including in a big data environment.

Today, there is an asymmetry of control over personal data between data subjects and the organisations processing the data.⁴⁵³ In a big data context, individuals indeed hardly control their personal data, and are sometimes not aware of the processing activities in which their data are involved, which may lead to decisions that individuals do not understand.⁴⁵⁴ In addition, data subjects may fear losing control over their digital identity by engaging in big data analytics as they are

not consulted anymore, nor taken into account in the decision-making process, which means that they might be discriminated without having the possibility to react.⁴⁵⁵

This is why giving more control to individuals, and ensuring transparency, should improve big data analytics, by allowing them to rectify mistakes, detect unfair decisions, and make better choices.⁴⁵⁶ In this way, they will benefit from the processing of their personal data in a big data context, and therefore feel more inclined to participate in data processing activities for big data purposes.



Illustration in the transport sector:

Civil drones collect data intentionally and unintentionally, especially pictures about individuals, which can give indications about their location, habits, physical characteristics, etc. In their survey about the use of civil drones and their related privacy, data protection and ethical implications, Finn and Wright explain that in some instances the images captured by drones are recorded, stored and shared with other organisations. Individuals are not aware of such processing and have therefore no control over their data. According to Finn and Wright, awareness and legal initiatives are necessary to improve knowledge about legal and ethical standards in order to be able to raise and tackle those issues.⁴⁵⁷

⁴⁵¹ GDPR, art 7(3)

⁴⁵² Caitlin A. Surakitbanharn and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles'

<https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf> accessed 23 August 2018

⁴⁵³ Melanie Swan, 'Philosophy of Big Data: Expanding the Human-data Relation with Big Data Science Services' in '15 Proceedings of the 2015 IEEE First International Conference on Big Data Computing Service and Applications (IEEE, 2015) 468-477 DOI: [10.1109/BigDataService.2015.29](https://doi.org/10.1109/BigDataService.2015.29)

⁴⁵⁴ EDPS, Opinion 7/2015; WP29, Opinion 3/2013

⁴⁵⁵ EDPS, Opinion 7/2015; Norwegian Data Protection Authority, 'Big Data Report' (Datatilsynet 2013) 7 <<https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf>> 23 August 2018; Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <<https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf>> accessed 23 August 2018

⁴⁵⁶ Ibid

⁴⁵⁷ Rachel L. Finn and David Wright, 'Privacy, Data Protection and Ethics for Civil Drone Practice: A Survey of Industry, Regulators and Civil Society' (2016) 32(4) Computer Law & Security Review 577

Setting the scene on personal data ownership

For some time already, the issue of ownership of data (whether it is personal or non-personal) has been heavily debated throughout the EU and in other parts of the world. While it could be labelled as a legal issue, given that ownership or property is traditionally a legal concept going back as far as the legal system of ancient Rome (see the twelfth Chapter Data ownership), the personal aspect of data ownership has an ethical connotation that is worth being looked into.

The EU Charter recognises the right to property or ownership in its Article 17 in the following terms:

"Everyone has the right to own, use, dispose of and bequeath his or her lawfully acquired possessions. No one may be deprived of his or her possessions, except in the public interest and in the cases and under the conditions provided for by law, subject to fair compensation being paid in good time for their loss."

Individuals seem to have a general sense that they own their personal data given that the data is about them or relates to them.⁴⁵⁸ Moreover, where the personal data is particularly sensitive in nature, individuals even more vehemently tend to claim it as their own.

'Personal data' is defined by Article 4(1) of the GDPR as "*any information relating to an identified or identifiable natural person ('data subject')*", whereas an 'identifiable natural person' is defined as "*one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*".

As explained above, the entry into force of the GDPR has increased the control individuals have over the collection, processing, and sharing of their personal data.⁴⁵⁹ This evolution seems to create a certain impression of personal data ownership. For instance, some scholars highlight the fact that the

⁴⁵⁸ World Economic Forum in collaboration with Bain & Company, Inc., 'Personal Data – The Emergence of a New Asset Class' (World Economic Forum 2011) 16 <http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf> accessed 23 August 2018

⁴⁵⁹ Alex Howard, *Data for the Public Good* (O'Reilly Media Inc. 2012) 23

GDPR "*recognises different levels of control rights to consumers in accordance with a 'proprietary' approach to personal data.*"⁴⁶⁰ More specifically, some have emphasised that in practice personal data is perceived as an individual's property.⁴⁶¹

Challenges and opportunities for personal data ownership

Even if the GDPR and some EU Member States' laws grant important rights to data subjects, they do not regulate the question of data ownership and therefore do not explicitly recognise a "property" right of individuals in their data. In our view, the GDPR only regulates the relationship between the data subject and the data controller(s)/processor(s), without creating and regulating the issues of commercially exploitable rights in personal data.⁴⁶²

This view is supported by the manner in which the right to property is recognised in the EU Charter; i.e. *the right to own [...] his or her lawfully acquired possessions*. Personal data is not a possession that can be acquired by the data subject, be it lawfully or not. It is information that attaches to an individual because of his/her persona. Consequently, personal data protection is not conditional upon an act of acquisition on behalf of the data subject. To claim otherwise would go against the data protection principles of the GDPR and the rights to respect for private and family life and to protection of personal data enshrined in the EU Charter.

Whereas personal data is something inherent to and indivisible from the individual, it may be lawfully – i.e. in compliance with the data protection rules – acquired by third parties, either directly from the data subject or through other sources. Such interpretation would fit within the definition of the right to property under the EU Charter. This being said, any such "ownership" right subsisting in personal data to the benefit of third-party natural or legal persons, would be

⁴⁶⁰ Gianclaudio Malgieri, 'Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data' (2016) 4 PinG 133; Jacob M. Victor, 'The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy' (2013) 123(2) Yale Law Journal 266

⁴⁶¹ Nadezhda Purtova, 'The Illusion of Personal Data as No One's Property' (2015) 7(1) Law, Innovation and Technology 83

⁴⁶² European Data Protection Supervisor, 'Ethics Advisory Group Report 2018' (EDPS 2018) 25 <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf> 23 August 2018

restricted by the application of the GDPR and notably by the rights of data subjects.⁴⁶³

In a big data ecosystem, this tension between data subjects wanting to "own" their personal data and third parties claiming ownership over entire datasets could stifle innovation. Indeed, as long as data subjects do not volunteer their personal data, they retain some type of *de facto* ownership or at least control. Therefore, data subjects may refrain from providing their personal data as soon as they realise this would entail forsaking "ownership" or control over such data. In addition, even if data subjects willingly provide their personal data, it proves highly difficult, if not impossible, to establish ownership of different data components, given that they are part of datasets containing data from various types and originating from various sources. Furthermore, taking into account the various actors involved in the big data ecosystem, many different entities may try to claim ownership in (parts of) the dataset, including in the personal data components.

An additional complicating factor is that the scope of what can be qualified as personal data is constantly evolving.⁴⁶⁴ Certain types of information (e.g. IP addresses) that would not necessarily have been qualified as personal data under the previous Data Protection Directive, are now recognised to be personal data under the GDPR. This is not only due to the fact that the legal definition of personal data has been broadened, but also because of continuous technological developments facilitating the identification or linking back to an individual.

In conclusion, a claim of ownership by a data subject in its personal data would be hard to sustain. This however does not mean that data subjects have to give up all control over their personal data. The advent of the GDPR, with its novel and/or strengthened data subject rights, has increased the means of data subjects to exercise control over the processing of their personal data.

⁴⁶³ See in the same vein, in the context of disclosure of chemical data, the Court Order of the EU General Court in case T-189/14 wherein the President examines in obiter dictum the question of privacy (Case T-189/14 R *Deza a.s v Agence européenne des produits chimiques* [2014] ECLI:EU:T:2014:686). More particularly, the President acknowledges the relevance of the question of privacy of legal entities but nevertheless reminds, on the basis of the decision of the Court of Justice of the EU in case C-450/06, that it may be necessary to prohibit the disclosure of information qualified as confidential in order to preserve the fundamental right to privacy of an undertaking (Case C-450/06 *Varec SA v État belge* [2008] ECLI:EU:C:2008:91).

⁴⁶⁴ Václav Janeček, 'Ownership of Personal Data in the Internet of Things' (2018) forthcoming in the *Computer Law & Security Review*



Looking beyond



In this concluding Chapter, we want to start looking beyond the issues and opportunities that were identified in the 17 chapters that make up the publication.

Over the course of 17 chapters, we have presented a summary of the findings from our research conducted in the LeMO Project concerning legal, ethical and social challenges and opportunities pertaining to big data. Where relevant, the chapters have been illustrated with examples from the transport sector, which is the focus of the LeMO Project. These findings had also been published in two reports, namely the 'Report on Legal Issues' and the 'Report on Ethical and Social Issues'. Both are available online at www.lemo-h2020.eu/deliverables/.

Chapter overview



Key questions raised and addressed throughout the publication included questions such as:

- *"do the privacy concepts of the GDPR fit with big data?"*;
- *"can anonymisation techniques be applied while keeping an acceptable level of predictability and utility of big data analytics?"*;
- *"is the current legal framework in relation to data ownership satisfactory?"*;
- *"what are the main areas in which competition law may have an impact on the use of big data?"*; and also
- *"can social differences in access to technology and education or skills lead to data-driven discrimination?"*.

Through the numerous questions raised and addressed throughout the publication, various legal, ethical and social issues and opportunities were identified in relation to big data. The paragraphs below offer a succinct summary of the various topics covered by each chapter.

1 General overview (p.6): In addition to introducing the publication, this chapter provides some

background information with respect to big data in the transport sector, useful to bear in mind while reading the other chapters.

2 Privacy and data protection (p.10): The second chapter focuses on some of the key privacy and data protection aspects in a big data context, showing how certain principles and requirements can be difficult to fit with some of the main characteristics of big data analytics. The chapter demonstrates that finding a balance between the various interests at stake is of paramount importance. In light hereof, it is essential to keep in mind that the right to the protection of personal data is not an absolute right, but must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality. Any guidance or administrative/judicial decision should carefully take into account all interests at stake as failing to do so would necessarily impede the development of disruptive technologies and prohibit the emergence of a true data economy.

3 Anonymisation / pseudonymisation (p.17): This chapter looks into the impact of anonymisation and pseudonymisation in a personal data protection context, and the possible use of anonymisation and pseudonymisation techniques as a way to protect non-personal data. Anonymisation and pseudonymisation techniques generally provide fertile ground for opportunities with respect to big data applications. Nevertheless, account must be taken of the challenges that may arise in this respect. Most importantly, a balance will need to be struck between, on the one hand, the aspired level of anonymisation (and its legal consequences) and, on the other hand, the desired level of predictability and utility of the big data analytics.

- 4 (Cyber-)security (p.24): Considering the increasingly devastating impact that cyber-threats and attacks may have on society, issues related to cyber-security have become increasingly important in recent years. The requirement to put in place security measures is imposed in various legislations at EU and national level, including key instruments like the GDPR and the NIS Directive. Such legislations however remain rather general and vague as to which specific measures are deemed appropriate. In order to comply with the relevant requirements, organisations generally need to rely on security experts and take into account the evolving guidance documents published by authorities such as ENISA. Also, relying on certification mechanisms, seals, marks and codes of conduct will enable companies to comply with their legal obligations in terms of security and demonstrate their compliance.
- 5 Breach-related obligations (p.31): In recent years, the EU has made significant progress in terms of cybersecurity and related incident notification requirements, with notable developments including the Cyber Security Strategy and the NIS Directive. It follows that organisations facing a security incident may need to notify such incident to one or more national competent authorities. The requirement to inform authorities will however depend on certain criteria laid down in the applicable legislations, as clarified by the guidance documents published at EU and national level. Accordingly, the various actors of the data value chain need to implement measures, procedures and policies in order to abide by the strict notification requirements and be prepared to provide the necessary information to the competent authorities, all within the imposed deadlines.
- 6 Supply of digital content and services (p.36): This chapter looks into the possible provision of personal data by a consumer in order to receive digital content. It assesses how this practice is dealt with in the recently adopted Digital Content Directive and looks into its interaction with the applicable data protection legislation, and in particular the GDPR. As demonstrated through this chapter, legalising this economic reality generates practical and legal concerns. Accordingly, clarifications and guidelines are necessary to allow a greater degree of predictability for digital market actors and to ensure the usefulness of big data.
- 7 Free flow of data (p.40): Free flow of data represents an ideal scenario in which no (legal) barriers to cross-border data flows remain. Efforts have been taken at EU level with the adoption of the Regulation on the free flow of non-personal data. A number of uncertainties remain, including a difficult interaction with the GDPR. Still, the Regulation remains an important step in the elimination of restrictions to cross-border data flows and their negative impact on business. Companies expect cost reductions to be the main benefit of eliminating data localisation requirements. Furthermore, start-ups in the European transport sector and in the EU in general increasingly rely on competitive cloud services for their products or services. Prohibiting localisation restrictions would therefore increase competitiveness of the EU cloud services market. This in turn could allow start-ups to go to market quicker, to increase their pace of innovation and would also support scalability and achieve economies of scale.
- 8 Liability (p.46): The EU institutions have been engaged in ongoing work regarding extra-contractual and statutory liability in the context of disruptive technologies. On such basis, it will be possible to determine whether regulatory intervention is required. In all likelihood, intervention should take place in two phases. In the short- and mid-term, non-regulatory intervention, such as the creation of model contract clauses or the identification of appropriate safety standards, should be pursued. In the long term, regulatory intervention should be considered in the form of sector-specific legislation on minimum liabilities to be borne by certain service providers in certain sectors, a general revision of liability law, and/or legislation on insurance-related obligations. Nonetheless, this chapter has shown that the current status of contractual liability rules, which may differ across the EU, is likely to limit the uptake of new technologies, including big data in the transport sector.
- 9 Intellectual property rights (p.52): This ninth chapter examined the aspects related to copyright, database rights and trade secrets and in particular to what extent such protection mechanisms can apply to (big) data. In this respect, it cannot be excluded that different actors in the big data analytics lifecycle will try to claim intellectual property rights or protection under trade secrets in (parts) of datasets intended to be used. These actors may try to exercise the

exclusive rights linked to the intellectual property right concerned or keep the information secret. Any unreasonable exercise of rights may stifle data sharing and thus innovation through big data, including in the transport sector. This is however mainly due to the inherent nature and purpose of intellectual property rights and trade secrets protection, which may at the same time provide an incentive for stakeholders to engage in data sharing for big data purposes.

10 Open data (p.59): The 'big data' required to feed big data analytics tools typically emanates from a variety of sources. One such source is the public sector, which has been opening up certain of its datasets to the public. The EU institutions have taken both legislative and non-legislative measures to encourage the uptake of open data, most notably through the PSI Directive, which attempts to remove barriers to the re-use of public sector information throughout the EU. Still, open data regimes also encounter a number of challenges – on a technical, economic and legal level – that cannot be ignored. The proposal for a recast of the PSI Directive aims to address some of these concerns. A major change concerns the expansion of the Directive's scope to include public undertakings. While information sharing has not been made mandatory for public undertakings (yet), the new regime constitutes a significant development for the transport sector, where services are often provided by public undertakings.

11 Data sharing obligations (p.65): This chapter addresses those legal instruments that impose specific data sharing obligations on private undertakings and therefore affect a company's control of, access to, or use of data. Such legislations are usually sector-focused and provide for an array of rights and obligations in relation to specific types of data in particular circumstances. The chapter offers a succinct examination of those pieces of legislation imposing data sharing obligations that are most relevant to the transport sector, showing that data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems. The EU should however carefully consider whether the imposition of such general data sharing obligations is in each case equally necessary.

12 Data ownership (p.71): If the numerous stakeholders in the (big) data analytics lifecycle cannot rely on any of the other exclusive rights discussed in this publication, they increasingly try

to claim "ownership" in (parts of) the datasets used in the analytics. No specific ownership right subsists in data and the existing data-related rights do not respond sufficiently or adequately to the needs of the actors in the data value cycle. Up until today, the only imaginable solution is capturing the possible relationships between the various actors in contractual arrangements. Nevertheless, we found that filling the data ownership gap with contractual arrangements is far from ideal.

13 Data sharing agreements (p.77): A critical analysis is made of the current-day common practice to use data sharing agreements to govern the access to and/or exchange of data between stakeholders in a big data analytics lifecycle. It is unclear, however, whether such practice enables covering all possible situations with the necessary and satisfactory legal certainty. Indeed, data sharing agreements entail numerous limitations in the absence of a comprehensive legal framework regulating numerous rights (e.g. ownership, access or exploitation rights) attached to data, the way in which such rights can be exercised, and by whom. While guidance has been issued by the European Commission recently, a more solid and legally secure solution might be desirable.

14 Competition (p.84): The final chapter addressing legal issues and/or opportunities in the context of big data in transport focuses on the impact of big data on different aspects of EU competition law and seeks to create more clarity on when and how the so-called ownership or (mis)use of (big) data can give rise to competition law issues. As such, big data aggregation in the transport sector can give rise to a variety of competition law issues that suggest that certain aspects of competition law may not be fit for purpose. Abuse of dominance, merger control and anticompetitive behaviour have all seen challenges in the face of big data, AI and digitisation. The recent public consultation on shaping competition policy in the age of digitisation has yielded some interesting insights on how to mould competition law to address these topical issues.

15 Trust, surveillance and free will (p.88): Moving away from legal issues and opportunities and into ethical and social aspects, the first ethical and social concepts examined in the context of big data and transport are those of trust, surveillance and free will. One of the main

dimensions of big data, describing consistency and trustworthiness, is veracity. In this respect, big data may present challenges in relation to its quality (e.g. heterogeneous and unstructured data). It can however also be used for trust assessment, including through so-called reputation systems. In relation to surveillance, two main issues arise, namely risks of asymmetries in the control over information on the one hand and privacy risks on the other hand. With respect to supporting free will of humans, increasing accessibility and personalisation for passengers can provide benefits to people in the form of more personalised or affordable services. Organisations use certain types of data like journey data to ensure a better understanding and serving of people's needs.

16 Discrimination (p.94): (Data-driven) discrimination is a particular social and ethical issue that may materialise in a big data context and is therefore addressed in a separate chapter. Big data analytics can be a tool to make existing discriminatory decisions visible, hence this social issue may be resolved by personalised services (as “positive discrimination”) based on big data analytics. In spite of this opportunity, there are

still biases because of the inherent characteristics of big data (e.g., heterogeneity, data size and quality, noise, etc.).

17 Transparency, consent, control and personal data ownership (p.99): Privacy is probably the most recurrent topic in the debate on ethical issues surrounding big data, which is not illogical given that the concepts of big data and privacy are *prima facie* mutually inconsistent. Indeed, the analysis of extremely large datasets may include personal data, and the more personal information included in the analytics, the more it might interfere with the privacy of the individuals concerned. In this context, the question of ownership over personal data is among others raised, as individuals tend to have a sense of ownership over their personal data. While a claim of ownership by a data subject in its personal data would be hard to sustain (given that legally no specific ownership rights subsist in data), this does not mean that data subjects have to give up all control over their personal data, particularly with the advent of the GDPR.

Authors

Benoit Van Asbroeck

Partner, IP/IT

Tel: +32 (0)2 282 6067
Benoit.Van.Asbroeck@twobirds.com



Simon Mortier

Associate, IP/IT

Tel: +32 (0)2 282 60 82
Simon.Mortier@twobirds.com



Julien Debussche

Senior Associate, IP/IT

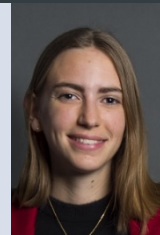
Tel: +32 (0)2 282 60 44
Julien.Debussche@twobirds.com



Charlotte Haine

Associate, IP/IT

Tel: +32 (0)2 282 60 87
Charlotte.Haine@twobirds.com



Jasmien César

Associate, IP/IT

Tel: +32 (0)2 282 6045
Jasmien.Cesar@twobirds.com



Brona Heenan

Senior Associate, EU & Competition

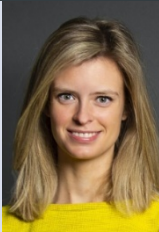
Tel: +32 (0)2 282 60 05
Brona.Heenan@twobirds.com



Isis De Moortel

Associate, IP/IT

Tel: +32 (0)2 282 60 85
Isis.DeMoortel@twobirds.com



Anthony Benavides

Trainee, EU & Competition

twobirds.com

Abu Dhabi & Amsterdam & Beijing & Bratislava & Brussels & Budapest & Copenhagen & Dubai & Dusseldorf & Frankfurt & The Hague & Hamburg & Helsinki & Hong Kong & London & Luxembourg & Lyon & Madrid & Milan & Munich & Paris & Prague & Rome & San Francisco & Shanghai & Singapore & Stockholm & Sydney & Warsaw

Bird & Bird is an international legal practice comprising Bird & Bird LLP and its affiliated and associated businesses. Bird & Bird LLP is a limited liability partnership, registered in England and Wales with registered number OC340318 and is authorised and regulated by the Solicitors Regulation Authority. Its registered office and principal place of business is at 12 New Fetter Lane, London EC4A 1JP. A list of members of Bird & Bird LLP and of any non-members who are designated as partners, and of their respective professional qualifications, is open to inspection at that address.