LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable D2.2 **Report on Legal Issues**

Julien Debussche, Jasmien César, Isis De Moortel

Bird & Bird

October 2018

Work Package 2

Project Coordinator

Prof. Dr. Rajendra Akerkar (Western Norway Research Institute)

Horizon 2020 Research and Innovation Programme

MG-8-2-2017 - Big data in Transport: Research opportunities, challenges and limitations

Distribution level	Public (P)
Due date	31/10/2018 (M12)
Sent to coordinator	29/10/2018
No. of document	D2.2
Title	*Report on Legal Issues*
Status & Version	*Final*
Work Package	*2: Institutional and governmental issues and barriers*
Related Deliverables	*D2.1, D2.3, D2.4*
Leading Partner	*Bird & Bird*
Leading Authors	*Julien Debussche, Bird & Bird, all Chapters* *Jasmien César, Bird & Bird, all Chapters* *Isis De Moortel, Bird & Bird, all Chapters*
Contributors	*Benoit Van Asbroeck, Bird & Bird, all Chapters* *Jehan De Wasseige, Bird & Bird, Sections on Privacy and Data Protection and (Cyber-)Security* *Brona Heenan, Bird & Bird, Section on Competition* *Anthony Benavides, Bird & Bird, Section on Competition* *Marianna Rantou, Bird & Bird, Section on Competition* *Charlotte Haine, Bird & Bird, Section on Supply of digital content* *Maruša Benkic, CORTE, illustrations on Anonymisation and Pseudonymisation* *Ivo Hindriks, Panteia, illustrations on Privacy and Data Protection*
Reviewers	*Kim Hee, GUF*
Keywords	*Big data, Transportation, Legal issues*

Disclaimer:

This report is part of the LeMO project which has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 770038.

The content of this report reflects only the authors' view. 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.

Executive summary

This Deliverable identifies and examines various legal issues that are relevant to the production of, access to, linking of and re-use of big data in the transport sector.

Chapter 2 sets the scene and introduces the concept of big data, its particular characteristics, its possible use in the transport sector, the existing policy framework, and the identified legal issues.

In Chapter 3, the authors examine the various identified legal issues and discuss the challenges and opportunities that may arise in this respect, coming up with notably the following findings:

- **Privacy and data protection**: Some concepts, principles and obligations under data protection law appear to be problematic for the uptake of big data. In particular, the broad definition of "personal data" and "processing", the qualification of the various actors involved as (joint-)controllers or processors, the core data protection principles, the need to identify a ground for processing, the requirement to conduct data protection impact assessments, the implementation of privacy by design and by default measures, the rights of data subjects, and the requirement to put in place adequate data transfer mechanisms seem difficult to reconcile with the concept of big data.
- **(Cyber-)Security**: 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. However, such legislative framework remains rather general and vague as to which specific measures are deemed appropriate. In order to comply with this requirement, organisations involved in big data analytics generally need to rely on security experts and take into account the evolving guiding documents published by authorities such as ENISA. Relying on certification mechanisms, seals, marks, and codes of conduct will enable companies complying with their legal obligations and demonstrate their compliance.
- **Breach-related obligations**: The various actors of the (big) data value chain need to implement measures, procedures and policies to abide by the strict notification requirements and be prepared to provide the necessary information to the authorities, 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.
- **Anonymisation / pseudonymisation**: Anonymisation and pseudonymisation techniques generally provide fertile ground for opportunities with respect to big data applications, including in the transport sector. Nevertheless, 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.

- **Supply of digital content and services (personal data as counter-performance)**: Personal data as a form of payment for the supply of digital content is an emerging reality. In this respect, the proposed EU legal framework on the supply of digital content and services will ensure an adequate level of protection for the consumer. Nevertheless, the obligations concerning data may make some current digital services inoperable. Some companies may also start to charge for digital content services that are currently free. On a wider scale the ecosystem of innovative services in the field of transport could be jeopardised.
- **Free flow of data**: The free flow of data presents a scenario in which no legal barriers hinder the cross-border flow of data. Such cross-border data flows may be restricted by data localisation requirements, which come in many shapes and forms. The new EU Free Flow Regulation should ensure the free flow of data across EU Member States, ensure data availability for regulatory control by EU authorities, and encourage the creation of codes of conduct for cloud services. The elimination of data localisation requirements is expected to create more innovation, which will positively impact big data analytics in the transport sector.
- **Intellectual property in big data environment**: All intellectual property rights examined may have, to some extent, an impact on the use of big data, including in the transport sector. Depending on the manner in which and the extent with which a right holder may exercise its exclusive rights attached to the intellectual property right concerned, intellectual property rights may pose a barrier to data access, interoperability, and exploitation.
- **Open data**: 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. Open data is a key component of most big data applications. A proposal for a revision of the PSI Directive intends to extend the scope of application to public undertakings, including actors in the transport sector such as ports and airports, public passenger transport services by rail and by road, and air carriers and EU ship owners fulfilling public service obligations.
- **Sharing obligations**: 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. This Deliverable succinctly examines the body of legislation specific to the transport sector that could impact a company's control of, the access to, or the rights in data. The analysis has shown that data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems.
- **Data ownership**: 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 Deliverable demonstrates however 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, i.e. data sharing agreements.

- **Data sharing agreements**: It is unclear whether 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 enables covering all possible situations with the necessary and satisfactory legal certainty. 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.
- **Liability**: The EU institutions have looked into and continue to examine issues related to extra-contractual liability, statutory liability, and safety requirements in the context of disruptive technologies, including in the transport sector. Based on their continued efforts, it will be possible to determine whether any regulatory intervention is required. The contractual liability legal framework, which differs across the EU, may however limit the uptake of new technologies, including big data in the transport sector. The present Deliverable further looks into the relevance for big data in the transport sector of the exemption of liability for intermediaries (the so-called safe harbour regime), and the proposed liability regime for suppliers of digital content and services under the Draft Directive on the Supply of Digital Content.
- **Competition**: 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. The present Deliverable considers three main areas in which competition law may have an impact on the use of big data. In view of the important role of big data in the transport sector, the Deliverable discusses the competition law issues that could arise with respect to organisations belonging to the broadly-defined "transport sector".

Finally, the last Chapter serves as a conclusion and introduces possible ways of moving forward to encourage the production of, access to, linking of and re-use of big data in the transport sector, with a particular focus on the EU. The several improvements suggested by this Deliverable vary between the different legal issues and range from avoiding restrictive interpretations by the relevant authorities or courts, over soft law measures (such as guidelines and codes of conduct), to regulatory intervention at EU level.

Table of contents

Executive summary ..... **II**
List of Figures ..... **XIV**
List of Tables ..... **XV**
Glossary ..... **XVII**
1 Introduction **1**
1.1 Abstract 1
1.2 Purpose of the document 1
1.3 Target audience 2
1.4 Methodology 3
1.4.1 Doctrinal method *3*
1.4.2 Fundamental and non-fundamental research constraints *3*
1.4.3 Sources of law *4*
2 Setting the scene **5**
2.1 The concept of "big data" 5
2.2 Big data in the transport sector 6
2.3 Policy framework 7
2.4 Assigning responsibilities 7
2.5 Identifying legal issues related to big data in the transport sector 8
3 Analysis of key legal issues **9**
3.1 Privacy and data protection 9
3.1.1 The concept of "personal data" *10*
3.1.1.1 Types of data 10
3.1.1.2 Sensitive Personal Data 10
3.1.1.3 Taxonomy of types of data 11
3.1.1.4 The “Processing” of Personal Data 12
3.1.1.5 The “processing” of “personal data” in the transport sector 13
3.1.2 Various actors, roles and responsibilities *14*
3.1.2.1 Data controller 14
3.1.2.2 Data processor 15

3.1.2.3 Sub-processors 16
3.1.2.4 Allocation of responsibilities in the context of big data and transport 17
3.1.3 Data protection principles 19
3.1.3.1 Lawfulness, fairness & transparency 19
3.1.3.2 Purpose limitation (and secondary use) 20
3.1.3.3 Data minimisation 24
3.1.3.4 Accuracy 25
3.1.3.5 Storage limitation 25
3.1.3.6 Accountability 26
3.1.4 Legal grounds to process personal data (lawful processing) 27
3.1.4.1 Overview of the legal grounds 27
3.1.4.2 Consent of the data subject 28
3.1.4.3 Processing necessary for the performance of a contract with the individual or to perform pre-contractual obligations 31
3.1.4.4 Processing necessary for compliance with a legal obligation to which the controller is subject 32
3.1.4.5 Processing necessary for the purposes of the legitimate interests of the controller 33
3.1.5 Core obligations under the GDPR 36
3.1.5.1 Overview of the core obligations 36
3.1.5.2 Data protection impact assessment (DPIA) 37
3.1.5.3 Data protection by design and data protection by default 42
3.1.6 Rights of individuals 45
3.1.6.1 Information 46
3.1.6.2 Access 48
3.1.6.3 Rectification 49
3.1.6.4 Erasure ("right to be forgotten") 49
3.1.6.5 Restriction 50
3.1.6.6 Portability 50
3.1.6.7 Objection 52
3.1.6.8 Profiling and automated decision-making 53
3.1.7 Data transfers 58
3.1.8 Summary 59
3.2 (Cyber-)Security 64
3.2.1 Security requirements under the GDPR 64
3.2.1.1 Personal data governance obligations 64

3.2.1.2 Security of personal data processing 65
3.2.2 Security requirements under the Network Information Security Directive 66
3.2.2.1 Context 66
3.2.2.2 Scope of Application of the NISD 69
3.2.2.3 Security requirements under the NIS Directive 72
3.2.3 Security requirements under other legislations 74
3.2.4 Security Standards 75
3.2.5 Security in practice: challenges and recommendations 75
3.2.6 Summary 81
3.3 Breach-related obligations 83
3.3.1 Breach Notification obligations in the telecommunications sector 83
3.3.2 Data and privacy breach notification obligation 84
3.3.2.1 Scope of the obligation 84
3.3.2.2 Notifications in practice 87
3.3.3 Notification obligation under the NISD 89
3.3.3.1 Notification in practice 89
3.3.3.2 Sanctions 92
3.3.4 Incident Response Plan 92
3.3.5 Summary 92
3.4 Anonymisation and Pseudonymisation 93
3.4.1 Introduction to the Key Concepts of Anonymisation and Pseudonymisation 93
3.4.2 Anonymisation and pseudonymisation of personal data 96
3.4.2.1 Anonymisation and pseudonymisation as a processing subject to data protection law 96
3.4.2.2 Anonymisation as a means to avoid the applicability of data protection law 97
3.4.2.3 Anonymisation and pseudonymisation as a means to avoid the applicability of specific data protection obligations 103
3.4.2.4 Anonymisation and pseudonymisation as a means to comply with data protection law 104
3.4.2.5 Examples of anonymisation and pseudonymisation of personal data in the transport sector 106
3.4.3 Techniques of anonymisation as a way to protect non-personal data 108
3.4.3.1 Trade secrets 108

3.4.3.2 Confidential information 109
3.4.3.3 Anonymisation of trade secrets and confidential information 109
3.4.3.4 Example of Anonymisation and pseudonymisation of non-personal data in the transport sector 111
3.4.4 Summary 111
3.5 Supply of digital content and services (personal data as counter performance) 114
3.5.1 Introduction to personal data as counter-performance 114
3.5.2 Economic dimension 114
3.5.2.1 Business model of free digital content 115
3.5.2.2 Quantifying personal data 115
3.5.3 Educational dimension 117
3.5.4 The desirability of legislating 118
3.5.4.1 Practical challenges 118
3.5.4.2 Legal challenges 119
3.5.4.3 Is there a need to monetize data? 120
3.5.5 Summary 121
3.6 Free flow of data 122
3.6.1 Types of restrictions to the free flow of data 122
3.6.1.1 Strict restrictions 122
3.6.1.2 Conditional restrictions 123
3.6.2 Rationale and impact of restrictions 123
3.6.3 Existing legal instruments to assess the validity of data localisation requirements 124
3.6.4 Proposed Regulation on a framework for the free flow of non-personal data 124
3.6.4.1 Main features of the proposed Regulation 125
3.6.4.2 Challenges encountered in the proposed Regulation 126
3.6.4.2.1 Scope of application: non-personal data 126
3.6.4.2.2 Mixed data sets 129
3.6.4.2.3 Availability of data to competent authorities 130
3.6.4.3 Opportunities of the proposed Regulation 130
3.6.5 Summary 133
3.7 Intellectual property in big data environment 135
3.7.1 Copyright 135

3.7.1.1 Legal Framework 135
3.7.1.1.1 International Legal Framework 135
3.7.1.1.2 European Union Legal Framework 136
3.7.1.1.3 The EU Copyright Reform 137
3.7.1.1.4 National Legal Framework 139
3.7.1.2 Copyright Protection: General Overview 139
3.7.1.2.1 Scope of Copyright Protection 139
3.7.1.2.2 Ownership of Rights 141
3.7.1.2.3 Works in the Public Domain 142
3.7.1.3 Copyright Protection of Data 142
3.7.1.4 Exclusive Rights and Copyright Exceptions 144
3.7.1.4.1 Exclusive Rights 144
3.7.1.4.2 Copyright Exceptions and Limitations 147
3.7.1.5 Contractual Aspects 149
3.7.1.5.1 Transfer of Copyright 149
3.7.1.5.2 Licence Agreements 150
3.7.1.5.3 Free and Open Licences 151
3.7.1.6 Conclusion on copyright in a big data environment 151
3.7.2 Database Rights 153
3.7.2.1 Legal Framework 153
3.7.2.1.1 International Legal Framework 153
3.7.2.1.2 European Union Legal Framework 154
3.7.2.2 General Principles of Database Protection in the EU 155
3.7.2.2.1 Definition of Database 155
3.7.2.2.2 Types of Protection 156
3.7.2.3 Copyright Protection for Databases 157
3.7.2.3.1 Protection Requirements 157
3.7.2.3.2 Ownership of Rights 158
3.7.2.3.3 Copyright on Database 159
3.7.2.3.4 Exceptions 160
3.7.2.4 Sui Generis Protection of Databases 160
3.7.2.4.1 Protection Requirements 161
3.7.2.4.2 Application to the Data Economy 162
3.7.2.4.3 Ownership of Rights 163
3.7.2.4.4 Sui Generis Rights to Database 164
3.7.2.4.5 Rights and Obligations of "Lawful Users" 165

3.7.2.4.6 Exceptions to the Sui Generis Protection 166
3.7.2.5 Possibility to Protect Data under Database Rights 167
3.7.2.6 Contractual Aspects 168
3.7.2.7 Conclusion on database rights in a big data environment 169
3.7.3 Trade Secrets and Confidentiality 171
3.7.3.1 Legal Framework 171
3.7.3.1.1 International Legal Framework 171
3.7.3.1.2 European Union Legal Framework 171
3.7.3.2 Possibility to Protect Data as Trade Secrets 172
3.7.3.2.1 Definition and Scope of Protection 172
3.7.3.2.2 Who Owns Trade Secrets? 172
3.7.3.2.3 Rights Conferred 173
3.7.3.2.4 Exceptions 174
3.7.3.2.5 Data Protected as Trade Secrets 174
3.7.3.3 Confidentiality: Contractual Aspects 175
3.7.3.4 Conclusion on trade secrets and confidentiality in a big data environment 176
3.7.4 Summary 177
3.8 Open data 178
3.8.1 Concept of open data and PSI 178
3.8.2 Non-legislative measures on open data 179
3.8.3 Legislative measures on open data: Directive on the re-use of public sector information 180
3.8.3.1 PSI Directive of 2003 180
3.8.3.2 PSI Directive of 2013 180
3.8.4 Challenges and opportunities of sharing public sector information 182
3.8.4.1 Opportunities of sharing public sector information 182
3.8.4.2 Challenges of sharing public sector information 186
3.8.5 Towards a new PSI Directive? 189
3.8.5.1 Proposal for a revision of the PSI Directive 189
3.8.5.2 Benefits and shortcomings of the Recast Proposal 191
3.8.5.3 Limits to the desirability of opening up PSI: the case of essential services and critical infrastructure 193
3.8.6 Summary 194
3.9 Data sharing obligations 197
3.9.1 Introduction 197

3.9.2 Data sharing obligations in the transport sector 198
3.9.2.1 Travel Information Services 199
3.9.2.2 eCall 200
3.9.2.3 Minimum universal traffic information 200
3.9.2.4 Information services for parking for trucks and commercial vehicles 201
3.9.2.5 EU-wide real-time traffic information services 201
3.9.2.6 Infrastructure for Spatial Information in the European Union 202
3.9.2.7 Advance Passenger Information 202
3.9.2.8 Rail Passengers' Rights 203
3.9.2.9 Vehicle Emissions 203
3.9.2.10 Car Labelling 204
3.9.2.11 Vessel Traffic Monitoring 204
3.9.3 Other data sharing obligations 205
3.9.3.1 Unfair Contract Terms and Unfair Commercial Practices 205
3.9.3.2 Platform-to-Business Transparency 206
3.9.3.3 Competition Law 207
3.9.3.4 Data sharing obligations imposed through public tendering 208
3.9.4 Summary 208
3.10 Data ownership 211
3.10.1 The "Ownership" Concept 211
3.10.2 Actors in the Data Value Chain who Could Claim Ownership in Data .. 212
3.10.2.1 Internet Service Providers 213
3.10.2.2 IT Infrastructure Providers 213
3.10.2.3 Data Providers 213
3.10.2.4 Data Analytics Service Providers 214
3.10.2.5 Data-driven Entrepreneurs 214
3.10.2.6 A Layered Approach of the Key Roles of Actors 214
3.10.3 Legislation on data ownership 215
3.10.4 Case law addressing the issues of "ownership" of data 215
3.10.5 Commission Communications having an impact on the Data Ownership Debate 217
3.10.5.1 "Towards a Thriving Data-Driven Economy" (2014) 217
3.10.5.2 "A Digital Single Market Strategy for Europe" (2015) 218
3.10.5.3 "Building a European Data Economy" (2017) 218
3.10.5.4 "Towards a Common European Data Space" (2018) 219

3.10.6 Legal doctrine related to data ownership *220*
3.10.7 Summary *223*
3.11 Data Sharing Agreements 224
3.11.1 Data sharing agreement definition *224*
3.11.2 General rules applicable to data sharing agreements *225*
3.11.3 Completion and execution of a data sharing agreement *226*
3.11.4 Terms and conditions set forth by the parties *226*
3.11.5 Guidance from the European Commission *227*
3.11.5.1 B2B data sharing agreements 228
3.11.5.2 B2G data sharing agreements 232
3.11.6 Data Sharing Agreements: a Critical Analysis *233*
3.11.7 Summary *234*
3.12 Liability 236
3.12.1 Setting the scene *236*
3.12.2 Extra-contractual and statutory lability and safety regimes *238*
3.12.3 Contractual liability *242*
3.12.4 Limitation of liability for intermediaries – Safe Harbour *243*
3.12.5 Liability aspects of the draft Directive on the supply of digital content *246*
3.12.5.1 Overview of the liability aspects of the Draft Directive on the supply of digital content 246
3.12.5.2 Analysis of the potential applicability and consequences on big data 247
3.12.6 Summary *248*
3.13 Competition 250
3.13.1 The role of big data in competition law analysis – the approach of different competition authorities in recent years *252*
3.13.2 Big Data & Abuse of Dominance *255*
3.13.2.1 Overview 255
3.13.2.2 Examples 256
3.13.3 Big Data & Mergers *258*
3.13.3.1 Overview 258
3.13.3.2 Examples 260
3.13.4 Big Data & Agreements between undertakings *262*
3.13.5 Big Data and Transport: competition law issues *263*

3.13.6 Summary 273
3.14 Conclusion 277
References 281

List of Figures

Figure 1: Taxonomy of types of data 12
Figure 2: Potential variety of the actors involved in a C-ITS context 18
Figure 3: Overview of data protection principles 19
Figure 4: Graphic overview of the re-purposing assessment 22
Figure 5: Overview of the data subjects' rights 46
Figure 6: Legal criteria triggering the application of Article 22 GDPR and restrictions and permitted acts 55
Figure 7: Risk-based approach 102
Figure 8: ICO Anonymisation Code of Practice – Case study 3 107
Figure 9: Three-pronged legal regime of the Recast Proposal for a revision of the PSI Directive. 190
Figure 10: Data value cycle 212
Figure 11: The data ecosystem as layers of (key roles of) actors 215
Figure 12: Technical mechanisms for data sharing according to the Commission 230
Figure 13: Overview of the components and layers of disruptive technologies 237
Figure 14: The interplay between data protection, competition and consumer protection (as well as [unfair] trade practice) 251
Figure 15: Overview of key areas of competition law relevant to big data 252

List of Tables

Table 1: Overview of the different relationships that may arise in a data protection context 17
Table 2: Concepts of "profiling" and "automated decision-making" 53
Table 3: Summary table of opportunities and challenges in relation to privacy and data protection in the context of big data in the transport sector 63
Table 4: Illustration of the complexity of the situation in certain Member States that have already transposed the NISD 69
Table 5: Transport modes targeted by the NISD 70
Table 6: Key challenges related to the secure use of gig data identified by ENISA 76
Table 7: Security issues and mitigation measures 77
Table 8: Key recommendations identified by ENISA 78
Table 9: Summary table of opportunities and challenges in relation to (cyber-)security in the context of big data in the transport sector 82
Table 10: Breach notification requirements under the GDPR 86
Table 11: WP29 examples of start 72-hour period for notification 87
Table 12: Factors to determine the significance of an incident 89
Table 13: Overview of EU guidelines related to NISD notification requirements 91
Table 14: Anonymisation techniques 94
Table 15: Pseudonymisation techniques 95
Table 16: Key legal notions of anonymisation, pseudonymisation and encryption 96
Table 17: Strengths and weaknesses of anonymisation techniques 99
Table 18: Advantages of the use of anonymisation to protect non-personal information 110
Table 19: Summary table of opportunities and challenges in relation to anonymisation / pseudonymisation in the context of big data in the transport sector 112
Table 20: Key measures to valuate personal data 116
Table 21: Summary table of opportunities and challenges in relation to the supply of digital content and services in the context of big data in the transport sector 121
Table 22: Summary table of opportunities and challenges in relation to free flow of data in the context of big data in the transport sector 134
Table 23: Overview of the author's exclusive rights 144
Table 24: Main findings of the second evaluation of the Database Directive 155
Table 25: Database protection thresholds in selected EU Member States 162
Table 26: The concept of "lawful users" in Belgium, France and Germany 166
Table 27: Overview of the differences of implementation of the scientific research exception 167
Table 28: Reasonable measures to secure confidentiality of information 175
Table 29: Summary table of opportunities and challenges in relation to intellectual property in the context of big data in the transport sector 177
Table 30: Most important aspects of open data 178
Table 31: Graphic overview of Creative Commons licence CC0 181
Table 32: Summary table of opportunities and challenges in relation to open data in the context of big data in the transport sector 196
Table 33: Summary table of opportunities and challenges in relation to data sharing obligations in the context of big data in the transport sector 210
Table 34: Moving forward on access to machine-generated data 219
Table 35: Summary table of opportunities and challenges in relation to data ownership in the context of big data in the transport sector 223
Table 36: Principles for B2B data sharing 228
Table 37: Principles for B2G data sharing 233
Table 38: Summary table of opportunities and challenges in relation to data sharing agreements in the context of big data in the transport sector 235
Table 39: Schematic overview of the EU legal framework on liability 239
Table 40: Expected interventions regarding the EU liability legal framework 242

Table 41: Summary table of opportunities and challenges in relation to liability in the context of big data in the transport sector 249
Table 42: Summary table of opportunities and challenges in relation to competition in the context of big data in the transport sector 276

Glossary

Abbreviation	Expression
ACM	Autoriteit Consument & Markt (Dutch competition authority)
Agcom	Autorità per le Garanzie nelle Comunicazioni (regulator and competition authority for the communication industries in Italy)
API	Application Programming Interface
B2B	Business-to-Business
B2G	Business-to-Government
BCRs	Binding Corporate Rules
BMVI	Bundesministerium für Verkehr und digitale Infrastruktur (German Federal Ministry of Transport and Digital Infrastructure)
CCAM	Cooperative, Connected and Automated Mobility
CCTV	Closed-Circuit Television
CDN	Content Delivery Network
C-ITS	Cooperative Intelligent Transport Systems
CJEU	Court of Justice of the European Union
CMA	UK's Competition & Market Authority
CSIRTs	Computer Security Incident Response Teams
Database Directive	Directive on the legal protection of databases
DG COMP	European Commission's Directorate-General for Competition
DPA	Data protection authority
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officier

DSA	Data Sharing Agreement
DSM	Digital Single Market
DSM Directive	Directive on copyright in the Digital Single Market
DSP	Digital Service Provider
EDPS	European Data Projection Supervisor
EEA	European Economic Area
ENISA	European Agency for Network and Information Security
FCA	French competition authority
FOT	Field Operational Test
GDPR	General Data Protection Regulation
GIS	Geographical Information Systems
ICO	Information Commissioner's Office (UK DPA)
IEC	International Electrotechnical Commission
InfoSoc Directive	Information Society Directive
IoT	Internet of Things
IP	Intellectual property
IPR	Intellectual Property Rights
ISO	International Standards Organisation
ISPs	Internet Service Providers
ITS	Intelligent Transport Systems
LeMO	Leveraging Big Data for Managing Transport Operations
LIA	Legitimate Interests Assessment

LINC	Laboratoire d'Innovation Numérique de la CNIL
MaaS	Mobility as a Service
McKinsey	McKinsey Global Institute
NCAs	National Competent Authorities (in the meaning of the NIS Directive)
NDA	Non-Disclosure Agreement
NIS Directive (NISD)	Network and Information Security Directive
OECD	Organization for Economic Co-operation and Development
OEM	Original Equipment Manufacturers
OES	Operators of Essential Services
PET	Privacy-enhancing technologies
PIA	Privacy Impact Assessment
PSDs	Public Sector Databases
PSI	Public Sector Information
PSI Directive	Public Sector Information Directive
RAM	Random-Access Memory
RIM	Repair and Maintenance Information
SCRA	Small Cells Risk Analysis
SMEs	Small and Medium-sized Enterprises
SSCs	Standard Contractual Clauses
TFEU	Treaty on the Functioning of the European Union
TRIPS	Trade-Related Aspects of Intellectual Property Rights
WIPO	World Intellectual Property Organisation

WP29	Article 29 Working Party

1 Introduction

1.1 Abstract

This Deliverable identifies the various legal issues that are relevant to the production of, access to, linking of and re-use of big data in the transport sector.

The legal issues that have been identified as most relevant in light of the performed research, and which are further examined in this Deliverable, are as follows:

- Privacy and data protection
- (Cyber-)Security
- Breach-related obligations
- Anonymisation / pseudonymisation
- Supply of digital content and services (personal data as counter performance)
- Free flow of data
- Intellectual property in big data environment
- Open data
- Data sharing obligations
- Data ownership
- Data sharing agreements
- Liability
- Competition

In a first chapter (Chapter 2), this Deliverable sets the scene by reiterating the concept of big data with its particular characteristics and highlighting its interaction with legal issues.

Chapter 3 examines the identified legal issues and discusses the challenges and opportunities related thereto. Whenever possible, the various Sections of Chapter 3 provide illustrations in the transport sector.

Finally, the conclusion of this Deliverable (Chapter 4) introduces possible ways of moving forward, mainly at EU level.

1.2 Purpose of the document

The LeMO project will contribute to developing a strategy that defines the research efforts necessary for the realisation of the big data economy through a consideration of the opportunities, limitations and challenges associated with big data in the transport sector. It will thus aid European stakeholders in improving adoption of technology and support actions that amplify constructive opportunities (e.g. new products and services, efficiencies, economic competitiveness, etc.) associated with big data, while diminishing limitations (e.g. privacy infringements, legal barriers, etc.). As such, the LeMO project has three main objectives:

1. To produce a **research and policy roadmap towards data openness, collection, exploitation and data sharing** to support European transport stakeholders in capturing

and addressing issues, which range from technical to institutional, including legitimacy, data privacy and security.

2. To **involve European transport sector actors** in order to identify and analyse concrete opportunities, barriers and limitations of the transportation systems to exploit big data opportunities.
3. To **disseminate the LeMO findings, recommendations and the contribution of the LeMO** project to evidence–based decision making by improving knowledge on methodological and exploitation issues taking also into account economic, legal, social, institutional and technical aspects.

Task 2.2 aims to identify and examine the different legal issues that are currently relevant to big data, and which may be relevant to big data as opportunities for the production of, access to, linking of and re-use of big data in the transport sector. As part of this task, partners of the LeMO Project examined EU legislative instruments, national regulations where needed, EU case law, national case law where needed, legal doctrine, academic journal articles, project reports, and any other relevant information to identify potential legal issues relevant to big data. On the basis thereof, the partners produced a high-level discussion of legal issues, including both constructive information and recommendations, which will serve as a baseline for further development in relation to other Work Packages of the Project. Particular legal issues to be examined include but are not limited to privacy and data protection, intellectual property (including copyright), open data and liability. Especially, issues of legitimacy and public acceptance (e.g. privacy, data security, etc.) were adequately addressed.

More particularly, this Deliverable provides an overview of the context (big data, with a particular focus on the transport sector) and identified relevant legal issues. Whenever possible, the various Sections provide illustrations with practical considerations and introduce possible ways to move forward and legal initiatives that may exist, focusing mainly on the EU.

1.3 Target audience

This document will be made publicly available. The results of the research are especially interesting for people working for organisations in the public sector (e.g. politicians, policy makers, policy consultants etc.) and in the private sector (e.g. managers, directors, and consultants).

More specifically, the target audience for this deliverable includes:

- European Commission
- EU Parliament
- Public and private transport organisations
- Authorities (regional and national level) that develop and enforce policies and legislation
- Horizon 2020 projects and related transport projects (cf. clustering activities)
- Organisations and experts involved in the LeMO case studies
- Partners and Advisory & Reference Group in the LeMO project

1.4 Methodology

The following Deliverable approaches the legal issues relevant to big data in the transport sector primarily using the doctrinal legal research method (see below).[1] Although this deliverable applies several known legal methodologies, they are not followed rigidly. Where applicable, aspects of other methodologies providing a useful lens with which to evaluate the legal problems considered are also incorporated.

1.4.1 Doctrinal method

The doctrinal method is closely tied to the common law approach of evaluating legal precedents and their application in future circumstances. The method has been described as being at the "core of practice" and is applied by attorneys in the field as well as academics.[2] The doctrinal method applies quantitative as well as qualitative research methods.[3] Taking a quantitative approach, the method takes a positive view of the world evaluating law in a manner that is "objective, neutral, and fixed."[4]

In this Deliverable, B&B evaluates the application of existing laws to a disruptive technology – i.e. big data – that is (currently) not specifically regulated. In researching these issues, the doctrinal method plays several roles. Primarily, the doctrinal method encompasses research in law, as opposed to research about law. In this Deliverable, the state of the law in the EU is evaluated by considering how legal rules will likely apply to big data. Although the technology may be new (or a new spin on existing technology) there are existing laws in place that are applicable to big data, even if they are not specific. As a starting point, the Deliverable considers whenever relevant what the existing law is, and how it applies to big data. In particular, the Deliverable evaluates current and future problems with applying current regulations. In some circumstances, B&B will examine, in some cases, what the law ought to be (*de lege ferenda*).

1.4.2 Fundamental and non-fundamental research constraints

The broadly defined research questions structured above are not without some constraints.[5] Certain Member States may have different views or have a stricter approach regarding certain legal issues. Accordingly, illustrations of national laws specifically applicable are given in some relevant cases. It is also not excluded that B&B relies on its network of legal experts across the EU to consider the rules applicable to certain topics. However, even when relying on other countries, the legal studies provided in the LeMO project will not map big data regulations in the entire EU/EEA.

[1] *See e.g.* Terry Hutchinson and Nigel Duncan, 'Defining and Describing What We Do: Doctrinal Legal Research' [2012] 17/1 Deakin Law Review 83
[2] Ibid
[3] Ibid
[4] Ibid
[5] Dan Svantesson, 'A Legal Method for Solving Issues of Internet Regulation' [2011] 19(3) International Journal of Law and Information Technology 243

1.4.3 Sources of law

As indicated in the description of work of the LeMO project, and in particular in relation to Working Package 2 (Institutional and governmental issues and barriers), the legal analysis will be provided with an EU perspective, taking into consideration the case-law from the Court of Justice of European Union. It will provide, where necessary and applicable only, national law examples and case-law generally highlighted by the legal literature. A similar approach will be adopted in relation to the opinions or guidance provided by authorities.

2 Setting the scene

2.1 The concept of "big data"

Although this Deliverable 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 will consider the legal issues having big data in mind.

More concretely, what is to be understood by "big data"? While there is no real consensus on a definition, the initial logical observation is that it is often described as a large dataset comprising different types of data that have grown beyond the ability to manage and analyse them with traditional tools.[6] Hence, handling variable (un-)structured data in real-time requires the adoption and use of new methods and tools (e.g., processors, software, algorithms, etc.).[7]

One could however not 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 stored in different locations, connected by networks and brought together by software.[8]
- **Velocity**: refers to the speed, which is of 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.[9]
- **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.[10] 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*."[11] This typically poses ethical issues but also legal ones, which are examined in this Deliverable.

---

[6] Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3
[7] Commission, 'Towards a thriving data-driven economy' (Communication) COM (2014) 442 final, 4
[8] 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 16 October 2018
[9] James R. Kalyvas and Michael R. Overly, *Big Data: A Business and Legal Guide* (Auerbach Publications 2014) 5
[10] Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3
[11] 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 16 October 2018

The "V" of "**Value**" has also been highlighted to refer to the ability of turning data into value.[12] While it could be argued that data, per se, has no value, processing it creates value. In other words, all data collected and stored would not likely generate any value unless it is analysed or used by some "intelligent" software algorithms, which analyse data, learn from data, and make/suggest decisions or predictions. Moreover, the value related to data may also flow from the time spent by humans to create the algorithms, to train such algorithms with human-generated examples and answers, or to organise the data. Similarly, the (personal) data provided by individuals in their day-to-day life (by using social media platforms or using an itinerary application for instance), also has a value. Deliverable D2.1 of the LeMO Project has notably look into the "value" aspect of data in the context of a broader examination of the different economic issues that are relevant to big data currently, and which may be relevant to big data as opportunities for the production of, access to, linking of and re-use of big data in the transport sector.

Finally, when looking into the legal issues related to big data, it is worth considering other disruptive technologies such as Artificial Intelligence ("**AI**"), including Machine Learning, Deep Learning, or Neutral Networks, which are all algorithm-based. Such algorithmic methods rely on a vast amount of data (big data) to produce desired results and to find trends, patterns and predictions. In some instances, this Deliverable will provide illustrations relying on such other technologies.

2.2 Big data in the transport sector

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

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

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. 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 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] 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 16 October 2018

2.3 Policy framework

The legislator, at EU and/or national level, has adopted policies in order to regulate several aspects related to (big) data, the transport sector, but also to combat the most conspicuous and persistent ethical issues or to set social norms.[13]

While there are no policies specific to big data, lawmakers have adopted some legislations aimed notably at protecting the privacy of its 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.[14]

While this Deliverable aims to examine the most relevant general legal issues related to "data", it will nonetheless rely, where relevant, on specific transport legislations.

In addition to public policies, companies – including in the transport sector – have adopted or adhered 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. The potential applications in the transport sector are diverse, as digitalisation is a major trend of the transport sector.[15]

Despite the existence of public and private policies, the use of new technologies, such as in this case big data-driven technologies, raises new issues that may require adopting new or replacing existing policies. This Deliverable will attempt to propose, where possible, ways to move forward and determine whether regulatory intervention and/or soft law measures are desirable.

2.4 Assigning responsibilities

The data value cycle can be rather complex and involves numerous stakeholders. Many of such 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, analyse, purchase, take a licence on, or add value to the data.

This complexity increases the difficulties in determining who could be legally responsible and liable for any wrongdoing and damage. Are 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 other actors responsible?

[13] Deliverable D2.3 of the LeMO project identifies and examines various societal and ethical issues that are relevant to the production of, access to, linking of and re-use of big data in the transport sector.
[14] 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.
[15] 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.

2.5 Identifying legal issues related to big data in the transport sector

This Deliverable addresses how the use of (big) data and the deployment of new data-driven technologies may raise discussions in relation to the legal intricacies, putting a particular emphasis on big data in the transport sector. The issues presented here may nonetheless be valid for other domains.

More specifically, the research conducted in the context of the LeMO Project and this Deliverable has enabled identifying the following key legal issues, deemed to be particularly relevant to big data, including in the transport sector:

- Privacy and data protection
- (Cyber-)Security
- Breach-related obligations
- Anonymisation / pseudonymisation
- Supply of digital content and services (personal data as counter performance)
- Free flow of data
- Intellectual property in big data environment
- Open data
- Data sharing obligations
- Data ownership
- Data sharing agreements
- Liability
- Competition

This Deliverable focuses on the above issues, which are detailed in the following Chapter.

This, however, does not mean that other legal issues are not relevant. Indeed, the development of new services in the transport sector that rely on data-driven technologies raise a myriad of technical, economic, legal, ethical and social issues.

3 Analysis of key legal issues

3.1 Privacy and data protection

The analysis of privacy and security 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 Section. 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."*[16] This multitude of sources, actors and purposes cannot always be reconciled with the legal requirements related to data protection and security[17]. 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 among which the new General Data Protection Regulation (the "GDPR") that came into effect on 25 May 2018, introducing a raft of changes to the current 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.

The newly introduced instruments are especially relevant in light of – and may be perceived to be contradictory to – the EU Commission strategy on "Building a European Data Economy", which strives towards a true EU data-driven economy. In this respect, the Civil Liberties, Justice and Home Affairs Committee of the European Parliament has pointed out in its Study on "Big Data and Smart Devices and Their Impact on Privacy" that *"two interlinked, and to some extent conflicting, initiatives are relevant here: the development of EU strategies promoting a data-driven economy and the current reform of the EU personal data protection legal framework, in the context of the adoption of a General Data Protection Regulation."*[18]

This Section will not delve into all rights and obligations included in the GDPR. The following sub-Sections will however examine some of the core principles and concepts put forward by the GDPR, as well as the different interpretations of such concepts made by the many actors active in the fields of privacy and data protection at European level, will be confronted with the context of big data analytics, with a particular focus on the transportation sector.

[16] 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> accessed 16October 2018

[17] "*Les principes de nombreux projets de big data sont en contradiction avec les principes de respect de la vie privée et de protection des données*" Commission de la protection de la vie privée, 'Rapport Big Data' (CPVP 2017) 14 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 18 October 2018

[18] 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) 5 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018

3.1.1 The concept of "personal data"

3.1.1.1 *Types of data*

In the EU, the concept of "personal data" is rather wide-ranging. According to the GDPR, the concept refers to any information relating to an identified or identifiable natural person ('data subject'): *"An identifiable natural person is 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"*. [19] The GDPR particularly expanded this concept to take account of the online environment.

The fact that the definition refers to any information relating to an "identified or identifiable" individual[20] essentially means that it includes the name of a person, mobile phone number, e-mail, location, contacts, credit card and payment data, browsing history, pictures, videos, temperature, blood pressure, insulin level, etc.

3.1.1.2 Sensitive Personal Data

The GDPR distinguishes ordinary and special categories of personal data, also known as "sensitive data". The processing of such types of data is restricted and prohibited in most cases. Accordingly, in order to process such special categories of data, the data controller must find a proper legal ground exhaustively listed in the GDPR.

More specifically, the GDPR includes the following concepts, which are defined: "special categories of personal data", "data concerning health", "genetic data"[21], "biometric data", "data relating to criminal convictions and offences".

In the context of big data, it cannot be excluded that the data analysis concerns sensitive data or even to have a "transformational impact" on data. For instance, the processing of non-sensitive personal data could lead – for instance, through data mining – to the generation of data that reveals sensitive information.[22]

---

[19] GDPR, art 4(1)

[20] In addition, it shall be noted that additional requirements apply to "special categories" of personal data, meaning data related to racial or ethnic origins, political opinions, religious or philosophical beliefs, trade union membership and data concerning health or sex life. Special restrictions may be set out under EU national laws for the processing of data relating to offences, criminal convictions, as well as, and under certain national laws, data relating to administrative sanctions or judgments in civil cases.

[21] Genetic data was not expressly included as a sensitive data within the Directive. This was considered as a major backdrop within the Directive for failing to extend such protection to new but important data as genetic information.

[22] 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) 30 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018; Commission de la protection de la vie privée, 'Rapport Big Data' (CPVP 2017) 47 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 16 October 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) 7 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018; European Data Protection Supervisor, 'Opinion 4/2015. Towards a New Digital Ethics. Data, Dignity and Technology' (EDPS 2015) 6 <https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf> accessed 16 October 2018

Privacy and data protection in the transport sector – Example 1

The Article 29 Working Party ("**WP29**") observed in its opinion 3/2017 on Cooperative Intelligent Transport Systems ("**C-ITS**")[23] 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 as over speeding data, signal violation / intersection safety).

Article 10 of the GDPR specifies that such data relating to criminal convictions and offences may only be processed under the control of official authority or when the processing is authorized by Union or Member State law providing for appropriate safeguards for the rights and freedoms of data subjects. WP29 concluded 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*".

Based on this strict opinion of WP29, it would thus not be possible to collect data such as car speed, CCTV footages, etc., as potentially, this could constitute criminal data.

3.1.1.3 Taxonomy of types of data

The development of new technologies, such as big data analytics and the Internet of Things, has somewhat complicated the notion of "personal data" and led to the emergence of various types of data.[24] The Data Protection Authority of the United Kingdom (the "Information Commissioner's Office" or "ICO") has provided an overview of the taxonomy that shows the complexity of data and which can be depicted as follows:[25]

[23] 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

[24] See in for instance 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) 19 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 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) 10 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018

[25] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 1 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

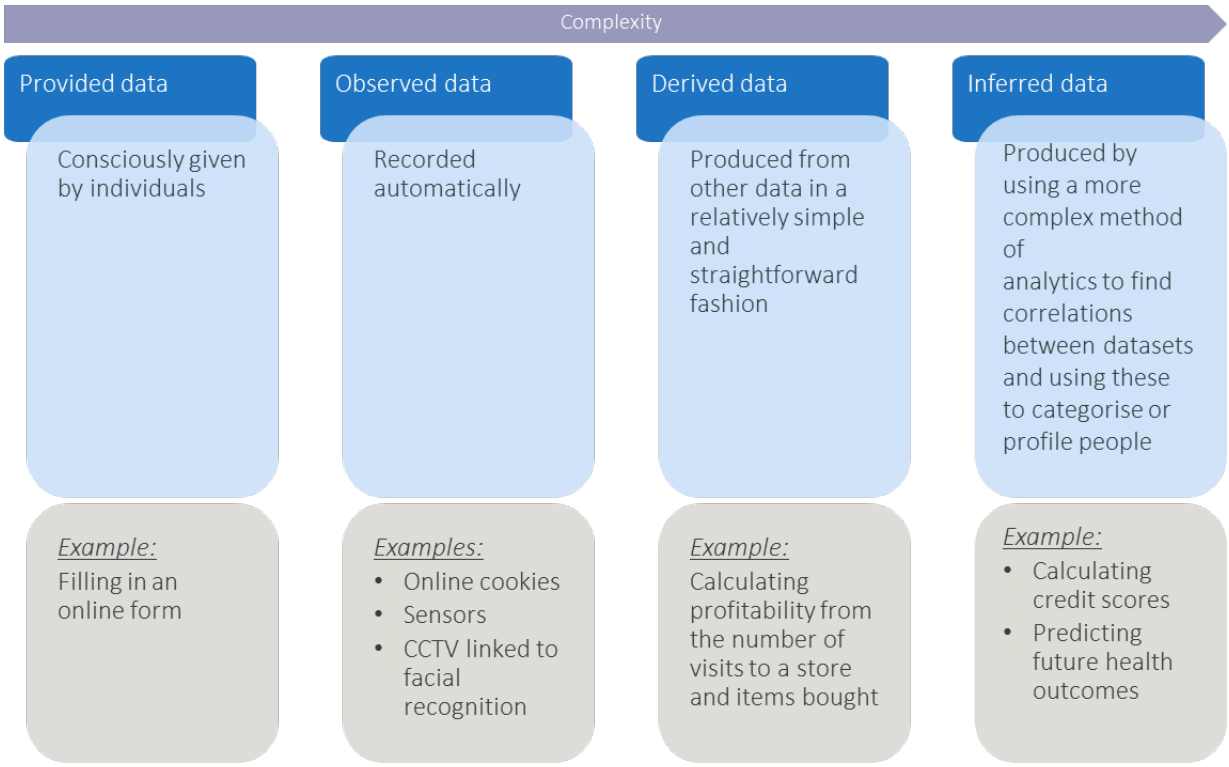


Figure 1: Taxonomy of types of data[26]

Despite the complexity of data, the GDPR does not make any distinction and the general concept of "personal data" will apply, even to inferred data.

3.1.1.4 The "Processing" of Personal Data

The GDPR applies when there is a "processing" of personal data which it defines as "*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*".[27]

It goes without saying that big data analytics involving "personal data" necessarily implies a processing within the meaning of the GDPR. Hence, the various principles and obligations set therein will need to be carefully assessed and complied with by the involved stakeholders.

[26] Ibid 14
[27] GDPR, art 4(2)

3.1.1.5 The "processing" of "personal data" in the transport sector

In light of the above definitions, the use of big data applications in the transport sector will in many circumstances require complying with the strict obligations of the GDPR.

Privacy and data protection in the transport sector – Example 2

C-ITS can be defined as systems that allow vehicles to be connected with each other and with the infrastructure and to the cloud. The main purpose is to render traffic more efficient and safer by shortening travel durations, helping drivers in difficult weather conditions, preventing road accidents such as with blind spot detection technologies. This technology implies the interaction of many different actors collecting a huge amount of different types of data (sensor based, Closed-Circuit Television ("**CCTV**") cameras, etc.) among which personal data.

The car can incorporate intelligent sensors providing the following info[28]: inductive loops (presence, count and speed), pneumatic tubes (counts, speed), cameras (counts, classification, speed, presence), infrared sensors (counts, speed, classification), passive acoustic and microwave (counts, speed, presence), etc.

Mobile-sourced data can also provide the following information: GPS/mobile (speed, presence, count and location), Bluetooth (speed, presence, count), etc.

The above information is all likely to be considered as personal data.

Privacy and data protection in the transport sector – Example 3

In the context of the Transforming Transport EU project, the "Smart Passenger Flow Pilot" focusses "on the full passenger process, analysing and describing the passenger behaviours in order to anticipate the number of resources required to manage the expected volume of passengers and predicting when passenger process might affect the aircraft departure times."[29]

Such types of analysis would usually require identifying passengers and examining their gender, location, boarding pass scanning time, etc. Except if such data is anonymised (see Section 3.4 above on the issues of anonymisation), it will be necessary to comply with the GDPR.

It follows from the foregoing that the broad definition of "personal data", the existence of the notion of "special categories of data" and the extensive concept of "processing" will necessarily require applying the numerous obligations under the GDPR when performing big

[28] Matthew Clarke, 'Big Data in Transport' (*The Institution of Engineering and Technology*, 2016) 4 <https://www.theiet.org/sectors/transport/topics/intelligent-mobility/articles/big-data.cfm?origin=carousel> accessed 16 October 2018

[29] Juan Antonio Ubeda and others, 'Transforming Transport. Summary of deliverable' (Transforming Transport 2018) 3 <https://transformingtransport.eu/sites/default/files/2018-08/D8.3_PUBLIC.pdf> accessed 16 October 2018

data analysis, including in the context of the transport sector. The interpretation to be given can be extensive to such extent that it leads to the conclusion that "data relating to criminal convictions and offences" is being processed. This would thus limit the processing activities or unnecessarily increase the processing restrictions by applying stringent rules included in the GDPR.

3.1.2 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.

As already specified by the Article 29 Working Party in 2010, 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.[30]

3.1.2.1 Data controller

Based on Article 4(7) of the GDPR, the following requirements are essential for an actor to be considered as a "controller":

- First, a controller can be a natural or legal person, public authority, agency or any other body. This implies that the form or nature of the entity is irrelevant.
- Second, the controller determines the purposes, conditions and means of the processing. This is a crucial element and one of the main factors in assessing this aspect is the level of influence that someone has in determining "why" (i.e., purposes) and "how" (i.e., means) certain processing activities should be performed. In establishing controllership, it has to be noted that the factual circumstances are a more relevant factor than a 'fine tune' designation based on contract or law.[31] This means, for example, that a clear contractual provision excluding a party from being a controller is not relevant if all the other circumstances indicate otherwise.

The decision regarding the 'purpose and means' can be made jointly with others, where several legally separate entities process data together or jointly with others for a shared purpose.

The GDPR has clarified the rules with respect to co-controllership by laying down under Article 26 the rules related to the responsibility of joint controllers. Particularly, it contains an elaborate provision on the obligations when more than one controller determines the purpose and means of processing. First, the controllers are required to determine – in an arrangement – the respective responsibilities in light of compliance with the GDPR, and

[30] Article 29 Data Protection Working Party, 'Opinion 1/2010 on the concepts of "controller" and "processor"' (2010) WP 169, 2
[31] Ibid 15

particularly the rights of data subjects. Second, the arrangement should reflect the effective roles and relationships of each controller, and the essence of the arrangement must be made available to data subjects. Third, the data subjects are granted the right to exercise their rights under the GDPR against each controller, irrespective of the arrangement among the controllers.

Finally, the situation of joint-controllership should not be confused with the situation of "controllers in common", which is not as such regulated under the GDPR. While joint controllers decide together on the purposes and the means of processing, controllers in common process a same data set independently of each other.

3.1.2.2 Data processor

To be qualified as a "processor"[32], a natural person or an entity must fulfil the following two elements:

- First, it must be a person or legal entity legally separate from the controller.
- Second, it must *process personal data on behalf of the controller*. This implies that decisions on the 'purpose' and 'essential means' should be made by the controller.[33]

  The concept of 'essential means' gives a margin of manoeuvre to processors (such as cloud providers) to determine technical and organisational issues without being considered 'controllers'.

The GDPR has reinforced the responsibilities imposed on processors.[34] This development has significant implications for service providers in the information technology sector, which are often considered as data processors so far as the provider adheres to the instructions of the controller and does not process the data for its own purposes. This entails that processors, such as cloud providers, become directly accountable vis-à-vis regulators as well as data subjects.

More specifically, the GDPR has extended the scope of application of EU data protection law, where certain requirements apply for the first time to "processors". The GDPR also introduces new rules that apply when engaging processors and when the latter engage sub-processors.

More particularly, data controllers may only appoint data processors that provide sufficient guarantees to implement appropriate technical and organisational measures to ensure

[32] GDPR, art 4(8)
[33] Ibid 13
[34] When considering big data, the Civil Liberties, Justice and Home Affairs Committee of the European Parliament highlighted that "*the 2014 European Commission Communication* [on data-driven economy] *makes clear that the fundamental right to personal data protection applies to Big Data when the data processed can be qualified as personal. Referring to the Commission's Data Protection Reform package (…), the Commission underlines that it will work with member states and stakeholders to ensure that business, and in particular SMEs, receive adequate guidance, notably on issues such as data anonymisation and pseudonymisation, data minimisation, personal data risk analysis, as well as tools and initiatives enhancing consumer awareness. The European Commission also announces its support to projects aiming to regulate personal data breaches and to ensure that data is used in a manner compatible with its initial collection, recognising that 'these measures will build the trust that is necessary to exploit the full potential of the data-driven economy*." 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) 18 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018; COM (2014) 442 final, 11

processing meets the requirements of the GDPR.[35] Also, processors are required to process personal data in accordance with the controller's instructions.

Finally, the relationship between the controller and the processor must also be governed by a binding contract the content of which must meet a minimum content. The obligations placed on the processor must cover the duration, nature and purpose of the processing, the types of data processed and the obligations and rights of the controller. There are also a number of specific requirements, including the processing of personal data only on documented instructions from the controller, and requirements to assist the controller in complying with many of its obligations. The data processor has an obligation to inform the controller if it believes an instruction breaches the GDPR or any other EU or Member State law.

3.1.2.3 Sub-processors

In many instances, including in the context of big data analysis, multiple service providers are included in a chain of processors.

The GDPR gives data controllers a wide degree of control in terms of the ability of the processor to sub-contract (engage "sub-processors"). In effect, data processors require prior written consent. This can be general, but even where general consent has been given, the processor is still required to inform the controller of any new sub-processors, giving the controller the opportunity to object. Also, the lead processor is required to reflect the same contractual obligations it has vis-à-vis the controller in a contract with any sub-processors and remains liable to the controller for the actions or inactions of any sub-processor.[36]

[35] GDPR, art 28(1)
[36] GDPR, arts 28(2) and 28(4)

3.1.2.4 Allocation of responsibilities in the context of big data and transport

On the basis of the previous sub-Sections, the table below provides an overview of the different relationships that may arise in a data protection context:

Controller-Processor relationship	Processor-Subprocessor relationship	Joint-Controller relationship	Controllers 'in common' relationship
			
Processor processes data on behalf of controller; controller solely determines the purposes and the means of the processing	Subprocessor is engaged by the processor to carry out specific processing activities on behalf of the controller	Two controllers are acting together to determine the purposes and the means of the processing: 1 pool of data and 1 set of same purposes	Two controllers share 1 pool of data; each controller defines the purposes and means of their own processing: 1 pool of data but multiple purposes
Article 28 GDPR	Article 28(2) and (4) GDPR	Article 26 GDPR	Undefined / Unregulated
A contract must be put in place (the minimum content is imposed by Art. 28 GDPR)	A contract must be put in place (the same obligations as set out in the contract between the controller and the processor shall be imposed on the subprocessor)	An arrangement must be put in place, the essence of which must be made available to data subjects. The arrangement must define the respective obligations and determine who's responsible towards the data subjects when they exercise their rights	Each controller must individually comply with the obligations under the GDPR

Table 1: *Overview of the different relationships that may arise in a data protection context*

The distinction between "controller" and "processor", taking into account the concepts of joint-controllership, controllers in common and sub-processors, can quickly become complex

in a big data context. This is mainly due to the fact that many actors may be involved in the data value chain.[37]

The intricacies are also due to the objective of big data analytics which "*is about finding correlations, making predictions and aiding decision-making; all of which blur the lines between who is actually determining the purposes and manner of the processing when an organisation has chosen to outsource the analytics to another company*".[38]

Privacy and data protection in the transport sector – Example 4

As already demonstrated in our previous example, data collection and processing in a C-ITS context involve the interaction of a multitude of actors (see diagram). Determining which actor is controller (joint or in common) and for which processing activity, and which other qualifies as a (sub-)processor and put in place the right contracts can be quite complex.

Figure 2: Potential variety of the actors involved in a C-ITS context

It follows that guidance in determining the precise roles of those involved in complex data processing activities and template agreements, compliant with the strict requirements of the GDPR, should be further developed by the competent authorities at EU and national levels. This would increase legal certainty to those involved in the data value chain, and ultimately benefit data subjects.

[37] The underlying strategy of such value chain generally aims at extracting the maximum value from data in order to provide benefits for the economy and citizens (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) 17 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018). The Civil Liberties, Justice and Home Affairs Committee of the European Parliament criticises the European Commission, stating that "the successive European Commission Communications fail to clarify what is personal data and how the personal data life cycle may contain conflicting purposes and priorities". 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) 18 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018

[38] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 57 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

3.1.3 Data protection principles

The data protection principles are at the core of the rules related to the processing of personal data. Article 5(1) of the GDPR lists six key principles relating to the processing of personal data and Article 5(2) provides for a general principle of "accountability", according to which the controller shall be responsible for, and able to demonstrate compliance with, the other six principles. The data protection principles may be depicted in Figure 3 below.

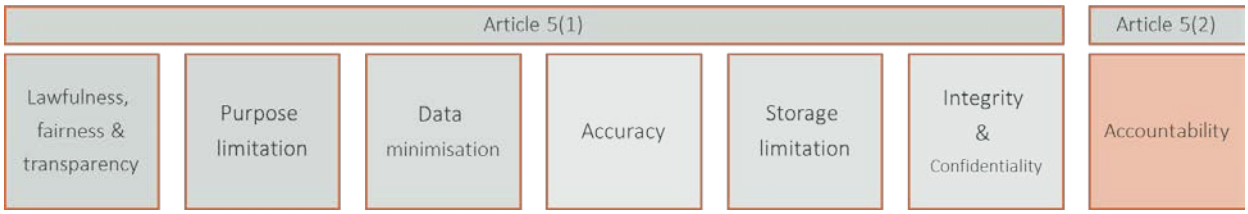


Figure 3: Overview of data protection principles

The following sub-Sections examine these principles more in depth, as well as the challenges and opportunities they may pose in relation to big data.

3.1.3.1 Lawfulness, fairness & transparency

Personal data must be processed lawfully, fairly, and in a transparent manner in relation to the data subject[39]. The "lawfulness" of the processing will be examined in sub-Section 3.1.4 above.

One of the requirements is for the processing of data to be "fair", meaning that the data subject must be in a position to learn of the existence of a processing operation and must be given accurate and full information (for instance about the identity of the controller, the purposes of the processing of data, etc.). Fairness is therefore about being open on the processing in order to empower individuals by making them aware of what information about them is being collected and processed.

More particularly, in a big data context, one must examine the effects of the processing on data subjects and the individuals' reasonable expectations with respect to the processing of their personal data.[40] Indeed, certain big data analytics projects are more likely to directly affect individuals than others. For instance, in case there is a profiling of users, the effects on individuals will be higher.

The principle of "fair and transparent" processing means that the controller must provide information to individuals about its processing of their data, unless the individual already has this information.

The information to be provided is specified under Articles 13 and 14 of the GDPR. The controller may also have to provide additional information if, in the specific circumstances and context, this is necessary for the processing to be fair and transparent. Also, the GDPR affirms

[39] GDPR, art 5(1)(a)

[40] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 20 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

that the information must be provided in a concise, transparent, intelligible and easily accessible way, using clear and plain language (in particular where the data subject is a child).

The transparency principle in a big data context – where the complexity of the analytics renders the processing opaque – 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*."[41]

Privacy and data protection in the transport sector – Example 5

In its recent guidelines on automated individual decision-making and profiling adopted on 3 October 2017, the Article29 Working Party takes the example of car insurances to illustrate the possible issues of fair, lawful and transparent processing of personal data in the transportation sector.[42]

It indicates that some insurers offer insurance rates and services based on an individual's driving behaviour. The elements they take into account in such cases possibly include the following elements: distance travelled, time spent driving, journey undertaken as well as predictions based on other data collected by the sensors in a (smart) car. The data collected would then be used for profiling to identify bad driving behaviour (such as fast acceleration, sudden braking, and speeding). This information can be cross-referenced with other sources (e.g., the weather, traffic, type of road) to better understand the driver's behaviour.

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.

3.1.3.2 Purpose limitation (and secondary use)

Personal data must be collected for specified, explicit and legitimate purposes; and must not be further processed in a way incompatible with those purposes.

Foremost, this requires any processing of personal 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."[43]

---

[41] 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 16 October 2018; 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 Degrave, Cécile de Terwangne, Séverine Dusollier and Robert Queck (eds), *Law, Norms and Freedoms in Cyberspace / Droit, Normes et Libertés dans le Cybermonde* (Larcier 2018)

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

[43] 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)

Furthermore, the principle includes a second building block: i.e. the prohibition to further process personal data in a way incompatible with the initial purposes (re-purposing). The Article 29 Working Party published a lengthy opinion in 2013 – under the Data Protection Directive – on the purpose limitation principle,[44] focuses on this second building block. It follows that distinguishing between compatible and incompatible processing of personal data is often a complex and delicate exercise.[45] The Article 29 Working Party affirms that compatibility must be assessed on a case-by-case basis, looking at the relevant circumstances and taking into account certain factors.[46] More specifically, transparency towards individuals must be preserved in case of further processing, encompassing not only the aim of the processing but also the manner in which it takes place.[47]

Article 6(4) of the GDPR has codified some elements of Opinion 03/2013. It sets out the rules on factors a controller must take into account to assess whether a new processing purpose is compatible with the purpose for which the data were initially collected. Where such processing is not based on consent[48], the GDPR lists five factors that should be taken into account in order to determine compatibility[49]:

- any link between the purposes for which the personal data have been collected and the purposes of the intended further processing
- the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller
- the nature of the personal data, in particular whether special categories of personal data are processed, (…), or whether personal data related to criminal convictions and offences are processed, (…)
- the existence of appropriate safeguards, which may include encryption or pseudonymisation.
- the possible consequences of the intended further processing for data subjects

The ICO further highlights that a key factor to take into consideration with respect to the compatibility assessment is whether the new purpose is "fair". This would entail considering "*how the new purpose affects the privacy of the individuals concerned and whether it is within their reasonable expectations that their data could be used in this way*."[50]

---

[44] Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation' (2013) WP203, 16 and 52-53 examples 9-10 and 11 in Annex 3

[45] 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) 30 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018

[46] See also European Data Protection Supervisor, 'Preliminary Opinion of the European Data Protection Supervisor. Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy' (EDPS 2014) 14 <https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf> accessed 16 October 2018

[47] Ibid

[48] Or on EU or Member State law relating to matters specified in Article 23 (general article on restrictions relating to the protection of national security, criminal investigations etc.).

[49] Further processing of personal data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes shall not be considered incompatible with the original processing purposes. However, the conditions of Article 83(1) (which sets out safeguards and derogations in relation to processing for such purposes) must be met.

[50] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 38 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

It follows that the further processing of the same personal data for another purpose (re-purposing) must be carefully assessed. The issues of data repurposing is by essence very relevant in the context of big data as performing algorithmic analysis upon a dataset is usually not the initial purpose for which the data was collected. Figure 4 aims to summarise the legal analysis that must be performed:

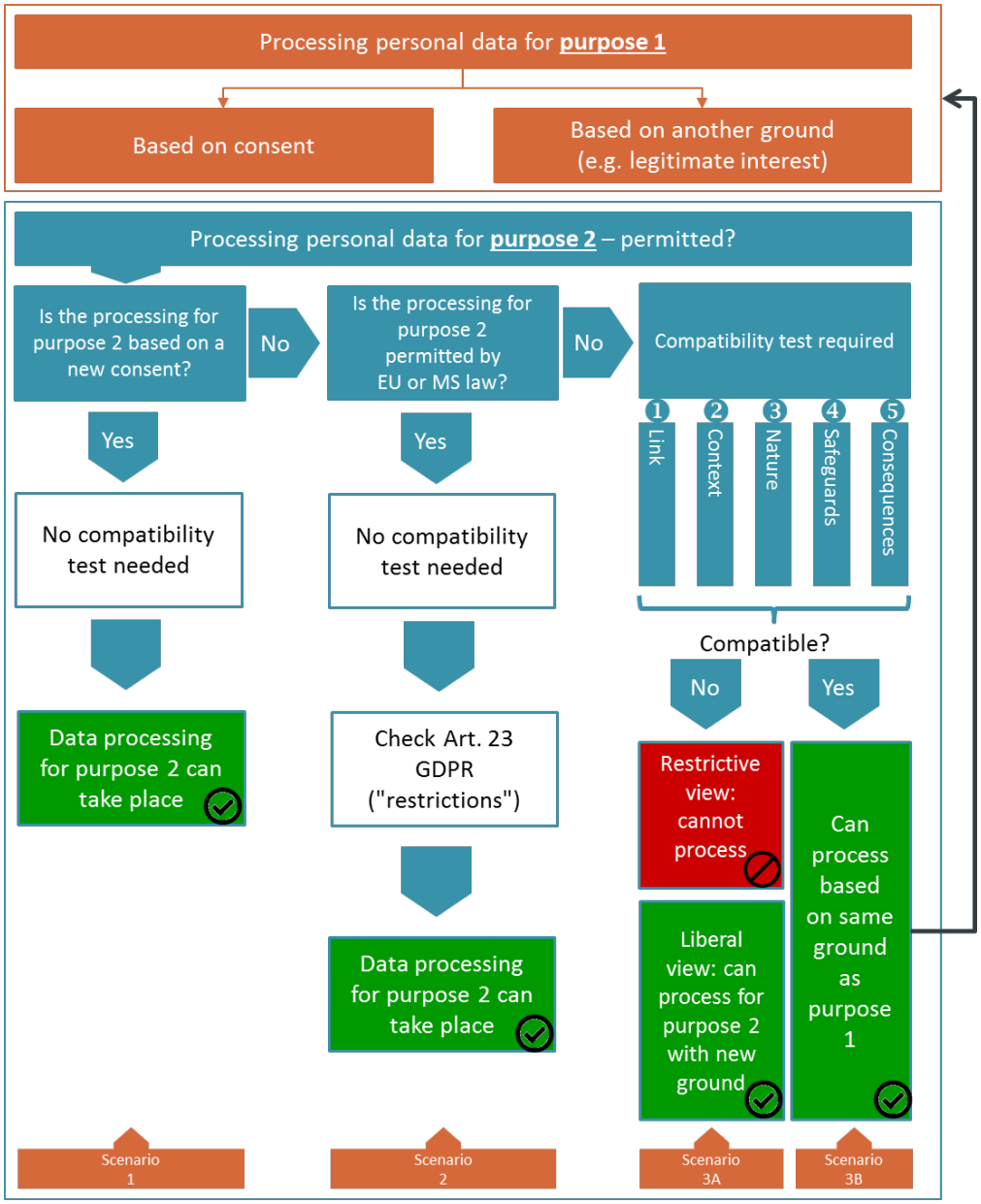


Figure 4: Graphic overview of the re-purposing assessment

With respect to **Scenario 1** depicted in the diagram above, it follows from the wording of Article 6(4)[51] as well as Recital 50 of the GDPR that "*where the data subject has given consent*

[51] *A contrario* reading of Article 6(4): "Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent (…)".

(…), the controller should be allowed to further process the personal data irrespective of the compatibility of the purposes." Accordingly, a compatibility test on the basis of the factors listed in Article 6(4) should not be carried out. In any event, all other principles and requirements of the GDPR remain fully applicable, including the strict requirements related to 'consent' but also the transparency principle and the information obligation in relation to the additional purpose(s).

In the event that the processing of personal data for another purpose is not based on consent, **Scenario 2** provides for another possibility foreseen in Article 6(4), *i.e.* the Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1). The latter provides for certain 'restrictions' such as national and public security, defence, judicial proceedings, etc. In those – rather exceptional – cases, the data controller will be entitled to further process the personal data for such other purpose.

The residual category of **Scenario 3** applies in the event that the processing for another purpose is not based on a new consent or permitted on the basis of a restriction provided under EU or Member State law. In such event, a compatibility test must be performed; taking into account the factors listed under Article 6(4) of the GDPR (see above).

- If the compatibility test is positive, the further processing may be carried out, on the basis of the same ground as for the first initial purpose (**Scenario 3B**).
- If, on the contrary, the compatibility test is negative, the situation becomes more problematic (**Scenario 3A**).
    - The Article 29 Working Party applies a conservative and restrictive view in its Opinion 03/2013 (issued before the adoption of the GDPR). It concludes that the further processing of personal data in any way that is incompatible with the initial purpose is unlawful and therefore not permitted. Following such view, the Working Party further clarifies that the data controller cannot simply consider the further processing to be a new processing activity disconnected from the previous one, and thus circumvent this prohibition by relying on another ground to legitimise the processing.[52]
    - In our view, a more liberal approach should be adopted in order to permit the further processing of personal data. Indeed, following the restrictive view of the Article 29 Working Party would not permit the uptake of new technological evolutions and would therefore not lead to a true data-driven economy.[53] Hence, it is recommendable to permit the data controller to base the further processing on another ground (however exhaustively listed in Article 6(1) GDPR) and to apply the necessary test (including, for instance, the legitimate interests test - see below) in order to legitimise the further processing. It goes without saying that all pertinent principles and obligations provided under the GDPR remain fully applicable even if a liberal approach would be chosen. Such

[52] Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation' (2013) WP203, 40

[53] See in the same direction Lokke Moerel and Corien Prins, 'On the Death of Purpose Limitation' (*IAPP*, 2 June 2015) <https://iapp.org/news/a/on-the-death-of-purpose-limitation/> accessed 16 October 2018

view may be substantiated by an *a contrario* reasoning deducted from Recital 50 of the GDPR which provides that "*no legal basis separate from that which allowed the collection of the personal data is required*" if the further processing is compatible with the purposes for which the personal data were initially collected. This seems to suggest that in case the further processing is not compatible with the initial purposes, a new legal basis – separate from the legal basis which allowed the collection of the personal data – is required.

The latter view is particularly relevant when considering big data analytics because applying the purpose limitation principle in a big data context can be particularly challenging.

3.1.3.3 Data minimisation

The general principle of "data minimisation" enshrined in Article 5(1)(c) of the GDPR provides that personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Also, the period for which the data are stored should be limited to a strict minimum. Finally, personal data should only be processed if the purpose of the processing cannot be fulfilled by other means.[54]

It is clear that the concepts of "data minimisation" and of 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*."[55] Against such background, some advocates of big data demand derogations from the central principles, particularly those of purpose limitation and data minimisation, and argue that these principles should not (or not fully) be applied to big data processing.[56] On the other hand, the EDPS affirms that "*Data protection authorities need to enforce data minimisation, which requires personal information only to be processed where 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed'*[57] (…)."[58]

It is important to stress the issues related to the quantity and quality of data but also in relation to the sources of such data. More particularly, even in a big data context, it is essential to determine whether the processed data is necessary and relevant for the purposes of the processing, or whether it is excessive.[59]

---

[54] The Belgian Privacy Commission puts emphasis on Recital 39 of the GDPR which provides that the data minimisation principle "*requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum. Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means*." See <https://www.autoriteprotectiondonnees.be/node/19242> accessed 22 October 2018

[55] 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/14-03-26_competitition_law_big_data_en.pdf> accessed 16 October 2018; See also Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 40 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[56] Ibid

[57] GDPR, art 5(1)(c)

[58] European Data Protection Supervisor, 'Opinion 8/2016. EDPS Opinion on Coherent Enforcement of Fundamental Rights in the Age of Big Data' (EDPS 2016) 7 <https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf> accessed 18 September 2018

[59] See in that sense Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 40 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

On such basis, the ICO is of the opinion that having well-managed, up-to-date and relevant data – rather than acquiring and keeping data just in case it may be useful – helps to improve data quality and contributes to the analytics.[60] In such context, the ICO provides some recommendations to abide by the data minimisation principle, in line with the concepts of "privacy by design" and "privacy by default". More particularly, organisations should[61]:

- articulate at the outset why they need to collect and process particular datasets;
- clarify what they expect to learn or be able to do by processing that data;
- ensure that the data is relevant and not excessive in relation to the purposes.

It follows that the data minimisation principle is closely linked to the purpose limitation requirement as any organisation must first determine the purposes of the processing and then establish that the data will be relevant and thus not excessive.

3.1.3.4 Accuracy

Personal data must be accurate and, where necessary, kept up-to-date; every reasonable step must be taken to ensure that inaccurate personal data, having regard to the purposes for which they are processed, are erased or rectified without delay. This principle should be read in conjunction with the data subjects' rights (see sub-Section 3.1.6 above).

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*."[62] It should be borne in mind that, in addition to three main features of big data (i.e., Volume, Velocity and Variety), several authors also refer to "Veracity" which relates to the ability of analysing datasets that comprise less controllable and accurate data.[63]

Despite the technical ability of analysing datasets that may comprise inaccurate data, compliance with the GDPR requires implementing measures to disregard the elements of a database that would be inaccurate. While this would require additional investments at different stages of the big data analysis in order to improve the quality of the data, it also provides the opportunity to improve the data management and ultimately contribute to a better analytics outcome.

3.1.3.5 Storage limitation

Personal data must 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. Personal data may be stored for longer periods insofar as the data will be processed solely for archiving

[60] Ibid 42
[61] Ibid 41
[62] 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 16 October 2018
[63] Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3; See also Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 43 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

purposes in the public interest, or scientific and historical research purposes or statistical purposes in accordance with Article 83(1) and subject to implementation of appropriate technical and organisational measures.

The GDPR does not specify the exact data retention periods given that these are necessarily context-specific. This being said, considering that data can be kept for "no longer than is necessary" in light of the purpose for which it was originally collected "assumes that each data element is collected only for a single purpose (or perhaps a small number of discrete purposes), and that this purpose was immediately apparent at the outset".[64] In reality, this is seldom the case.[65]

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.

Despite such challenges, the storage limitation principles requires any organisation to carefully assess the retention periods and determine whether data can be erased, but also whether it can be anonymised or pseudonymised (see also Section 3.4). The requirement to retain data for "no longer than is necessary" indeed only applies to *personal data*. Data which is not personal falls outside of data protection law and so, in principle, can be retained indefinitely. "Anonymisation throws up its own challenges, especially given European data protection authorities' strict views on what qualifies as effective anonymisation, but it is for many organisations often more achievable than full deletion".[66] If anonymisation is not achievable, an organisation may consider pseudonymising the data, which will still qualify as personal data but considered to be inherently less intrusive than 'ordinary' data, as examined in Section 3.4 above.

3.1.3.6 Accountability

The accountability principle relates to the ability to demonstrate compliance with the GDPR's principles, notably through the adoption of certain technical measures, the implementation of policies, the keeping of paper trails of decisions relating to data processing, the introduction of staff training programs, the performance of audits and impact assessments, or the adherence to approved codes of conduct.

The GDPR starts from the postulate that the processing of personal data is a risk for the rights and freedoms of individuals. Such risk must be taken into account and continuously re-assessed.[67] In this context, the GDPR imposes a risk-based approach. Companies are therefore

[64] Phil Lee, 'Privacy, Security and Information Law. To Keep or not to Keep: Data Retention Challenges and Solutions' (*Fieldfisher*, 30 July 2018) <https://privacylawblog.fieldfisher.com/2018/to-keep-or-not-to-keep-data-retention-challenges-and-solutions> accessed 16 October 2018
[65] Ibid
[66] Ibid
[67] The risk-based approach is enshrined in Article 24 relating to the responsibilities of the data controller. See also in that context Article 29 Data Protection Working Party, 'Statement on the role of a risk-based approach in data protection legal frameworks' (2014) WP218

required to appreciate in an objective manner the likelihood and severity of the risk to the rights and freedoms of individuals, taking into consideration the nature, scope, context and purposes of the processing.[68]

Recital 75 of the GDPR provides several examples of risky processing activities which could lead to physical, material or non-material damage. Among such examples, two are particularly relevant to big data analytics. Indeed, the processing is deemed risky where the processing involves a large amount of personal data and affects a large number of data subjects, as well as where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data.

This implies that when an organisation relies on big data analytics, an objective risk assessment method should be applied[69], keeping in mind that the GDPR appears to consider that big data analytics presents a risks for individuals.

Finally, it is worth mentioning that several scholars have discussed the concept of "algorithmic accountability", which can be particularly relevant in a big data context. In a nutshell, it relates to the ability to check that the algorithms used and developed are actually doing what is legitimately expected, and that they are not producing "discriminatory, erroneous or unjustified results".[70]

3.1.4 Legal grounds to process personal data (lawful processing)

3.1.4.1 Overview of the legal grounds

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 grounds exhaustively listed in GDPR. The processing of personal data is lawful only if[71]:

1. The data subject has given its consent;
2. It is necessary for the performance of a contract with the data subject or to take steps prior to entering into a contract;
3. It is necessary for the purposes of legitimate interests of the controller or a third party;
4. It is necessary for compliance with a legal obligation to which the controller is subject.
5. It is necessary to protect the vital interests of a data subject or another person where the data subject is incapable of giving consent;
6. It is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

[68] GDPR, art 24 and Recital 76. See also in relation to the risk-based approach Articles 32(1) and 33 to 35

[69] Commission de la protection de la vie privée, 'Rapport Big Data Rapport' (CPVP 2017) 14-15 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 16 October 2018

[70] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 52 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[71] Member States are permitted to introduce specific provisions to provide a basis under Articles 6(1)(c) and 6(1)(e) (processing due to a legal obligation or performance of a task in the public interest or in the exercise of official authority) for other specific processing situations (e.g. journalism and research). This is likely to result in a degree of variation across the EU. Also, article 9(2) of the GDPR sets out the grounds on the basis of which the processing of "sensitive personal data", which is otherwise prohibited, may take place.

The following paragraphs examine more in detail the various legal grounds that are the most likely to permit the processing of personal data in a big data context.[72]

3.1.4.2 Consent of the data subject

While 'consent' is the first ground that can permit the processing of personal data, it can quickly become a particularly difficult concept to comply with in light of its definition and the many conditions that must be met. Similarly, if an organisation decides to base its processing activity/ies on the individuals' consent, it shall keep in mind the strict view of the Article 29 Working Party which concluded that *"(…) if a controller chooses to rely on consent for any part of the processing, they must be prepared to respect that choice and stop that part of the processing if an individual withdraws consent. Sending out the message that data will be processed on the basis of consent, while actually some other lawful basis is relied on, would be fundamentally unfair to individuals. In other words, the controller cannot swap from consent to other lawful bases.*"[73]

Article 4(11) GDPR defines 'the consent of the data subject' as "*any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she by statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.*" It follows that for consent to be qualified as "consent" under the GDPR, it must fulfil the following four key cumulative conditions:

- *Freely given*: consent implies that a real choice is given to the data subject. Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment.[74] Consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority, or in an employment relationship. Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance.[75] When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.[76]
- *Specific*: the consent of the data subject must be given in relation to "one or more specific" purposes. Hence, consent should cover all processing activities carried out for the same purpose or purposes; when the processing has multiple purposes, consent should be given for all of them. Also, a data subject must be given a choice in relation to each of the purposes. Not only does this refer to the purpose limitation principles,

[72] The last three grounds being unlikely to permit big data analytics, they will not be analysed in this Deliverable.
[73] Article 29 Data Protection Working Party, 'Opinion 17/2017 on Consent under Regulation 2016/679' (2017) WP259, 23
[74] GDPR, Recital 42
[75] GDPR, Recital 43
[76] GDPR, art 7(4)

but it also aims to ensure a level of user control and transparency for the data subject.[77] In its opinion on consent, the Article 29 Working Party sums the elements that a controller must comply with in order to meet the requirement of 'specificity': the controller must apply (i) purpose specification as a safeguard against function creep, (ii) granularity in consent requests, and (iii) clear separation of information related to obtaining consent for data processing activities from information about other matters.[78]

- *Informed*: the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended. In line with the key principle of "transparency" and the data subject right of information, providing information to data subjects prior to obtaining their consent is essential in order to enable them to make informed decisions, understand what they are agreeing to, and for example exercise their right to withdraw their consent.[79]
- *Unambiguous*: the definition of consent itself requires a statement from the data subject or a clear affirmative act. This entails that it must always be given through an active motion or declaration in order to ensure that the data subject has clearly consented to the particular processing. Recital 32 suggests that this may be covered by: *"ticking a box when visiting a […] website, choosing technical settings […] or by any other statement or conduct which clearly indicates […] the data subject's acceptance of the proposed processing of their personal data. Silence, pre-ticked boxes or inactivity should therefore not constitute consent."*[80]
  The aspects related to the unambiguity of consent relates to the form of consent, which is further clarified by Article 7 and Recitals 32 and 42 of the GDPR. The latter stipulate that the request for consent should be provided in an intelligible and easily accessible form, using clear and plain language and it should not contain unfair terms.

The GDPR further includes an article 7 which enumerates two additional conditions related to consent:

- *Proof of consent*: the controller should be able to demonstrate that the data subject has given consent to the processing operation. In particular in the context of a written declaration on another matter, safeguards should ensure that the data subject is aware of the fact that and the extent to which consent is given.[81]
- *Withdrawal right*: the data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.[82]

It follows from the foregoing summary that the various conditions of consent are stringent and may be particularly difficult to meet in many instances. Therefore, relying on consent can

[77] Article 29 Data Protection Working Party, 'Opinion 17/2017 on Consent under Regulation 2016/679' (2017) WP259, 23
[78] Ibid 11
[79] Ibid 12
[80] GDPR, Recital 32
[81] Ibid
[82] GDPR, art 7(3)

be particularly difficult or may prove to be unpractical[83]or even impossible in a big data context, especially in its more complex applications.[84]

Some of the critiques and issues made regarding consent in relation to big data – already depicted in Deliverable D2.3 'Report on Ethical and Social issues' – can be summarised as follows:

- The opaque nature of data analysis can make it difficult for meaningful consent to be provided.[85]
- "*Citizens can find it difficult to recognise the connection between the different steps of Big Data processing practices that in some circumstances, which affects their ability to balance advantages (often short-term) vs. disadvantages (often long-term).*"[86]
- To what extent can an individual provide a valid informed consent with respect to big data analytics of his/her personal data?
- To what extent is it possible to anticipate the different data processing activities in big data analytics processes and have a valid informed consent?
- Consent is seen as something binary (yes/no choice), which is seen as incompatible with big data analytics notably due to its experimental nature and its propensity to find new uses for data.

This being said, the ICO is of the opinion that it is possible in certain cases to proceed with big data analytics on the basis of an individual's consent. More particularly, it pleads for a non-binary consent approach, which can be achieved both legally and technically: "*it may be possible to have a process of graduated consent, in which people can give consent or not to different uses of their data throughout their relationship with a service provider (…).*"[87]

In the same vein, the European Agency for Network and Information Security ("**ENISA**") does not exclude relying on consent in the context of big data. It is of the opinion that "*practical implementation of consent in big data should go beyond the existing models and provide more automation, both in the collection and withdrawal of consent. Software agents providing consent on user's* [sic] *behalf based on the properties of certain applications could be a topic to explore. Moreover, taking into account the sensors and smart devices in big data, other types of usable and practical user positive actions, which could constitute consent (e.g. gesture, spatial patterns, behavioural patterns, motions) need to be analysed.*"[88]

---

[83] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 30 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[84] 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) 11 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018

[85] Giovanni Buttarelli, 'A Smart Approach: Counteract the Bias in Artificial Intelligence' (*EDPS*, 8 November 2016) <https://edps.europa.eu/press-publications/press-news/blog/smart-approach-counteract-bias-artificial-intelligence_de> accessed 16 October 2018

[86] 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) 21 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018

[87] Ibid

[88] Giuseppe D'Acquisto and others, '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 16 October 2018. The Article 29 Working Party provides further examples of disruptive ways to obtain consent in its Guidelines on consent under Regulation 2016/679: *"Swiping a bar*

3.1.4.3 Processing necessary for the performance of a contract with the individual or to perform pre-contractual obligations

The ground provided under Article 6(1)(b) 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 must however be stressed that such contract must be concluded directly between the data controller and the data subject. Also, this particular ground is limited by the criterion of 'necessity', which therefore requires a direct and objective link between the processing itself and the contractual performance expected from the data subject.

In its guidelines on C-ITS[89] the Article 29 Working Party does not exclude relying on this particular ground in the context of disruptive technologies like connected cars, bearing in mind that it must concern processing that is "necessary for the performance of a specific and freely chosen contract to which the data subject is party"[90]. Also, such ground must be interpreted strictly and there must be a clear connection between assessment of necessity and compliance with the purpose limitation principle. In a C-ITS context, the Article 29 Working Party notably highlights two aspects of primary importance[91]:

- It is important to clearly determine beforehand the parties involved in the contract, in order to constraint the processing within the restricted perimeter of the sole actors involved in the scope of C-ITS, and avoid any further use by undetermined other parties. Accordingly, limiting the analysis to the contract between data subjects and a private road operator would be incomplete, since there may be other parties involved in the processing (car manufacturers and software developers, for instance) - either acting as joint controllers according to Article 26 of the GDPR, or as a whole in the context of a single consortium bearing the role of full controller - that may establish a contract with data subjects.[92]
- The rationale of the contract, its substance and goals must precede the processing itself, and controller(s) must test against these rationale and goals whether the data processing is necessary for the performance of the contract with each individual user, taking into account that cars may be driven by owners or other users.[93]

---

on a screen, waiving in front of a smart camera, turning a smartphone around clockwise, or in a figure eight motion may be options to indicate agreement, as long as clear information is provided, and it is clear that the motion in question signifies agreement to a specific request (e.g. if you swipe this bar to the left, you agree to the use of information X for purpose Y. Repeat the motion to confirm). The controller must be able to demonstrate that consent was obtained this way and data subjects must be able to withdraw consent as easily as it was given. "

[89] 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

[90] It does not cover situations where the processing is not genuinely necessary for the performance of a contract, but rather unilaterally imposed on the data subject by the controller. Also the fact that some data processing is covered by a contract does not automatically mean that the processing is necessary for its performance (Article 29 Data Protection Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (2014) WP217, 7).

[91] 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

[92] Ibid

[93] Ibid

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; and
- It implies a complex chain of actors (joint controller / processors) and multiple contracts (but little directly with the data subjects themselves).

Indeed, "*big data analytics, by its nature, is likely to represent a level of analysis that goes beyond what is required simply to sell a product or deliver a service. It often takes the data that is generated by the basic provision of a service and repurposes it*"[94].

3.1.4.4 Processing necessary for compliance with a legal obligation to which the controller is subject

The GDPR provides under Article 6(1)(c) a legal ground in situations where "processing is necessary for compliance with a legal obligation to which the controller is subject".

For such ground to apply, the obligation must be imposed by law. The law must fulfil all relevant conditions to make the obligation valid and binding, and must also comply with data protection law, including the requirement of necessity, proportionality and purpose limitation. More specifically, Article 6(3) and Recitals 41 and 45 make it clear that the legal obligation in question must be: an obligation of Member State or EU law to which the controller is subject and "clear and precise" and its application foreseeable for those subject to it. The Recitals emphasise that the relevant "legal obligation" need not necessarily be statutory. A legal obligation could cover several processing operations carried out by the controller so that it may not be necessary to identify a specific legal obligation for each individual processing activity.

Generally, it is unlikely that the processing in a big data analytics context can be based on a "legal obligation" as foreseen in the GDPR.

This being said, according to the Article 29 Working Party, such legal ground should not automatically be set aside in a technology context. Indeed, in its opinion on C-ITS, it is of the following opinion:

> *"Given the scope of C-ITS to improve road safety, foster transport efficiency and promote environmental sustainability, also through the implementation of this European wide interoperable system, the Article 29 Working Party finds that the long term legal basis for this type of processing is the enactment of an EU wide legal instrument (art.6(1)c of the GDPR). It is likely, given the projected prevalence of (semi-)autonomous cars that the inclusion of this technology in vehicles will become mandatory at some point in time, comparable to the legal obligation on car manufacturers to include e-call functionality in all new vehicles. Such a legal obligation should not allow for blanket collection and processing of personal data. The scope of*

[94] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 35 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

> *the legal obligation needs to be properly assessed, and validated as proportionate and strictly necessary in a democratic society, as is required under the protection offered by the applicable fundamental rights. This assessment and law making process should be initiated by the Commission as soon as possible, in order to prevent that the processing of location data and other personal data of EU citizens within C-ITS will take place without a legal basis, and would not be fully covered by an adequate level of protection."*

It follows that the EU authority appears to believe that the other grounds of Article 6 of the GDPR do not adequately fulfil some relevant elements in the context of C-ITS, which would therefore require the implementation of "*sector-specific Regulations for collecting and processing data in the field of Intelligent Transport System*".[95] Not only does this strict view ignore the complexity of the data value chain, but it would more generally require the adoption of numerous legislative instruments. Such adoption process would necessarily be slower than the development of mature technologies and would in any event unlikely be able to provide the legal certainty to all stakeholders involved in the analytics and processing chain associated to the disruptive technologies.

3.1.4.5 Processing necessary for the purposes of the legitimate interests of the controller

The protection of privacy and personal data is not absolute and often requires a balance of interests. Given the difficulties to rely on other grounds, such as consent, in a big data context (see above), the legitimate interests of an organisation may be a good alternative.[96] This ground does however not apply to processing carried out by public authorities in the performance of their tasks.

The GDPR includes Article 6(1)(f) which permits the processing of personal data for the purposes of "legitimate interests". It applies where "*processing 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, in particular where the data subject is a child.*"

It follows that in addition to exercising a balance between the interests of the organisation and those of individuals, the processing must be "necessary" for the legitimate interests of the controller (or third party). Hence, a certain threshold must be met (the processing on such basis must be more than just potentially interesting[97]), and that there is no other way of meeting the legitimate interest that interferes less with people privacy.[98] Relying on the

[95] 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, 13

[96] "*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 16 October 2018

[97] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 33 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[98] Ibid

balance of interests "*should, however, not be over-stretched so as to encompass any possible third-party interest.*"[99]

When a controller wishes to rely on its legitimate interests to process personal data, the GDPR further imposes to be transparent towards the data subjects by informing them in the privacy notices of the legitimate interests pursued by the controller or by a third party. It is also required to document the balancing test "*in a sufficiently detailed and transparent way so that the complete and correct application of the test could be verified - when necessary - by relevant stakeholders including the data subjects and data protection authorities, and ultimately, by the courts.*"[100]

The GDPR Recitals give examples of processing that could be necessary for the legitimate interest of a data controller.[101] These include:

- Recital 47: processing for direct marketing purposes or preventing fraud;
- Recital 48: transmission of personal data within a group of undertakings for internal administrative purposes, including client and employee data (please note that international transfer requirements still apply);
- Recital 49: processing for the purposes of ensuring network and information security, including preventing unauthorised access to electronic communications networks and stopping damage to computer and electronic communication systems; and
- Recital 50: reporting possible criminal acts or threats to public security to a competent authority.

Recital 47 also states that controllers should consider the expectations of data subjects when assessing whether their legitimate interests outweigh the interests of data subjects. The interests and fundamental rights of data subjects "could in particular override" that of the controller where data subjects "do not reasonably expect further processing." Several factors including the nature of the interests, the impact of the processing and any safeguards which are or could be put in place, must be considered when making a decision regarding whether an individual's rights would override a controller's legitimate interest.

Undeniably, relying on such ground requires a thorough assessment (a so-called "*Legitimate Interests Assessment*" or "*LIA*"). The Article 29 Working Party[102], as well as the Data Protection Network[103], have broken down the LIA into the following steps, providing illustrations, tips

[99] 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) 30 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018. In such context, the LIBE Committee also states that "the concept of data minimisation is relevant in this case, acting as a reminder that data processing shall always be as limited as possible."

[100] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (2014) WP217, 43

[101] The Article 29 Working already published numerous examples in its opinion 06/2014 on legitimate interests under Directive 95/46/EC.

[102] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (2014) WP217

[103] Data Protection Network, 'DPN Legitimate Interests Guidance – GDPR (version 2.0)' (*DPN*, 2018) <https://www.dpnetwork.org.uk/dpn-legitimate-interests-guidance/> accessed 16 October 2018

and key questions to help conducting the assessment, which must always be conducted fairly[104]:

- Identify and qualify an interest as 'legitimate' or 'illegitimate'.
- Carry out a necessity test in order to determine whether the processing is necessary to achieve the interest pursued.
- Carry out a balancing test to assess whether the data controller's interest override the fundamental rights or interests of the data subject; and
- Take into account (additional) safeguards such as to minimise data collection, to be transparent towards data subject, to provide control mechanisms to data subjects (e.g. for opting-out), etc.

The latter step has led the European Data Projection Supervisor ("**EDPS**") to adopt a rather optimistic view of the legitimate interest in the context of IoT and big data analytics: *"in big data cases where it is difficult to strike a balance between the legitimate interests of the organisation and the rights and interests of the data subject, it may be helpful to also give people the opportunity of an opt-out"*[105]. Such opt-out solution is a typical example of safeguards (i.e. compensating controls or measures) which may be put in place to protect the individual, or to reduce any risks or potentially negative impacts of processing.[106]

However, even when safeguards can be put in place, relying on legitimate interests requires any *"(…) big data organisation […] to have a framework of values against which to test the proposed processing, and a method of carrying out the assessment and keeping the processing under review. It will also have to be able to demonstrate it has these elements in place, in case of objections by the data subjects or investigations by the regulator."*[107]

It follows that a Legitimate Interest Assessment will prove to be rather difficult in a big data context, and more generally whenever new technologies are involved. Such difficulty is notably voiced by the Article 29 Working Party in its opinion on the recent developments on the Internet of Things ("**IoT**")[108]. More specifically, it takes a rather strict view by concluding in an IoT context that *"the processing of an individual's personal data is likely to affect significantly his/her fundamental rights to privacy and to the protection of personal data in situations where, without IoT devices, data could not have been interconnected or only with great difficulty".* It further added that "*In the light of the potential seriousness of that interference, it is clear that such processing will hardly be justified by merely the economic interest which an IoT stakeholder has in that processing. Other interests pursued by the*

[104] The controller should not attempt to make the assessment unfair or biased, and must always give due regard and weighting to the rights and freedoms of individuals.

[105] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 33-34 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 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 16 October 2018

[106] Data Protection Network, 'DPN Legitimate Interests Guidance – GDPR (version 2.0)' (*DPN*, 2018) 18 <https://www.dpnetwork.org.uk/dpn-legitimate-interests-guidance/> accessed 16 October 2018,

[107] 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 16 October 2018

[108] Article 29 Data Protection Working Party, 'Guidelines on the Recent Developments on the Internet of Things' (2014) WP223, 15

controller or by the third party or parties to whom the data are disclosed must come into play".

3.1.5 Core obligations under the GDPR

Data controllers have to implement appropriate measures in order to comply with the GDPR.

3.1.5.1 Overview of the core obligations

The list below aims to summarise the different obligations imposed upon data controller and/or data processors:

- *Records of processing activities:* controller and processors have a general obligation to keep internal records (registers) of data processing activities and to make such records available to the supervisory authority on request. It is therefore necessary for any company to map all its data processing activities and to keep detailed records.
- *Data Protection Officer:* a company may be obliged to appoint a Data Protection Officer ("**DPO**"), who will notably have to be independent and report to the highest management level of the company. The DPO will have to inform and advise the company and its employees, monitor compliance with the GDPR and internal policies, as well as act as the point of contact with the authorities and cooperate with them. In case the company concludes it is not required to appoint a DPO, it will need to document its legal assessment and make it available to the authorities upon request. In the event that the organisation does not have an establishment in the EU, but that the GDPR nevertheless applies, it must in principle designate a "representative" in the EU.
- *Security:* all personal data processed shall be subject to appropriate technical and organizational security measures. Such measures shall take into account: (i) the state of the art, (ii) the costs of implementation, (iii) the nature, scope, context and purposes of processing; and (iv) the risk of varying likelihood and severity for the rights and freedoms of individuals. It is therefore necessary to assess all security measures in place and review them where necessary (see Section 3.3 above for a more in-depth analysis of the breach-related obligations).
- *Data breach notifications*: the GDPR requires notification of a security breach leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data. Thus, in the case of data breach, the organisation must inform, within a short period of time, as the case may be, the (co-)controller, the supervisory authority and/or the affected individuals. It is therefore necessary to anticipate possible incidents and adopt/review internal policies (see Section 3.3 above for a more in-depth analysis of data breach-related obligations).
- *Data Protection Impact Assessments:* in certain circumstances, a company will have to carry out Data Protection Impact Assessments ("**DPIA's**") and in particular when the processing is "likely to result in a high risk" for individuals, such as for instance when implementing new technologies or in case of systematic and extensive evaluation of personal aspects relating to individuals which is based on automated processing. It is

therefore necessary to adopt the necessary internal procedure to determine when DPIA's are required, and carry out such assessment where needed.

- *Data protection by design and data protection by default:* any company must adopt adequate technical and organizational measures aimed at effectively implementing the GDPR and ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. This requires, in particular, conducting an audit and, where necessary, modifying internal policies and IT means related to the processing of personal data. Similarly, the introduction of a new technology, process or activity that involves the processing of personal data requires determining if it integrates privacy by design and by-default.

The below sub-Sections further analyse the last two obligations of the above list, which are particularly relevant in a big data context.

3.1.5.2 Data protection impact assessment (DPIA)

The GDPR formally codifies the requirement for Data Protection Impact Assessments to be carried in case a processing presents a "high risk" to the rights and freedoms of natural persons. Although DPIAs are mandated for high risk processing, the Article 29 Working Party recommends that they should be seen as a tool for accountability and could be used in somewhat wider situations as well. In the view of the Working Party, conducting a DPIA will help organisations build compliance (at the outset) and demonstrate compliance at a later date.

This being said, strictly speaking, DPIAs are required 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.[109] While Article 35(1) clearly indicates that processing "using new technologies" is likely to result in a high risk, Article 35(3) and Recital 91 provide a non-exhaustive list of occasions when DPIAs are required:

- systematic monitoring of a publicly accessible area on a large scale (e.g. CCTV) (Article 35(3);
- monitoring publicly accessible areas on a large scale, especially when using optic-electronic devices or for any other operations where the competent supervisory authority considers that the processing is likely to result in a high risk to the rights and freedoms of data subjects, in particular because they prevent data subjects from exercising a right or using a service or a contract, or because they are carried out systematically on a large scale (Recital 91);
- systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person (Article 35(3);

[109] The Article 29 Working Party suggests that if a controller concludes that processing does not need a DPIA, because it is not likely to result in a high risk, then this should also be documented.

- processing on a large scale of special categories of data or data relating to criminal convictions and offences (Article 35(3);
- large-scale processing operations which aim to process a considerable amount of personal data at regional, national or supranational level and which could affect a large number of data subjects and which are likely to result in a high risk, for example, on account of their sensitivity, where in accordance with the achieved state of technological knowledge a new technology is used on a large scale as well as to other processing operations which result in a high risk to the rights and freedoms of data subjects, in particular where those operations render it more difficult for data subjects to exercise their rights (Recital 91);
- processing for taking decisions regarding specific natural persons following any systematic and extensive evaluation of personal aspects relating to natural persons based on profiling those data or following the processing of special categories of personal data, biometric data, or data on criminal convictions and offences or related security measures (Recital 91).

For other processing activities, the organisation should determine whether it 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. More specifically, it is considered that a risky processing may:

- lead to physical, material or non-material damage;
- give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage;
- occur where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data;
- occur where special categories of data and data relating to criminal convictions is processed;
- occur where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles;
- occur where personal data of vulnerable natural persons, in particular of children, are processed;
- occur where processing involves a large amount of personal data and affects a large number of data subjects.

In addition to the above illustrations and elements provided by the GDPR to determine whether a DPIA may be required, Articles 35(4) and 35(5) of the GDPR allow national supervisory authorities to establish a list of processing operations that are necessarily subject to the requirement to conduct a DPIA ("black list") and a list of processing activities for which

no DPIA shall be required ("white list")[110]. Although such lists must be submitted to the European Data Protection Board (EDPB) and are subject to the consistency mechanism[111], the content of such lists may differ between the various Member States.[112] In the first five months of the application of the GDPR, twenty-two competent supervisory authorities had submitted their draft lists to the EDPB.

Finally, the national and European authorities[113] may published guidelines ("grey list") to assist controllers determining whether a DPIA may be required, as well as methodologies and tools to allow them carrying out the required DPIA.

It follows that a myriad of guidance, criteria and methodologies have been published since the entry into application of the GDPR.

An analysis of the various lists and guidance published from 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”, “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 the Internet of Things) should lead to the requirement to carry out a DPIA.

Even if interpretation issues may arise due to the possible discrepancies between the lists and guidance, it is very likely that data controllers active in the context of disruptive technologies will need to conduct one or more DPIA’s prior to the processing.

The GDPR provides for strict rules in case a DPIA must be carried out, which must be documented to identify and evaluate the possible risks as well as to determine and propose how risks can be limited or reduced.

First, with respect to the content (and the methodology) of the DPIA, the Article 29 Working Party recommends an iterative process[114], based on Article 35(7) of the GDPR[115].

---

[110] Recital 91 of the GDPR already considers that "*The processing of personal data should not be considered to be on a large scale if the processing concerns personal data from patients or clients by an individual physician, other health care professional or lawyer. In such cases, a data protection impact assessment should not be mandatory*".

[111] GDPR, art 63

[112] The WP29 states that while the draft lists of the competent supervisory authorities are subject to the consistency mechanism, this does not mean that the lists should be identical. The competent supervisory authorities have a margin of discretion with regard to the national or regional context and should take into account their local legislation. The aim of the EDPB assessment/opinion is not to reach a single EU list but rather to avoid significant inconsistencies that may affect the equivalent protection of the data subjects.

[113] Article 29 Data Protection Working Party, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining whether Processing is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679' (2017) WP248, 7

[114] Ibid 14

[115] GDPR, art 35(7): *“The assessment shall contain at least:*
(a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
(b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes;
(c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and
(d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.”

Second, a DPIA should be carried out at an early enough stage so that recommendations can be acted on. This may entail a need to re-assess later on. Periodic review is also likely although the Article 29 Working Party suggests that they should both be carried out continuously and re-assessed at least every three years or perhaps sooner, if circumstances require doing so.

Third, the data controller is rather free regarding the form of the DPIA. The Article 29 Working Party is not prescriptive and notes that there are various templates available for this. The guidelines take account of two relevant ISO documents: (i) risk management: ISO 31000:2009 on Risk Management – Principles and Guidelines;[116] and (ii) Protection Impact Assessments ("**PIAs**") in an information security context: ISO/IEC 29134 on Information technology – Security techniques – Guidelines for privacy impact assessment (project).[117]

In such context, the Article 29 Working Party proposes criteria to assess whether a DPIA template or methodology is sufficiently comprehensive to comply with the GDPR.

More particularly, the ICO also published in its 2017 report dedicated to big data, artificial intelligence and machine learning a specific annex related to "privacy impact assessment for big data analytics", in light of the GDPR.[118] In such guidance it relies on its PIA framework and their related code of practice (known as "PIA COP")[119], but includes a checklist of the key points for conducting a PIA/DPIA for big data analytics. The six steps identified by the ICO are as follows, in which we include some examples from the various checklists[120]:

- Step 1: Identify the need for a PIA
    - We have a DPO available for consultation on PIAs.
    - Our big data analysts use appropriate screening questions to help identify the need for a PIA.
    - If the direction of a big data project seems unclear, we err on the side of caution and begin the PIA process anyway.
- Step 2: Describe the information flows
    - Where possible, we clearly describe the predicted information flows for our big data project.
    - If the purposes of the processing are uncertain: we use only anonymised data, or we describe the information flows as the project progresses.
- Step 3: Identify the privacy and related risks
    - We ask ourselves questions about the proposed big data analysis to identify and record the associated privacy risks.
    - As the project develops we regularly return to these questions and develop new questions to identify and record any new risks.

[116] ISO 31000:2009(en), Risk management – Principles and guidelines
[117] ISO/IEC 29134, Information technology – Security techniques – Guidelines for privacy impact assessment
[118] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 99ff <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018
[119] Information Commissioner's Office, 'Conducting Privacy Impact Assessments Code of Practice' (ICO 2014) <https://www.pdpjournals.com/docs/88317.pdf> accessed 16 October 2018
[120] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 99-113 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018.

    - We assess whether the proposed big data analytics is the only method by which the project could be conducted.
    - We assess whether the proposed big data analytics is justified in relation to its potential benefits.
    - We consult internally and externally throughout the big data project.
- Step 4: Identify and evaluate privacy solutions
    - We identify and record appropriate measures to address the privacy risks previously identified.
    - As the big data analysis progresses and new risks are identified, we continue to identify and record measures to address these risks.
    - If the direction of a big data project is unclear, we use novel methods of obtaining consent and providing privacy notices.
- Step 5: Sign off and record the PIA outcomes
    - We obtain board-level sign-off for the measures identified to address the privacy risks of the proposed big data analytics.
    - We keep a record of the sign-off and the whole PIA process.
    - If we have identified high risks but not the measures to mitigate them, we consult the ICO before starting any data processing.
    - We produce and publish a PIA report.
- Step 6: Integrate the PIA outcomes back into the project plan
    - We ensure that the agreed privacy solutions are folded back into the big data project.
    - We regularly review our big data processing operations to check whether the privacy solutions are working as expected.

Fourth, although the Article 29 Working Party notes that the GDPR does not require this, it states that publication should be undertaken, either in full or in part, to demonstrate trust and accountability (in particular where members of the public could be impacted by the processing).

Finally, if the DPIA demonstrates a high risk for the data subjects involved, the controller must notify the supervisory authority and obtain that authority's opinion on the adequacy of the measures proposed in the framework of the DPIA. This is the case whenever risks cannot be mitigated and remain high - such as where individuals may encounter significant or even irreversible consequences, or when it is obvious that a risk may occur. In addition, Member State law may require that data controllers consult the authority in some cases (e.g. processing in the public interest in relation to public health), irrespective of the level of residual risk.

While the data controller is ultimately responsible for the DPIA, it may be required to seek external assistance. Hence, the data processors involved in the processing activity may be required to help. Also, the Data Protection Officer must be involved and must monitor performance of the DPIA. 'Where appropriate' the controller should seek the views of data subjects (e.g. via a survey or study). Finally, although not mentioned in the GDPR, others

should be involved if they are a relevant stakeholder (e.g. business unit responsible for the processing) and/or relevant expert (lawyer, security expert etc.).

3.1.5.3 Data protection by design and data protection by default

The GDPR includes, in the section related to the obligations of data controllers, a dedicated article related to the requirement to implement "data protection by design" and "data protection by default" measures. Such measures are linked to the core principles of the GDPR, and in particular the accountability principle and the related requirement to implement measures to demonstrate compliance with the GDPR, but also the purpose limitation, storage limitation and data minimisation principles (see also sub-Section 3.1.3 above).

More particularly, the requirement to adopt "data protection by design" measures 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). Said measures must be implemented in an effective way so as to integrate the necessary safeguards into the data processing in order to meet the requirements of the GDPR and to protect the rights of data subjects.

These obligations must be respected both at the time of the determination of the means for processing and at the time of the processing itself. Some elements to take into account while implementing the measures are (i) the state of the art; (ii) the cost of implementation; (iii) the nature, scope, context and purposes of the processing; and (iv) the risks of varying likelihood and severity for rights and freedoms of individuals posed by the processing.

As for the compliance with the "data protection by default" requirement[121], 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, rethinking physical design of certain products and services as well as data sharing initiatives. Moreover, organisations must focus on the requirement to take technical measures to meet individuals' expectations in order to notably delimit what data will be processed for what purpose, only to process the data strictly necessary for the purpose for

[121] GDPR, art 25(2)

which they are collected, implement a "privacy-first" approach with default settings, to appropriately inform individuals and provide them with sufficient controls to exercise their rights, and implement measures to prevent personal data from being made public by default.

The GDPR imposes requirements related to "data protection by design" and "data protection by default" only on data controllers. However, in practice, the entire data value chain is impacted.

Indeed, the GDPR imposes upon controllers a general duty to "*use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject*."[122] This imposes an indirect obligation to have processors develop services, products and processes in line with data protection by design and by default. In the same vein, those providers that do not process data but develop product and technology for controllers and processors are *"encouraged to take into account the right to data protection when developing and designing […] to make sure that controllers and processors are able to fulfil their data protection obligations"*.[123] Failing to develop GDPR-compliant tools would in practice push these providers out of the market.

The long-standing research carried out under the concept of "privacy by design" may provide useful insights on how to comply with the "data protection by design" and "data protection by default" obligations. Indeed, the concept of "privacy by design" was developed in the 90's, notably by the Information & Privacy Commissioner of Ontario in Canada, where "privacy by default" was considered to be one of the seven foundational principles of privacy by design[124]: "*Privacy by design seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice. If an individual does nothing, their privacy still remains intact. No action is required on the part of the individual to protect their privacy — it is built into the system, by default."*[125]

The development of these concepts has served as a fundamental source of inspiration for the creation of dedicated requirements in the EU. In such context, the European Network and Information Security Agency – the only institution at EU level which has been equipped with the competence and resources to perform dedicated research regarding privacy and data protection by design and by default – has published in 2014 and 2015 (prior to the GDPR) some practical insights on the requirements of 'privacy by design':

- "Privacy and Data Protection by Design – from policy to engineering", December 2014;[126]
- "Privacy by design in big data. An overview of privacy enhancing technologies in the era of big data analytics", December 2015.[127]

[122] GDPR, art 28(1)
[123] GDPR, Recital 78
[124] It is thus not a standalone principle, like data protection by design and data protection by default.
[125] Ann Cavoukian, 'Privacy by Design: The 7 Foundational Principles' (PbD 2011) <https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf> accessed 16 October 2018
[126] George Danezis and others, 'Privacy and Data Protection by Design – from Policy to Engineering' (ENISA 2014) <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design> accessed 16 October 2018

In the latter document, ENISA examines in detail various measures that permit the effective implementation of the privacy by design obligation in a big data context. It notably emphasises the need to shift the discussion from 'big data versus privacy' to 'big data with privacy'.[128] In order to achieve such change, it is key to identify the privacy requirements as early as possible in the big data analytics value chain. ENISA provides in its report two useful tables. First, it is important to determine the most appropriate strategies, and second, to apply such strategies in the different phases of the big data value chain.

More concretely, in a big data analytics context, the various measures that may be implemented in order to abide by the data protection by design requirement can be summarised as follows[129]:

- Anonymisation and pseudonymisation measures (see Section 3.4 above for more details)[130];
- Security measures to prevent data misuse (e.g. access controls, audit logs, encryption, etc.);
- Data minimisation measures (see sub-Section 3.1.3.3 below for more details);
- Purpose limitation measures (see sub-Section 3.1.3.2 below for more details);
- Data segregation measures;
- etc.

In such context, relying on so-called "Privacy-enhancing technologies" or "PETs" allows implementing to a certain extent data protection by design and by default requirements on a technical level. Indeed, such technologies embody fundamental data protection principles by minimising personal data use, maximising data security and aim to empower individuals. Such PETs notably include the following:

- Do Not Track features: a setting used by all major browsers so that users can indicate to websites, advertisers and social media plugins that they don't want to be tracked.
- End-to-end encryption: a system of communication where only the communicating users can read the messages.
- Differential privacy: differential privacy makes it possible for companies to collect and share aggregate information about user habits, while maintaining the privacy of individual users.
- PIMS: systems that help give individuals more control over their personal data. PIMS allow individuals to manage their personal data in secure, local or online storage systems and share them when and with whom they choose.

---

[127] Giuseppe D'Acquisto and others, '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 16 October 2018

[128] Ibid 10

[129] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 72-74 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[130] "*The concept of privacy by design is often associated with the implementation of techniques to anonymise and pseudonymise personal data*", Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 72 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

- Sticky policies: machine-readable policies are attached within metadata to define individual's preferences and corporate rules as data travels across multiple parties, enabling users to improve control over their personal information.[131]

Although a lot of research has been carried out in the field of PETs, their uptake remains rather limited. In such context, it is worth mentioning the report published by ENISA regarding "PETs maturity assessment methodology and online repository", which aims to facilitate maturity evaluation of PETs and their presentation to end users.[132]

Privacy and data protection in the transport sector – Example 6

The past decade has seen the rise of new transportation modes such as ridesharing. Ridesharing services, such as Blablacar and Carma, 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.[133] 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.

It follows from the foregoing that the new requirements of "data protection by design" and "data protection by default" will require changes within organisation in order to adopt new approaches in the development of processes, services and products. These new obligations are also an opportunity for stakeholders in the data value chain to improve their offering by integrating or further developing privacy-enhancing technologies and solutions, and ultimately comply with many other requirements of the GDPR.

3.1.6 Rights of individuals

The GDPR aims to protect natural person in relation to the processing of personal data and therefore recognises several rights to such persons. A snapshot of the various rights of data subjects can be depicted in Figure 5:

[131] "Sticky policies" refers to the sticking of machine-readable policies to data to define allowed usage and obligations as it travels across multiple parties, enabling users to improve control over their personal information; Marco Casassa-Mont and Siani Pearson, 'Sticky Policies: An Approach for Managing Privacy across Multiple Parties' (2011) 44(9) Computer 60; Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 73 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[132] Marit Hansen, Jaap-Henk Hoepman and Meiko Jensen, 'Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies: Methodology, Pilot Assessment, and Continuity Plan' (ENISA 2015) <https://www.enisa.europa.eu/news/enisa-news/enisa-proposes-pets-maturity-assessment-methodology> accessed 16 October 2018

[133] 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

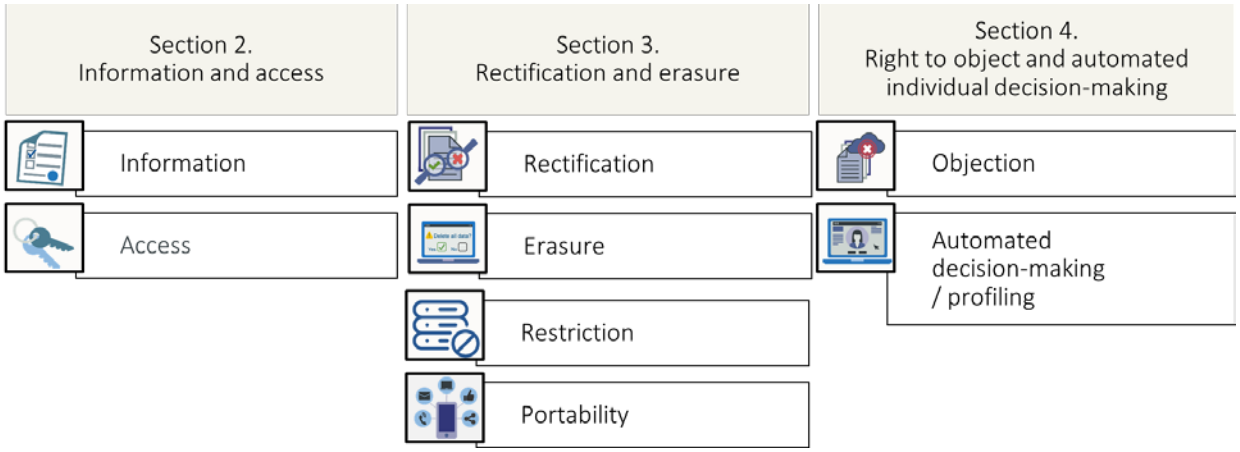


Figure 5: Overview of the data subjects' rights.

In addition to the above rights, the GDPR further provides for the strict procedures to respond to any data subject request, notably regulating issues with respect to the timing and the form of responses, 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.

It is particularly important to carefully consider the above rights and anticipate their concrete application. Although this can be difficult in a technology-rich environment, new technologies can however also be seen as an opportunity to allow individuals exercising their rights as "*innovative and responsible engineering can facilitate, among others, the exercise of individuals' rights of access, objection, opt-out, correction, as well as data portability*."[134]

The following sub-Sections aim to highlight some of the most important characteristics of each right when considering big data.

3.1.6.1 Information

One of the most important rights of data subjects is the right to information.

In order to ensure that personal data are processed fairly and transparently, data controllers must provide certain minimum information to data subjects regarding the collection and further processing of their personal data, such as for instance the purposes for which the data

[134] 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) 14 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018. The EDPS further details in this context the concept of 'functional separation', which is an area "where innovative engineering solutions are to be encouraged". More specifically, "in case an organisation processing data only wants to detect trends and correlations in the information rather than directly applying any insights they gained to the individuals concerned, 'functional separation' may potentially play a role in reducing the impact on the rights of individuals, while at the same time allowing organisations to take advantage of secondary uses of data. The objective of functional separation is to take technical and organisational measures to ensure that data used for research purposes cannot then be used to 'support measures or decisions' with regard to the individuals concerned (unless specifically authorised by these individuals)." 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) 15 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018

are processed. The data controller must provide information in a concise, transparent, intelligible and easily accessible form, using clear and plain language[135]. It should be borne in mind that new notices may be needed if personal data is processed for a new purpose not covered in the initial notice (e.g. for big data analytics purposes).

The GDPR however foresees certain (limited) situations where the controller does not have to provide the abovementioned information. This is the case if it would be impossible or if it would involve a disproportionate effort. In these cases, appropriate measures must be taken to protect individuals' interests and the information notice must be made publicly available.[136]

The requirement to inform individuals may be particularly relevant in a big data context. Indeed, as a matter of principle, national DPAs agree that big data analyses and the use of algorithms in general, require transparency and thus providing clear information to individuals.[137] Also, given that big data analytics often involve automated decision-making and profiling (see sub-Section 3.1.6.8 above for further details), it is important to provide information to individuals regarding the logic involved.[138]

In order to give the information in the most adequate manner, the data controller must carefully assess the best means. In such context, "*data protection authorities have long been recommending a 'layered' notice informing the data subjects about their data being processed step by step. This means providing the individual with the essential information about the processing at the point where the individual needs to make a decision based on the information (for example, an individual needs to know whether an app downloaded will have access to his location data before he chooses to install it), and providing further information in other formats, for example, via more detailed information on a website*."[139]

In a big data environment, the UK ICO believes that new technologies oblige organisations to be innovative and to find new ways of conveying the required information concisely. Although this can be rather challenging, it is "important to consider at an early stage of development

[135] The EU Commission may introduce in the future standardised icons. In such case these would then also need to be displayed to individuals.
[136] There is also no need to provide the information notice if there is an obligation under EU or Member State law for the controller to obtain/disclose the information; or if the information must remain confidential, because of professional or statutory secrecy obligations, regulated by EU or Member State law.
[137] See for instance Commission de la protection de la vie privée, 'Rapport Big Data' (CPVP 2017) <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 16 October 2018. Also, "*beyond the preservation of existing transparency and information obligations, it might be necessary to assess whether increasingly rapid and possibly unpredictable data processing practices need to be accompanied by other protection mechanisms that acknowledge that asymmetries in knowledge cannot be (easily) overcome*." 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) 30 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018
[138] European Data Protection Supervisor, 'Opinion 8/2016. EDPS Opinion on Coherent Enforcement of Fundamental Rights in the Age of Big Data' (EDPS 2016) 7 <https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf> accessed 16 October 2018
[139] 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) 11 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018; see also references made: Article 29 Data Protection Working Party, 'Opinion 10/2004 on More Harmonised Information Provisions' (2004) WP100; Article 29 Data Protection Working Party, 'Opinion 2/2009 on the Protection of Children's Personal Data (General Guidelines and the special case of schools)' (2009) WP160; and Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (2013) WP203, 16 and 52-53 examples 9-10 and 11 in Annex 3

how this information will be provided, and to look at the relationship between usability and privacy by design."[140]

3.1.6.2 Access

Data subjects have the right to file a subject access request and obtain from the data controller a copy of their personal data. The GDPR further requires that the controller responds to any request with supplementary information exhaustively listed under Article 15.

If the request is made in electronic form, the information should be provided in a commonly used electronic form (unless the data subject requests otherwise). This could impose costs on controllers who use special formats, or who hold paper records. Recital 63 also encourages controllers, where possible, to provide a secure system which would grant the data subject direct access to his/her data.[141]

The GDPR nevertheless recognises some limits in the event that the individual's access may adversely affect third parties. Hence, the GDPR provides that the right to receive a copy of the data shall not adversely affect such rights. Recital 63 notes that this could extend to protection of intellectual property rights and trade secrets (e.g., if release of the logic of automated decision-making would involve release of such information). However, the Recital also notes that a controller cannot refuse to provide all information on the basis that access may infringe third party rights.

Finally, the GDPR also contains a useful limiting provision in Recital 63, particularly relevant in a big data context.[142] If the controller holds a large quantity of data, it may ask the data subject to specify the information or processing activities to which the request relates. Such provision may thus constitute a ground for the controller to refuse a blanket access to all personal data processed and to grant a limited access to the data specifically identified by the individual. It remains however to be seen what amount of data constitutes a "large quantity of information" as required by Recital 63. In this respect, it shall also be noted that the Recital does not go on to say explicitly that there is any exemption due to large volumes of relevant data: the limitation seems to have more to do with the specificity of the request, rather than the extent of time and effort required on the controller's part – although the two may, of course, be linked.

This being said, applying the data access right in a big data context will remain relevant, yet difficult in practice. The UK ICO acknowledges such difficulty due to "*the volume and variety of*

[140] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 64,66 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[141] In this context, the concept of "data stores" or "data vaults" is relevant. The European Commission Communication on Big Data (COM (2014) 442 final) refers to and encourages the use of 'personal data spaces' for use-centric, safe and secure places to store and possibly trade personal data. See also 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) 13 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018; European Data Protection Supervisor, 'Opinion 4/2015. Towards a New Digital Ethics. Data, Dignity and Technology' (EDPS 2015) 12 <https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf> accessed 16 October 2018; Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 84-85 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018.

[142] GDPR, Recital 63

big data and the complexity of the analytics". The ICO however believes that "*such reasons cannot be an excuse for disregarding legal obligations. The existence of the right of access compels organisations to practise good data management. They need adequate metadata, the ability to query their data to find all the information they have on an individual, and knowledge of whether the data they are processing has been truly anonymised or whether it can still be linked to an individual.*"[143]

3.1.6.3 Rectification

Individuals can require a controller, without delay, to rectify inaccuracies in personal data held about them. In some circumstances, if personal data are incomplete, an individual can require the controller to complete the data, or to record a supplementary statement.

The EDPS, made a link, in a big data context, between the rectification right and the right of access. It stated that "*individuals must be empowered to better detect unfair biases and challenge mistakes arising from the logic used in algorithms to determine assumptions and predictions and a strong right of access and correction is a precondition to this.*"[144]

3.1.6.4 Erasure ("right to be forgotten")

Pursuant to Article 17 of the GDPR, individuals have the right to have their data 'erased' in certain specified situations; in essence where the processing fails to satisfy the requirements of the GDPR. Such right applies in strictly defined cases such as for instance when data are no longer necessary for the purpose for which they were collected or processed, if the individual withdraws consent to processing or when processing is based on legitimate interests.[145]

If the controller has made personal data public, and where it is obliged to erase the data, the controller must also inform other controllers who are processing the data that the data subject has requested erasure of those data. In the same vein, if the controller has to erase personal data, then it must notify anyone to whom it has disclosed such data, unless this would be impossible or involve disproportionate effort.

While these obligations are intended to strengthen individuals' rights (especially in an online environment), they can be rather burdensome in a context of new technologies such as big data where numerous stakeholders may be involved. The obligation is to take *reasonable steps* and account must be taken of available technology and the cost of implementation. However, the obligation is potentially wide-reaching and extremely difficult to implement.

Finally, a few strict exemptions may apply where the obligation would then not apply, such as in particular if the processing is necessary for the exercise of the right of freedom of expression and information, for compliance with a legal obligation or if required for the establishment, exercise or defence of legal claims.

---

[143] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) 46 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

[144] 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) 12 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018

[145] If the individual objects and the controller cannot demonstrate that there are overriding legitimate grounds for the processing.

3.1.6.5 Restriction

The right to restrict processing is a new right created under the GDPR. It nonetheless replaces the provisions in the Data Protection Directive on 'blocking'. In some situations, this right provides an individual with an alternative to requiring data to be erased; in others, it allows the individual to require data to be held in limbo whilst other challenges are resolved.

The right to restriction applies in strict cases such as when the accuracy of the data is contested by the data subject and the controller needs time to verify, if the processing is unlawful but the data subject objects to erasure or in case the data subject has objected to the processing and the controller needs time to verify whether its legitimate interests override those of the data subject.

If personal data are 'restricted', then the controller may only store the data. It may not further process the data unless: (i) the individual consents; (ii) the processing is necessary for establishment, exercise or defence of legal claims; (iii) for the protection of the rights of another natural or legal person; or (iv) for reasons of important (EU or Member State) public interests. When the controller wishes to lift a restriction, it must notify the individual in advance.

Where the data are processed automatically, the restriction should be effected by technical means and noted in the controller's IT systems. This could mean moving the data to a separate system, temporarily blocking the data on a website or otherwise making the data unavailable. It is therefore required to take into consideration the possibility for data to be restricted at the conception phase of IT services (including big data analytics systems). This further allows abiding by the "privacy by design" requirement under the GDPR.

If the data have been disclosed to others, then the controller must notify those recipients about the restricted processing, unless this is impossible or involves disproportionate effort.

3.1.6.6 Portability

While the right to data access provided under the GDPR already gives individuals the right to require their data to be provided in a commonly used electronic form, the right to data portability goes further. It indeed requires the controller to provide information – to the data subject or to another controller – in a structured, commonly used and machine readable form.[146] There is some uncertainty as to whether the format must be interoperable, or whether this is a matter of best practice which controllers are encouraged to adopt.

Whereas data subject access is a broad right, portability is narrower. It only applies to personal data:

- which is processed by automated means (no paper records);
- which the data subject has provided to the controller; and

[146] The EDPS has already discussed shortly the portability right and what it would require organisations to do in its big data report "Meeting the challenges of big data" in 2015 (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) 13 <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 16 October 2018)

- which is processed on the basis of consent, or processed to fulfil a contract or steps preparatory to a contract.

The obligation to port the data is stated to be without prejudice to the rights of other data subjects. Presumably, a controller should not port data to another controller (or to the individual) if this would breach the rights of others.

Given the numerous questions raised by the creation of this new ambitious right, the Article 29 Working Party adopted guidelines on the right to data portability on 5 April 2017.[147] Such guidelines clarify the conditions under which this new right applies taking into account the legal basis of the data processing (either the data subject's consent or the necessity to perform a contract) and the fact that this right is limited to personal data provided by the data subject. The guidelines also provide concrete examples and criteria to explain the circumstances in which this right applies. In this regard, the Article 29 Working Party considers that the right to data portability covers data provided knowingly and actively by the data subject as well as the personal data generated by his or her activity. The Article 29 Working Party clearly advocates for a broad interpretation, where "*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.[148] This new right can therefore not be undermined and limited to the personal information directly communicated by the data subject, for example, through an online form.

The Article 29 Working Party further clarified that the right to data portability applies only to data controllers. However, the guidance underlines that data processors will have contractual obligations under the GDPR to assist the controller "*by appropriate technical and organisational measures"* with responding to requests by individuals to exercise their rights. Thus, the Article 29 Working Party concludes that the data controller should "*implement specific procedures in cooperation with its data processors to answer portability requests*".

As a good practice, data controllers (and processors to the extent necessary) should develop the means that will contribute to address data portability requests, such as download tools and "Application Programming Interfaces" ("**API**").

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

[148] 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.

Privacy and data protection in the transport sector – Example 7

If, in a big data analytics context, the exercise of individuals' right to access their personal data collected through intelligent cars (e.g., by various sensors, smart meters, connected objects, etc.) or related to C-ITS could be difficult to put in practice, the exercise of the right to portability 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 to data portability.

3.1.6.7 Objection

Under the GDPR, data subjects a have a right to object to processing of their personal data on certain grounds, in addition to the right to object to processing carried out for the purposes of profiling or direct marketing (see below). In case a data subject raises objections, the data controller is required to demonstrate that it either has compelling grounds for continuing the processing, or that the processing is necessary in connection with its legal rights. If it fails to demonstrate that the relevant processing activity falls within one of the permitted grounds, it must cease that processing activity.

There is no right for an individual to object to processing in general. Only three specific rights to object are recognised under the GDPR, all relating to processing carried out for specific purposes, or which is justified on a particular basis. More particularly, data subjects can object to:

- Processing for direct marketing purposes: this is an absolute right; once the individual objects, the data must not be processed for direct marketing any further.
- Processing for scientific / historical research / statistical purposes: less strong than the right to object to direct marketing – there must be "grounds relating to [the data subject's] particular situation". There is an exception where the processing is necessary for the performance of a task carried out for reasons of public interest.
- Processing based on two specific purposes: (i) legitimate interest or (ii) because it is necessary for a public interest task/official authority. The controller must then cease processing of the personal data unless (i) it can demonstrate compelling legitimate grounds which override the interests of the data subject; or (ii) the processing is for the establishment, exercise or defence of legal claims.

In a big data context, the Belgian Privacy Commission states that "*one may wonder what the objection right actually means in the context of big data analytics. Article 21 of the General*

Data Protection Regulation ("GDPR") does not open any general right not to be subjected to profiling or to big data analyses. The question as to whether there is a right to object will depend on the type of processing that is being sought and it is necessary to examine what legitimate interests prevail (…). In any event, the above will not prevent certain decisions based on the results of big data analyses from being perceived by the data subjects as incorrect, unfair or discriminatory."[149]

3.1.6.8 Profiling and automated decision-making

Concepts of "profiling" and "automated decision-making"

Profiling	Automated decision-making
Defined under Article 4(4) GDPR: *"'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements"*	Indirectly defined under Article 22(1) as "*a decision based solely on automated processing, including profiling, which produces legal effects concerning [the data subject] or similarly significantly affects him or her.*"

• 'Automated' process
• Three stages:
  o Data collection
  o Automated analysis to identify correlations
  o Apply correlations to an individual to identify characteristics or behaviour patterns
• The evaluation involves a form of assessment / judgment about a person
• May involve statistical deductions
• May be used to make predictions
	• 'Automated' process
	• Ability to make decisions by technological means
	• Absence of human intervention
	• Based on all types of personal data (provided, observed, derived or inferred)

• Can take place without automated decisions	• Can be made with or without profiling

Table 2: Concepts of "profiling" and "automated decision-making"

[149] In a big data context, the Belgian Privacy Commission states that "*one may wonder what the objection right actually means in the context of big data analytics. Article 21 of the GDPR does not open any general right not to be subjected to profiling or to big data analyses. The question as to whether there is a right to object will depend on the type of processing that is being sought and it is necessary to examine what legitimate interests prevail (…). In any event, the above will not prevent certain decisions based on the results of big data analyses from being perceived by the data subjects as incorrect, unfair or discriminatory.*" (free translation) CPVP, 'Big Data Rapport' (CPVP 2017) 54 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 16 October 2018

Privacy and data protection in the transport sector – Example 8

The Article 29 Working Party has provided an example, in the transport sector, to illustrate the difference between "profiling" and "automated decision making", and more particularly a process that may start as a rudimentary "automated decision-making" and becoming a decision based on profiling[150]:

- Automated decision-making: *"imposing speeding fines purely on the basis of evidence from speed cameras is an automated decision-making process that does not necessarily involve profiling."*
- Becoming a decision based on profiling: *"if the driving habits of the individual were monitored over time, and, for example, the amount of fine imposed is the outcome of an assessment involving other factors, such as whether the speeding is a repeat offence or whether the driver has had other recent traffic violations."*

These concepts are particularly relevant when considering disruptive technologies given that *"advances in technology and the capabilities of big data analytics, artificial intelligence and machine learning have made it easier to create profiles and make automated decisions with the potential to significantly impact individuals' rights and freedoms."*[151] Indeed, the vast amount of data that can be collected about individuals is often used to categorise people based on their characteristics, behaviours, interests and habits, but also to make predictions and decisions.

This being said, these particular concepts of profiling and automated decision-making are not generally regulated by the GDPR. Some specific Recitals or Articles of the GDPR however make explicit reference to these concepts[152], such as in particular:

- Applicability of the GDPR ("behaviour") (Recital 24)
- Transparency (additional information to provide) (Recital 60 and Articles 13(2)(f) and 14(2)(g))
- Access right (Recital 63 and Article 15(1)(h))
- Right to object (Recital 70 and Article 21(1) and (2))
- Profiling (Recitals 71-72 and Article 22)
- National restrictions for public security (Recital 73)
- Data Protection Impact Assessment (Recital 91 and Article 35(3)(a))
- Binding Corporate Rules (profiling specific) (Article 47(2)(e))
- Tasks of the Board (profiling specific guidance) (Article 70(1)(f))

More generally, the key principles and obligations of the GDPR will apply in the context of "profiling" and "automated decision-making".

[150] Ibid 8
[151] Ibid 5
[152] For a complete list, see Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and profiling for the purposes of Regulation 2016/679' (as last revised and adopted on 6 February 2018) WP251rev.01, Annex 2

Privacy and data protection in the transport sector – Example 9

In its attempt to provide concrete examples, the Article 29 Working Party illustrates the general requirement to abide by the principle of lawfulness, fairness and transparency in a context of profiling and automated decision making as follows: *"Some insurers offer insurance rates and services based on an individual's driving behaviour. Elements taken into account in these cases could include the distance travelled, the time spent driving and the journey undertaken as well as predictions based on other data collected by the sensors in a (smart) car. The data collected is used for profiling to identify bad driving behaviour (such as fast acceleration, sudden braking, and speeding). This information can be cross-referenced with other sources (for example the weather, traffic, type of road) to better understand the driver's behaviour.*

The controller must ensure that they have a lawful basis for this type of processing. The controller must also provide the data subject with information about the collected data, and, if appropriate, the existence of automated decision-making referred to in Article 22(1) and (4), the logic involved, and the significance and envisaged consequences of such processing."[153]

Specific rules on "solely automated decision-making"

The GDPR includes strict rules in relation to "solely automated decision-making" under Article 22. Several criteria must be met in order for the regime to apply, which can be summarised in the following diagram:

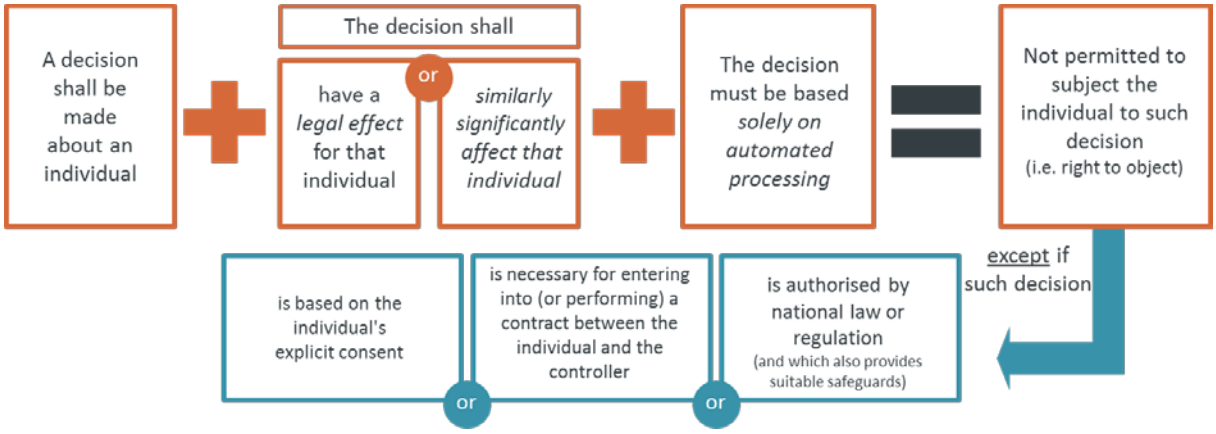


Figure 6: Legal criteria triggering the application of Article 22 GDPR and restrictions and permitted acts

[153] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and profiling for the purposes of Regulation 2016/679' (as last revised and adopted on 6 February 2018) WP251rev.01, 10

It follows from the above that several conditions must be met in order for the strict regime of Article 22 to apply:

- A decision shall be made: while the GDPR and its preparatory works do not provide guidance related to this first condition, the legal literature concludes nonetheless that "*making a decision about an individual person ordinarily involves the adoption of a particular attitude, opinion or stance towards that person. Such an attitude/stance can be of numerous kinds. (…) Alternatively, it can result in action being taken to influence the person with or without his/her knowledge*".[154]
- A decision that produces "legal effects" or similarly "significantly affects" the individual: the GDPR does not define and does not provide any guidance as to the meaning of these concepts. It is rather clear and straightforward to conclude that "legal effects" refers to effects that are able to alter or determine in part (or in full) an individual's rights and/or obligations.[155] The factual assessment can be somewhat more difficult. For instance, it can easily be concluded that activities such as credit monitoring fall within the scope of profiling as they could certainly have legal effects on a person. Targeted advertising is however more debatable as the 'legal effects' are not as straightforward.
  However, even in cases where the process does not have an "effect" on people's legal rights it could still fall within the scope of Article 22 if it produces an effect that similarly significantly affects data subjects.
  The idea of "significantly" affecting an individual presents more ambiguity. It remains indeed unclear whether this shall be understood in an objective or subjective manner. It can however be concluded that the strict regime does not apply to decisions that only affect the individual to a trivial or negligible extent.[156]
- A decision based "solely on automated processing": the right is narrow and only entitles data subjects to prevent "decisions made solely on automated processing", not the automated processing itself.[157]
  It follows that it is necessary to distinguish two possible concepts: (i) "fully automated decisions"; and (ii) "automated processing".[158] The question may however be more difficult in case the process includes a human intervention with the mere goal of

[154] Lee A Bygrave, *Data Protection Law: Approaching Its Rationale, Logic and Limits* (Information Law Series 10, Kluwer Law International 2002) 331. Based on the guidance provided by the UK Data Protection Authority, two simple illustrations can be given in this respect: (i) a bank/insurer uses algorithms to provide a yes or no decision relating to the online application of a loan/insurance; and (ii) an employer uses an automated productivity monitoring system to determine the worker's pay.
[155] Lee A Bygrave, *Data Protection Law: Approaching Its Rationale, Logic and Limits* (Information Law Series 10, Kluwer Law International 2002) 322. See also Lee A Bygrave, 21.
[156] Information Commissioner's Office, 'Automated Decision Taking' (*ICO*, 2016) <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-gdpr-say-about-automated-decision-making-and-profiling/> accessed 22 October 2018
[157] Eduardo Ustaran, *European Privacy: Law and Practice for Data Protection Professionals* (IAPP, 2011) 137
[158] Recital 58 of the GDPR provides as examples the "automatic refusal of an on-line credit application or e-recruiting practices without any human intervention."

circumventing the stricter legal regime of Article 22. This would require a case-by-case analysis of the decision-making process and the role of the "human" intervening.[159]

As depicted above, in certain circumstances clearly defined in three hypotheses under Article 22, decisions based solely on automated processing are permitted. It follows from Article 22(2) of the GDPR that the *decision* made solely on automated processing may nonetheless be made provided it is:

- authorised by EU or national law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests;[160] or
- necessary for entering into, or performance of, a contract between the data subject and a data controller; or
- based on the individual's explicit consent.[161]

In the event that automated decision-making is permitted on the basis of one of the three hypotheses, the controller shall abide by strict obligations clearly defined in the GDPR, which may be summarised as follows:

- Implement suitable measures to safeguard the rights of the individuals. For all permissible profiling, the GDPR (Recital 71, paragraph 2) recommends the controller to implement certain measures in order to ensure fair and transparent processing, taking into account the specific circumstances and context in which the personal data are processed: use appropriate mathematical or statistical procedures; implement technical and organisational measures to correct personal data inaccuracies and avoid errors; secure all personal data; and minimise the risk of "discriminatory effects against natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status, or sexual orientation."
  Furthermore, the GDPR reinforces the data subject's rights by suggesting that the controller provides specific information to the data subject, the right to obtain human intervention; the right to the individual to express his/her point of view, the right to obtain an explanation of the decision reached after such assessment, and the right to challenge/contest the decision.
- Inform the data subject and give access: in addition to the general principles and transparency obligations (see sub-Section 3.1.3.1 below), the GDPR provides specific requirements in case of 'profiling'. Articles 13 and 14 (information) and Article 15

[159] The Article 29 Working Party affirms in this context that "*The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing.*" (Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and profiling for the purposes of Regulation 2016/679' (as last revised and adopted on 6 February 2018) WP251rev.01, 21)

[160] Recital 71 of the GDPR specifies the following: "decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller".

[161] "Consent" represents one of the major improvements of the GDPR with respect to the legal regime relating to 'profiling'. It shall however be reminded that the GDPR approaches consent more restrictively. The requirements related to consent are further detailed in sub-Section 3.1.4.2.

(access rights) provide that the controller must give the following information to the data subject, at the time data is collected: the fact that profiling will occur, meaningful information about "the logic involved' and "the envisaged consequences of such processing for the data subject."

- Cease processing upon objection: restrictions on decisions based solely on automated processing (which could include profiling) apply if the decisions produce legal effects or similarly significantly affect the data subject. As detailed above, individuals have a right not to be subject to such decisions. Accordingly, even when profiling is permitted, a data subject has the right to object at any time.[162] Pursuant to Article 21, upon the data subject's objection to profiling that is otherwise authorised, the processing must cease unless: the controller can demonstrate compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject; or the processing is necessary for the establishment, exercise or defence of legal claims. Moreover, when processing is done for direct marketing purposes, including profiling, the data subject similarly has a right to object but in this case processing must cease and the controller is not authorised to continue under any circumstances.
- Conduct a data protection impact assessment. Article 35 of the GDPR imposes DPIAs to be performed in case of systematic evaluations or processing of personal data based on automated decision-making, or large scale processing of sensitive personal data (see sub-Section 3.1.5.2 below related to DPIAs).

3.1.7 Data transfers

The provision of big data analytics services may entail that the personal data collected and processed will be transferred internationally. This can be particularly true when relying on cloud computing services. The Civil Liberties, Justice and Home Affairs Committee of the European Parliament therefore concludes that "*thinking about the privacy and personal data protection implications of Big Data requires coordination with cloud computing policy*."[163]

Privacy and data protection in the transport sector – Example 10

One of the most obvious examples of international data transfers is airline companies. These indeed usually conduct business in multiple jurisdictions, some of which not being located in the European Economic Area (EEA). Not to mention the many actors they involve in each of the countries where these companies operate (e.g. airports, other airline carriers, booking websites and travel agencies, etc.).

There are also the foreign airlines that operate in the EEA, collect personal data from clients and transfer them to the non EEA country where their headquarters and the server where the data are to be processed are located.

[162] Indeed, in addition to the specific provisions relating to profiling (Article 22 mainly), Article 21 relating to the general "right to object" explicitly refers to profiling.

[163] 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) 31 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018

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.

The GDPR maintains the general principle that the transfer of personal data to any country outside the EEA[164] is prohibited unless that third country ensures an adequate level of privacy protection. Accordingly, transfers of personal data to "third countries" (i.e. outside the EEA, not ensuring an adequate level of protection) is restricted. In such cases, the data flow must be based on a particular means to allow the data transfer to take place, such as Standard/Model Contractual Clauses (SCCs)[165], Binding Corporate Rules (BCRs)[166], Code of conduct and certification, or derogations[167].

3.1.8 Summary

This Section dedicated to the analysis of some of the relevant issues related to privacy and data protection has shown the intricacies that some concepts, principles and obligations may cause in relation to a disruptive technology such as big data.

The main findings of the above sub-Sections may be summarised as follows:

- The concepts of "personal data" and "processing" are defined and interpreted very broadly, to such extent that the numerous obligations under the GDPR when performing big data analysis, including in the context of the transport sector, will apply in many circumstances. This may require limiting certain processing activities or technical developments to tackle the stringent rules included in the GDPR.
- The distinction between the concepts of joint-controllers, controllers in common and sub-processors is likely to be complex in relation to disruptive technologies, including in the big data value cycle. Hence, additional guidance and template agreements are more than welcome to clarify the relationships in certain intricate situations.
- 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 doing 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 contribute to the analytics.
- 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

[164] The European Economic Area includes the 28 EU countries and Iceland, Liechtenstein and Norway.
[165] A contract between the importer and exporter of the personal data, containing sufficient safeguards regarding data protection.
[166] Multinational corporations, international organisations and groups of companies wishing to transfer data within their corporate group comprising members established outside the EEA, can provide sufficient safeguards with respect to data protection using an internal code of conduct.
[167] 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.

to the grounds exhaustively listed in the GDPR are stringent and may limit or prohibit certain processing activities. Nonetheless, conducting thorough assessments, such as in the context of a Legitimate Interests Assessment, are likely to enable finding the most appropriate ground, while at the same time having the evidence to demonstrate the reasoning that has led to render the processing lawful, in accordance with the accountability principle.

- 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 particularly the case for the requirements to conduct Data Protection Impact Assessments and to implement privacy by design and privacy by default measures. Indeed, such obligations require thoroughly integrating privacy considerations in any processing activity and in the development of processes, services and products. It is ultimately necessary to adopt or amend internal procedures and to conduct audits on a regular basis.
- The numerous rights granted in the GDPR to individuals can be particularly challenging in 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, technologies can also be the way to allow individuals exercising their rights in a more innovative way, such as through Privacy Enhancing Technologies.
- As big data technologies generally rely on cloud computing services, it is necessary to carefully map the international data flows in order to put in place the most adequate solution to permit transfers to countries that do not provide an adequate level of protection.

Undeniably, the above summary, and this Section more generally, only 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 the following:

- 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; and
- The principle of proportionality must be taken into account.

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

Opportunities in relation to privacy and data protection in the context of big data in the transport sector	Challenges in relation to privacy and data protection in the context of big data in the transport sector
Having well-managed, up-to-date and relevant data may help improve data quality. This is also an opportunity to comply with the GDPR which requires implementing measures to disregard the elements of a database that would be inaccurate. Ultimately this is an opportunity to improve the data management and contribute to a better analytics outcome.	Certain key concepts (e.g. "personal data" and "processing") are defined and interpreted in such a broad way that the strictest rules of the GDPR would then apply, to such extent that would prohibit certain technologies of processing activities. Mere technical collected data could, due to the transformational impact of big data analytics, become personal data or even sensitive data and thus trigger the application of privacy and data protection laws
The development of coordinated and EU-wide guidance and templates, taking into account complex data processing activities, by EU and national levels are likely to be an opportunity to increase legal certainty to those involved in the data value chain, and ultimately benefit data subjects.	Some interpretations provided by certain authorities are conservative, too restrictive, and/or simplistic to such extent that it would prohibit certain technologies of processing activities (e.g. in relation to the further processing of personal data, to the extent of data portability, etc.).
The assessment requirements, and in particular Data Protection Impact Assessment and Legitimate Interests Assessments, are particularly relevant to big data analytics. They provide the opportunity to carefully examine the privacy implications and to identify the best measures to implement to comply with the GDPR, such as to minimise data collection, to be transparent towards data subject, to provide control mechanisms to data subjects (e.g. for opting-out), etc.	The distinction between “controller” and “processor”, taking into account the concepts of joint-controllership, controllers in common and sub-processors is complex in a big data context. Hence, identifying the role of each actor intervening in a big data context might prove to be difficult.

Opportunities in relation to privacy and data protection in the context of big data in the transport sector	Challenges in relation to privacy and data protection in the context of big data in the transport sector
Big data allows building and developing new consent models and provide more and novel automation, both in the collection and withdrawal of consent.	The core principles of the GDPR are particularly difficult to comply with, such as the storage limitation principle or the data minimisation principles. They are also somewhat contradictory to some big data analytics concepts as keeping data and being able to compare data from the past allow predictability.
The requirements of "data protection by design" and "data protection by default" will require changes within organisation in order to adopt new approaches in the development of processes, services and products. These new obligations can be an opportunity for stakeholders in the data value chain to improve their offering by integrating or further developing privacy-enhancing technologies and solutions, and ultimately comply with many other requirements of the GDPR.	Big data analytics often uses complex algorithms that are difficult to understand by data subjects. Such difficulty is poses issues in relation to the transparency principle and requirements.
Although considering the rights of data subjects and anticipating their concrete application can be difficult in a technology-rich environment, new technologies can also be seen as an opportunity to allow individuals exercising their rights as innovative and responsible engineering can facilitate, among others, the exercise of individuals' rights of access, objection, opt-out, correction, as well as data portability.	The various conditions of consent are stringent and may be particularly difficult to meet in many instances. Therefore, relying on consent can be particularly difficult or may prove to be unpractical or even impossible in a big data context, especially in its more complex applications.
	The grounds permitting the processing of personal data, exhaustively listed in the GDPR, will be generally difficult to apply in a big data context.
	This stakeholders active in the context of disruptive technologies, including big data analytics, are required to conduct one or more DPIA's prior to the processing, and continuously update such assessments.

Opportunities in relation to privacy and data protection in the context of big data in the transport sector	Challenges in relation to privacy and data protection in the context of big data in the transport sector
	The 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. This requires an extensive mapping of all international data flows, which can be rather challenging in a data-rich environment involving numerous actors.

Table 3: Summary table of opportunities and challenges in relation to privacy and data protection in the context of big data in the transport sector

3.2 (Cyber-)Security

For any organisation it is necessary to observe the legal obligations related to security and cyber-security. Such obligations not only derive from the GDPR, but also from other legislative instruments at both EU and national levels.

3.2.1 Security requirements under the GDPR

Foremost, the requirements relating to security under the GDPR will apply whenever personal data is processed. These obligations are however closely linked to those under the NIS Directive, examined below, and are in line with best practices applicable to information society systems that require adequate protection of assets.

3.2.1.1 Personal data governance obligations

Under the GDPR, any organisation must implement a wide range of measures to reduce the risk of non-compliance with the GDPR and to prove that it takes data governance seriously. Such measures create significant operational obligations and costs.

A general obligation is imposed upon data controllers[168] to adopt technical and organisational measures to meet the requirements set in the GDPR (and to be able to demonstrate that they have done so).[169] Operating a regular audit programme, implementing privacy by design and by-default measures, running Data Protection Impact Assessments, appointing a Data Protection Officer, etc. are all measures (examined above) 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.[170]

Big data may entail massive personal data processing operations, requiring therefore the implementation of adequate security measures. Pursuant to the Belgian DPA's Big Data Report, the greater the amount of personal data and processing used in a big data context, the greater will be the risk of violations associated with the security of these personal data, with the resulting adverse impact on the individual concerned.[171]

[168] The natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.
[169] GDPR, art 24
[170] GDPR, art 24(1)
[171] Data Protection Authority, 'Big Data Rapport' (CPVP 2017) 57 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 17 October 2018

(Cyber-)Security in the transport sector – Example 1

In the autumn of 2017, a USB key containing a number of security measures relating to the London airport (Heathrow) was found on the streets of West London[172]. This storage device contained highly confidential airport data, including maps (e.g., where the CCTV cameras and security gates are located), video-recordings, as well as multiple details about the security measures in place (e.g., security patrol schedules, types of badges required to access restricted areas, etc.).

Following the discovery of this USB stick and the disclosure of this information, the Heathrow airport authorities directly modified and strengthened the security measures in place to anticipate any threat and neutralise any risk.[173] In the hands of malicious individuals, these data could have been used as a weapon, which could most probably have had dramatic consequences. This example clearly shows (i) the risks to which the extreme computerisation of such dangerous means of transport as air transport is exposed; (ii) the importance of implementing an adequate (cyber-) security system, particularly in view of the said risk; and (iii) the consequences that the smallest, most humane security breach can have, notably from a financial perspective.

Furthermore, it shall be borne in mind that the GDPR imposes a high duty of care upon data controllers in selecting their personal data processing service providers, which will require procurement processes and request-for-tender documents to be regularly assessed, in particular on the security aspects.[174] In the context of data-rich environments, such as big data, the data controller should carefully reflect their security obligations in their respective agreements to be concluded with other actors, including processors and subprocessors.

3.2.1.2 Security of personal data processing

The GDPR requires data controllers and processors to "implement appropriate technical and organisational measures".[175] Such measures shall take into account the following elements:

- The state of the art;
- The costs of implementation;
- The nature, scope, context, and purposes of the processing; and
- The risk of varying likelihood and severity for the rights and freedoms of natural persons.

In assessing the appropriate level of security, account shall be taken in particular of the risks presented by the processing, notably from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise

[172] BBC, 'Heathrow Probe After 'Security Files Found on USB Stick'' *BBC News* (29 October 2017) <https://www.bbc.com/news/uk-41792995> accessed 17 October 2018

[173] Dan Wartburton, 'Terror Threat as Heathrow Airport Security Files Found Dumped in the Street' *Mirror news* (28 October 2017) <https://www.mirror.co.uk/news/uk-news/terror-threat-heathrow-airport-security-11428132> accessed 17 October 2018

[174] GDPR, art 28

[175] GDPR, art 32

processed.[176] In the context of big data, this entails that both data controllers and processors should continuously evaluate, manage and document those risks.[177]

Such risk-based approach, if carried out correctly, will not only lead to an effective and adequate security of the data processing, but may also be used to adhere to the accountability principle (see sub-Section 3.1.3.6 below) and which requires demonstrating compliance with the data protection principles and obligations laid down by the GDPR (see above).

The GDPR does not detail the security measures that can or should be put in place. It nonetheless provides the following specific suggestions for what types of security measures might be considered "appropriate to the risk":

1) the pseudonymisation and encryption of personal data (see Section 3.4 above for further details);
2) the ability to ensure the on-going confidentiality, integrity, availability and resilience of processing systems and services;
3) 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
4) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.[178]

The GDPR indicates that adherence to an approved code of conduct or certification mechanism may be used as an element to demonstrate compliance data governance obligations[179] as well as with security requirements.[180] Currently, such codes of conduct or certification mechanisms are being developed throughout the EU market. Such developments can only be encouraged in order to provide practical assistance to organisations.

3.2.2 Security requirements under the Network Information Security Directive

3.2.2.1 Context

The (minimal harmonisation) Network and Information Security Directive (the "NIS Directive" or "NISD") was adopted on 6 July 2016 to address the increasing challenges in relation to cybersecurity.[181] This EU legislation aims to respond to a drive to develop a common approach across Europe to address the potential for socio-economic damage caused by attacks on the

[176] GDPR, art 32(2)
[177] 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 17 October 2018
[178] GDPR, art 32(1)
[179] GDPR, arts 24(3) and 28(5)
[180] GDPR, art 32(3)
[181] 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 17 October 2018

network and information systems of operators of essential services and digital service providers.

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.[182] Some countries are however late in transposing the requirements of the NISD. It is therefore required to carefully consider the national obligations, which may be particularly relevant in a big data context, but also in the transportation sector.

In addition, the NIS Directive requires EU Member States to designate several new actors with the aim of attaining a high common level of security of network and information systems within the EU.[183] Thus, each EU Member State had to designate one or more national competent authorities ("**NCAs**") on the security of network and information systems, who shall monitor the application of the NIS Directive at national level.[184] Other key players coming onto the scene are the Computer Security Incident Response Teams ("**CSIRTs**").[185] Interactions with such entities notably include the requirement to notify security incidents either to the NCAs or to the CSIRTs (see Section 3.3 above dedicated to breach notification). The NCAs will have the necessary powers to urge essential and digital service providers to comply with their obligations under the NIS Directive.[186]

Furthermore, each EU Member State must select a national single point of contact, in order to facilitate the cross-border cooperation between the NCAs, the CSIRTs, and other relevant national authorities.[187] If an EU Member State decides to designate only one NCA, that NCA will also perform the function of single point of contact.[188]

It follows from the foregoing that the situation can become rather difficult, especially in case of cross-border services. For instance, in the context of the transport sector – which is particularly concerned by the NISD, as detailed below – many authorities may be involved, depending on the national systems[189]. In some countries, the competent authorities may depend on whether the service is public or private, or on the transportation mode. The table below aims to illustrate the complexity of the situation in certain Member States that have already transposed the NISD:

[182] 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
[183] NIS Directive, art 1(1)
[184] NIS Directive, art 8(1)
[185] NIS Directive, art 9
[186] NIS Directive, arts 15 and 17
[187] NIS Directive, art 8(3)
[188] Ibid
[189] Information based on the implementation status of the NIS Directive available at https://ec.europa.eu/digital-single-market/en/state-play-transposition-nis-directive

	Single point of contact	Competent authorities for DSPs	Competent authorities for OES (relevant to the transport sector)	National CSIRTs
United Kingdom 	National Cyber Security Centre (NCSC)	Information Commissioner's Office (ICO)	Transport (**Air**): Civil Aviation Authority (CAA), and Department for Transport Transport (**Rail**): Department for Transport (England, Wales, Scotland), The Department of Finance (Northern Ireland) Transport (**Water**): Department for Transport Transport (**Road**): Department for Transport (England, Wales), Scottish Ministers (Scotland), The Department of Finance (Northern Ireland) **Digital infrastructure**: Office of Communications (OFCOM)	National Cyber Security Centre (NCSC)
Germany 	Federal Office for Information Security / Bundesamt für Sicherheit in der Informationstechnik. It operates under the authority of the German Federal Ministry of the Interior.			
Sweden 	Swedish Civil Contingencies Agency (MSB)	Swedish Post and Telecom Authority	Transport (all sectors): Swedish Transport Agency	**Private** sector: Secretary of State for Information Society and Digital Agenda **Public** sector: Ministry of the Presidency and for the Territorial Administrations, through the National Cryptologic Centre

Spain 	National Security Council, through the National Security Department	**Private** sector: Secretary of State for Information Society and Digital Agenda **Public** sector: Ministry of the Presidency and for the Territorial Administrations, through the National Cryptologic Centre	Secretary of State for Security, -Ministry of Interior-, through the National Center for the Protection of Infrastructures and Cybersecurity (CNPIC)	**Private** sector: INCIBE-CERT, National Cybersecurity Institute **Public** sector: CCN-CERT, National Cryptologic Centre

Table 4: Illustration of the complexity of the situation in certain Member States that have already transposed the NISD

3.2.2.2 Scope of Application of the NISD

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

Operators of Essential Service (OES)

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 have to identify the operators of essential services established on their territory by 9 November 2018 based on the following criteria:

- entities providing a service which is essential for the maintenance of critical societal and/or economic activities;
- the provision of that service depends on network and information systems; and
- an incident would have significant disruptive effects on the provision of that service.

Regarding this last condition, the NIS Directive states that the EU Member States must consider the following factors when determining the significance of a disruptive effect:

- the number of users relying on the service concerned;
- the dependency of (one of) the sectors mentioned above on the service concerned;
- the impact incidents could have on economic and societal activities or public safety;
- the market share of the entity concerned;
- the geographic spread of the area that could be affected by an incident;

- the importance of the entity to maintain a sufficient level of the service, taking into account the availability of alternative means for the provision of that service; and
- any other appropriate sector-specific factor.[190]

Operators active in the following sectors may be included: energy, transport, banking, stock exchange, healthcare, utilities, and digital infrastructure.[191]

The transport sector is one of the services the NIS Directive considers to be essential. All kinds of transports are concerned[192], provided both by public and private entities[193]. The NIS Directive classifies in an Annex the following transportation modes:

Transport Mode	Sub-sector	EU sources of defined notions
Air 	Air carriers and airports	Point (4) of Article 3 of Regulation (EC) No 300/2008
	Airport managing bodies and entities operating ancillary installations contained within airports	Points (1) and (2) of Article 2 of Directive 2009/12/EC; Section 2 of Annex II to Regulation (EU) No 1315/2013
	Traffic management control operators providing air traffic control	Point (1) of Article 2 of Regulation (EC) No 549/2004
Rail 	Infrastructure managers and	Point (2) of Article 3 of Directive 2012/34/EU
	Railway undertakings and operators of rail related service facilities	Points (1) and (12) of Article 3 of Directive 2012/34/EU
Water 	Inland, sea and coastal passenger and freight water transport companies	Annex I to Regulation (EC) No 725/2004 (excluding individual vessels operated by those companies)
	Managing bodies of ports including their port facilities and entities operating works and equipment contained within ports)	Point (1) of Article 3 of Directive 2005/65/EC; point (11) of Article 2 of Regulation (EC) No 725/2004
	Operators of vessel traffic services	Point (o) of Article 3 of Directive 2002/59/EC
Road 	Road authorities responsible for traffic management control	Point (12) of Article 2 of Commission Delegated Regulation (EU) 2015/962
	Operators of Intelligent Transport Systems	Point (1) of Article 4 of Directive 2010/40/EU

Table 5: Transport modes targeted by the NISD

[190] NIS Directive, art 6
[191] NIS Directive, Annex II
[192] NIS Directive, Annex II
[193] NIS Directive, art 4(4)

It follows that, given their intimate ties in 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 NIS obligations set under the NISD and the implementing national legislations.

Digital Service Providers ("DSPs")

A digital service is described as "*any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services*".[194]

In contrast with the operators of essential services, 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 the following three different types of digital services: online marketplaces, online search engines, and cloud computing services.[195]

The fact that cloud computing services are targeted by the NISD is particularly relevant in a big data context, especially in light of its very broad definition. It is indeed defined as a digital service that enables access to a scalable and elastic pool of shareable computing resources, where the key elements are to be understood as follows[196]:

- 'Scalable': flexibly allocated by the cloud service provider in order to handle fluctuations in demand
- 'Elastic pool': provisioned and released according to demand
- 'Shareable': provided to multiple users who share a common access to the service

This being said, the other actors of the data value chain, taking an active role in the provision of services (such as in the transport sector), may also be concerned by the other concepts under the NISD. 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 OES), the NISD obligations may indirectly apply to suppliers of DSPs/OES as a result of flow down obligations.

[194] NIS Directive, art 4(5). A digital service provider without an establishment in the EU but providing services within the EU must appoint a representative. This representative will need to be established in one of the EU Member States where the digital services concerned are offered. In that case, the digital service provider shall be deemed to be under the jurisdiction of the Member State where the representative is established (NIS Directive, art 18(2)). Micro and small enterprises (as defined in Commission Recommendation 2003/361/EC) do not fall under the scope of the Directive.
[195] NIS Directive, arts 4(17)-(19)
[196] NIS Directive, Recital 17

Possible complex situations

(Cyber-)Security in the transport sector – Example 2

To illustrate the possible complexity that may arise from existing of OES and DSPs, one may rely on a Pilot from the Transforming Transport EU Project. The Tampere Pilot on Integrated Urban Mobility has three key objectives[197]:

- The first objective in the Tampere pilot is to provide tools for urban traffic management. This is 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.
- The second objective is to provide tools for informing drivers and public transport users regarding traffic status and traffic disruptions.
- The third objective is to improve urban logistics, 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 will be deployed and piloted.

In order to fulfil the above objectives, many players may come into play, which may qualify as OES and DSPs, or be obliged to take into account the NISD due to flow-down obligations.

Indeed, in such context, road authorities responsible for traffic management control and/or operators of Intelligent Transport Systems are likely to be involved. Similarly, cloud computing providers will be relied on. Finally, 'online market places' are likely to be involved and targeted by the NISD rules in the context of the third objective ('online market places' are being 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).

3.2.2.3 Security requirements under the NIS Directive

Under the new rules of the NID 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 operators of essential services and digital service providers 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

[197] Transforming Transport, 'Integrated Urban Mobility: Tampere Pilot' (*TT*, 2018) <https://transformingtransport.eu/domains/integrated-urban-mobility-tampere-pilot> accessed 17 October 2018

information systems.[198] The security measures shall take into account the state of the art, to ensure a level of security of network and information systems adequate to the risk.

Security of Operators of Essential Services

When considering the security aspects of Operators of Essential Services, it is worth looking into the following EU guiding documents:

- "Mapping of OES Security Requirements to Specific Sectors" published by ENISA in December 2017: such report provides a substantial and comprehensive mapping of the security requirements for OES, as they have been agreed in the "NISD Cooperation Group", to sector specific information security standards. It therefore associates the security requirements for OES, adopted by the Cooperation Group, with information security standards applicable to the sectors referred to in the Annex II of the NIS Directive. Moreover, "in order to achieve a common, baseline, cross-sector (horizontal) framework of security measures for the OES at EU level, the security requirements for the OES are primarily mapped to the most frequently used international information security standards by operators in each of these sectors." It therefore notably includes valuable insights on the security measures in the transport sector, referring notably to the 2012 Roadmap to Secure Control Systems in the Transportation Sector.[199]
- "Reference document on security measures for Operators of Essential Services" published by the NIS Cooperation Group in February 2018[200]: Such document doesn't aim at establishing 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.[201]

Security of Digital Service Providers

With respect to Digital Service Providers, the NISD stipulates that they must consider the following specific elements when determining the appropriate security measures[202]:

- the security of systems and facilities;
- incident handling;
- business continuity management;
- monitoring, auditing and testing; and
- compliance with international standards.[203]

[198] NIS Directive, arts 14 and 16

[199] The Roadmap to Secure Control Systems in the Transportation Sector Working Group, 'Roadmap to Secure Control Systems in the Transportation Sector' (August 2012) <https://ics-cert.us-cert.gov/sites/default/files/documents/TransportationRoadmap20120831.pdf> accessed 17 October 2018

[200] 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(0).pdf> accessed 17 October 2018

[201] Ibid 5

[202] 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)).

[203] NIS Directive, art 16(1)

In order to assist Member States and Digital Service Providers and to provide a common approach regarding the security measures for DSPs, ENISA has published "Technical Guidelines for the implementation of minimum security measures for Digital Service Providers".[204] This particular initiative has been achieved by examining current information and network security practices for the DSPs across the EU. It has brought light to some important findings that can add to existing security objectives and measures in information technology infrastructures in Europe.

Furthermore, with the aim of further clarifying some obligations under the NISD applicable to DSPs, the Commission adopted Implementing Regulation (EU) 2018/151 to specify 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.[205]

3.2.3 Security requirements under other legislations

It is important to note that other legal instruments may impose security requirements. 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 e-Privacy Directive[206]: it is required that providers of electronic communications services 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 Directive[207]: it complements the e-Privacy Directive by requiring providers of publicly available electronic communication networks and services to take appropriate measures to manage the risks posed to the 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 Directive[208]: privacy and data protection requirements apply to terminal equipment attached to public telecommunication networks. Radio equipment within certain categories or classes shall incorporate safeguards to ensure that the personal data and privacy of users and subscribers are protected.

[204] 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 17 October 2018

[205] 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

[206] 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 (e-Privacy Directive)

[207] 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)

[208] Directive 1999/5/EC of the European Parliament and of the Council on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity [1995] OJ L 91/10 (Radio Equipment Directive)

3.2.4 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 (the "**ISO**") and the International Electrotechnical Commission (the "**IEC**"), big data service providers can demonstrate to the regulator and to their customers that their systems are adequate, or at least that measures and processes have been implemented in terms of security.

Furthermore, several standards development organisations have created and are currently developing big data-specific standards: *e.g.*, the ISO/IEC, the Organisation for the Advancement of Structured Information Standards (OASIS), the International Telecommunication Union – Telecommunications sector (ITU-T), the World Wide Web Consortium (W3C), etc. It is essential for any big data service provider to follow up closely the evolutions in this respect.

It will however be necessary to determine whether such new standards sufficiently take into consideration the new requirements under the EU legal framework, and in particular the NIS Directive and the GDPR. Similarly, in the event an entity wishes to rely on existing standards, it must determine whether a review in light of the new EU rules is required.

3.2.5 Security in practice: challenges and recommendations

ENISA published in March 2016 a study related to big data security.[209] In its report, the Agency admits that big data security is not fundamentally different from traditional data security. However, big data raises specific security challenges because of (i) the data collected, aggregated, and analysed for big data analysis, (ii) the infrastructure used to store and house big data, and (iii) the technologies applied to analyse structured and unstructured big data.[210]

According to ENISA, security in a big data context poses the following three key challenges:

[209] Such report "*aims at identifying the key security challenges that the companies are facing when implementing Big Data solutions, from infrastructures to analytics applications, and how those are mitigated. The analysis focuses on the use of Big Data by private organisations in given sectors (e.g. Finance, Energy, Telecom).*"(Rossen Naydenov and others, 'Big Data Security. Good Practices and Recommendations on the Security of Big Data Systems (ENISA 2016) <https://www.enisa.europa.eu/publications/big-data-security> accessed 17 October 2018) More sectors (including the transport sector) and institutions (e.g. research centres, public organisations, and government agencies) are considered in the ENISA study.

[210] Ibid 13

Access control and authentication	*Secure data management*	*Source validation and filtering*
In a Big Data environment access to data is given to different people and entities in order to make computation and decisions. Maintaining the desired level of access control and authentication is a potential problem, as some of the entities might not have the capabilities to use the required security level.	The vast amount of logs the Big Data system collects is a key issue because the big volume of logs needs to be stored and protected. Proper protection is one issue, but there is also another – it should be possible to restore them in a secure way.	The essential use of a Big Data system is that it could collect information from many sources. Some of these sources may not be verified or trusted. Making analysis or decisions based on input that has not been verified could lead to potentially incorrect results.

Table 6: Key challenges related to the secure use of gig data identified by ENISA

In any event, the implementation of security measures obligations, notably to tackle the above challenges, can only make sense if they are implemented holistically, at all different stages of the data value cycle, to guarantee the continuity of services.[211] Concretely, such holistic approach entails that the following specific security issues and their possible mitigation measures ought to be considered throughout the different stages of the data value cycle.[212]

[211] Rossen Naydenov and others, 'Big Data Security. Good Practices and Recommendations on the Security of Big Data Systems' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-security> accessed 17 October 2018
[212] Ibid 8

Security issues	Mitigation measures
Integrity of the devices collecting data	Security testing procedures and audits, compliance with standards and certification mechanisms, access control and authentication.
Source validation	Encryption, security testing procedures and audits, risk assessment, source filtering, access control and authentication, monitoring and logging.
Infrastructure security	Security testing procedures and audits, compliance with standards and certification mechanisms, source filtering, access control and authentication, monitoring and logging.
Data security & secure data management	Encryption, security testing procedures and audits, access control and authentication, monitoring and logging.
Platform (e.g., cloud) security	Encryption, security testing procedures and audits, compliance with standards and certification mechanisms, risk assessment, access control and authentication, monitoring and logging.
Supply chain security	Security testing procedures and audits, compliance with standards and certification mechanisms, risk assessment.
Application software security	Security testing procedures and audits, compliance with standards and certification mechanisms, access control and authentication.
Interoperability of applications	Security testing procedures and audits, compliance with standards and certification mechanisms, access control and authentication.
Distributed denial-of-service attacks	Security testing procedures and audits, source filtering, monitoring and logging.
Unauthorised access	Encryption, security testing procedures and audits, compliance with standards and certification mechanisms, access control and authentication, monitoring and logging.

Table 7: Security issues and mitigation measures

In addition to applying mitigation measures internally, any company should ensure that safeguards are included in its contracts with, and can be enforced against, possible business

partners.[213] Any such agreement should therefore contain specific information security obligations as well as the warranties, indemnity provisions, and limitations of liability related thereto. In order to ensure the enforceability of such clauses, the contract should also provide for audit rights.[214] Furthermore, and inevitably, any agreement concluded for information security purposes should incorporate a comprehensive confidentiality clause.[215]

Better still, before entering into any business relations, an exhaustive due diligence of the envisaged business partner should be carried out, with a particular focus on information security.[216]

On the basis of its analysis of the security challenges in a big data environment, ENISA formulated the following five recommendations targeting various stakeholders:

Policy makers	**should focus on providing guidance for secure use of Big Data systems in the critical sectors.**
Authorities (of the critical sectors)	should encourage vendors to offer security authentication mechanisms.
Standardisation bodies	should adapt existing standards or create new security standards to include Big Data.
Big data providers or vendors	should invest in compliance with security standards for their products (devices, services, cloud etc.).
Industry players and vendors	should invest more into enhancing technical security skills of the staff on Big Data, through trainings and certifications.

Table 8: Key recommendations identified by ENISA

[213] Michael R Overly, 'Information Security in Vendor and Business Partner Relationships' in James R Kalyvas and Michael R Overly (eds), *Big Data: A Business and Legal Guide* (Auerbach Publications 2015) 27-30
[214] Ibid 30
[215] Ibid 29
[216] Ibid 23-27

(Cyber-)Security in the transport sector – Example 3

In 2015, Transportstyreisen, the Swedish Public Transport Agency, has outsourced to IBM its databases containing (personal) information on all vehicles in the country (e.g., driving licences, drivers addresses, etc.) including army and police vehicles, the identity of military and special forces workforce (state secret data), as well as technical specifications about certain bridges, roads, ports and subways.[217]

This outsourcing had to be done very quickly, which led to serious security and server implementation problems. All contractual precautions were not observed. All outsourced data were made available to persons unauthorised to process these data, including employees of IBM's Eastern Europe subcontractors, whose confidentiality had not been guaranteed. This data breach, entirely due to failed cyber security measures that were not implemented at all levels, has namely had significant political repercussions in Sweden.

This example highlights the importance of adopting this holistic approach. It further shows that the data value cycle, which involves many actors, increases the risk that personal data is insufficiently protected.[218] It is therefore important to identify all those involved in the big data analytics process and to put in place the necessary contractual arrangements including appropriate security obligations in order to comply with the integrity and confidentiality principle as well as other requirements under the GDPR.

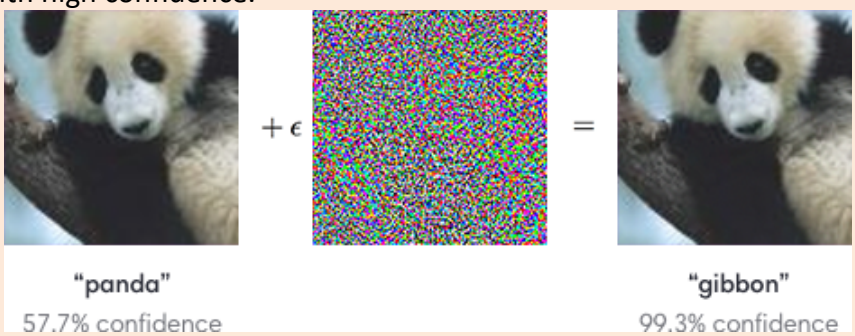
Despite the existence of guidance on the various security obligations and how to consider them practically, the security aspects remain difficult in reality and require further and continuous research. A good way to illustrate the complexities of applying appropriate security measures is through so-called “adversarial images”.

[217] Christina Anderson, 'Swedish Government Scrambles to Contain Damage From Data Breach' *The New York Times* (25 July 2017) <https://www.nytimes.com/2017/07/25/world/europe/ibm-sweden-data-outsourcing.html> accessed 17 October 2018

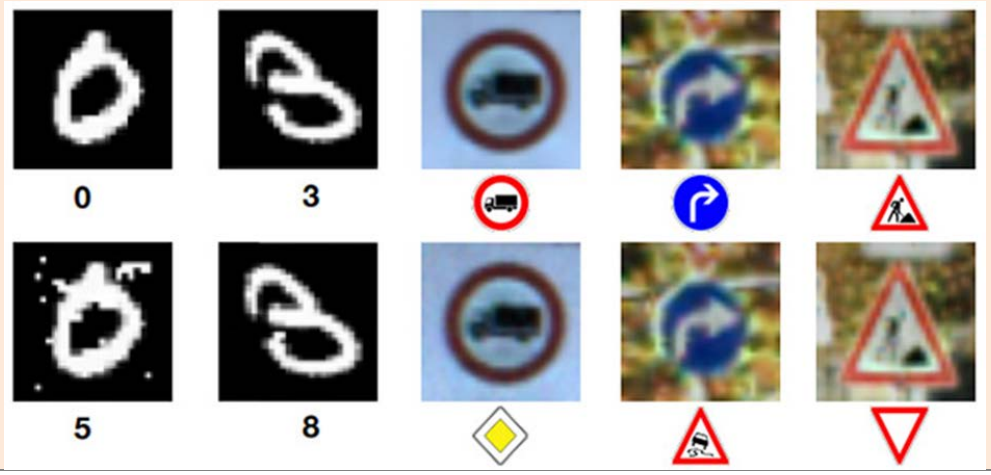
[218] Gloria González Fuster and Amandine Scherrer, 'Big Data and Smart Devices and Their Impact on Privacy. Study for the LIBE Committee' (European Parliament, Directorate-General for Internal Policies, Policy Department C Citizens' rights and constitutional affairs, 2015) 22-23 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16 October 2018; Article 29 Data Protection Working Party, 'Opinion 02/2013 on apps on smart devices' (2013) WP202 (In the case of mobile application the Article 29 Working Party details the fragmented nature of the app ecosystem, which includes app developers, app owners, app stores, manufacturers of Operating Systems (OS) and devices, and other third parties involved in the collection and processing of personal data from smart devices, such as analytics and advertising providers.)

(Cyber-)Security in the transport sector – Example 4

The concept of adversarial images consists in making minor changes to manipulate machine learning algorithms. To illustrate such specific security issue, OpenAI relies on the worked done by Cornell University[219]. 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."[220]



This 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. The Institute of Electrical and Electronics Engineers published an article to illustrate how "Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms".[221] 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."[222] 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."[223]



[219] Ian J. Goodfellow, Jonathon Shlens and Christian Szegedy, 'Explaining and Harnessing Adversarial Examples' (2015) arXiv:1412.6572
[220] Ian Goodfellow and others, 'Attacking Machine Learning with Adversarial Examples' (*OpenAI*, 24 February 2017) <https://blog.openai.com/adversarial-example-research/> accessed 17 October 2018
[221] 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 17 October 201
[222] Ibid
[223] Ibid

3.2.6 Summary

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

- Make a risk assessment (evaluate, manage and document the risks);
- Carefully assess the available security measures on the market;
- Continuously assess the adequacy of the implemented measures in light of the evolving risks and the available measure; and
- Adequately reflect the security aspects in the various contracts between the stakeholders.

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

Opportunities in relation to (cyber-)security in the context of big data in the transport sector	Challenges in relation to (cyber-)security in the context of big data in the transport sector
Big data analytics can contribute to the detection of security issues.	Big data analytics may pose specific security issues such in relation to access control and authentication, secure data management, and source validation and filtering.
The cyber-security requirements can ultimately help creating a cyber resilient culture across the organisations, notably giving the opportunity to train and certify staff.	The NIS Directive has increased the security requirements for some actors of the data value chain, including those indirectly impacted due to flow-down obligations.[224] This notably requires to (i) enable the on-going confidentiality, integrity, availability and resilience of systems and services (including those processing personal data); (ii) enable the ability to restore the availability and access to data in a timely manner in the event of incidents; and (iii) regularly test, assess and evaluate the effectiveness of security measures.

[224] Jasmien César and Julien Debussche, 'Novel EU Legal Requirements in Big Data Security: Big Data – Big Security Headaches?' (2017) 8(1) JIPITEC 79 <https://www.jipitec.eu/issues/jipitec-8-1-2017/4534> accessed 17 October 2018

Opportunities in relation to (cyber-)security in the context of big data in the transport sector	Challenges in relation to (cyber-)security in the context of big data in the transport sector
Actors of the data value chain, including authorities, standardisation bodies, service providers, vendors and industry players can develop together standards, certification mechanisms, seals, marks and codes of conduct, which could be tailored to the transport sector.	Threats to security are an ever-evolving issue which requires keeping up-to-date risks assessment and upgrading measures.
The heightened security requirement permit to enhance supply chain trust and resilience by engaging third party suppliers and customers in cybersecurity processes and business continuity measures[225]	The NISD is a minimal harmonisation Directive with possible discrepancies across the EU. It further requires Digital Service Providers to self-assess whether they fall within the scope of the NISD. Moreover, the interaction with authorities may prove difficult given the possible existence of multiple competent authorities in each country. Ultimately, the legal assessment may be complex in cases of multi-modal and cross-border transport services involving numerous actors.

Table 9: Summary table of opportunities and challenges in relation to (cyber-)security in the context of big data in the transport sector

[225] 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-BE/insights/2017/08/guide-to-the-nis-directive-for-transportation-companies> accessed 17 October 2018

3.3 Breach-related obligations

The present Section focuses on the applicable legal obligations, which derive from the GDPR, but also, where relevant, from other legislative instruments at both EU and national level.

Firstly, it should be noted that the legal concept of "data breach" does not coincide with the technical definition of "data breach".

As elaborated by E. Damiani in a big data context, there exist two sub-categories of threats on a technical level; *i.e.* (big) data breach and (big) data leak.[226] In such context, data breach refers to the theft of a data asset by intruding into the IT infrastructure, whereas data leak covers the disclosure of a data asset at a certain stage of its lifecycle.[227]

The legal notion of data breach, however, encompasses both technical definitions of data breach and data leak. Indeed, data breach in a legal context does not necessarily entail the malicious behaviour of a third party, but is also established in case (personal) data is disclosed without interference of a threat actor – *e.g.*, losing an unencrypted device.

3.3.1 Breach Notification obligations in the telecommunications sector

The Directive concerning the processing of personal data and the protection of privacy in the electronic communications sector[228] (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.[229]

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.[230] 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.[231]

The e-Privacy Directive is currently being reviewed in the framework of the EU Digital Single Market ("**DSM**") strategy. In this respect, the EU Commission held a public consultation, the report of which was made available in August 2016.[232] 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

[226] Ernesto Damiani, 'Toward Big Data Risk Analysis' in IEEE (ed), *Proceedings of the 2015 IEEE International Conference on Big Data* (IEEE, 2015) DOI:10.1109/BigData.2015.7363966

[227] Ernesto Damiani and others, 'Big Data Threat Landscape and Good Practice Guide' (ENISA 2016) 18 <https://www.enisa.europa.eu/publications/bigdata-threat-landscape> accessed 17 October 2018

[228] e-Privacy Directive

[229] e-Privacy Directive, art 4(3)

[230] 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

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

[232] Summary report available online at <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-evaluation-and-review-eprivacy-directive>

Directive given their "overlap" with the breach notification obligations under the GDPR (see below).[233] On 10 January 2017, the EU institutions adopted a draft e-Privacy Regulation, which would be directly applicable in all EU Member States.[234] 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.[235]

3.3.2 Data and privacy breach notification obligation

3.3.2.1 Scope of the obligation

The GDPR requires the notification 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*."[236]

It follows from such definition that many types of incidents will be considered as data breaches within the meaning of the GDPR. It 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 adequate data protection authorities across the EU (as well as to possible other authorities across the world in certain large breaches).

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

[234] 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

[235] 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.

[236] GDPR, arts 4(12) and 33

Breach-related obligations in the transport sector – Example 1

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 around 57 million Uber users around the world data such as names, email addresses and phone numbers.[237]

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."[238] 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.[239]

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.[240] The conclusions of such investigations have not yet been published.

[237] Dara Khosrowshahi, '2016 Data Security Incident' (*Uber Newsroom*, 21 November 2017) <https://www.uber.com/newsroom/2016-data-incident/> accessed 17 October 2018 and Austin Carr, 'Uber to Pay $148 Million in Settlement Over 2016 Data Breach' (*Bloomberg*, 26 September 2018) <https://www.bloomberg.com/news/articles/2018-09-26/uber-to-pay-148-million-in-settlement-over-2016-data-breach> accessed 17 October 2018

[238] Julia Apostle, 'The Uber Data Breach Has Implications for us all' *Financial Times* (27 November 2017) <https://www.ft.com/content/e2bf6caa-d2cb-11e7-a303-9060cb1e5f44> accessed 19 October 2018

[239] Austin Carr, 'Uber to Pay $148 Million in Settlement Over 2016 Data Breach' (*Bloomberg*, 26 September 2018) <https://www.bloomberg.com/news/articles/2018-09-26/uber-to-pay-148-million-in-settlement-over-2016-data-breach> accessed 17 October 2018

[240] European Commission, 'WP29 Has Established a Taskforce on the UBER Data Breach Case' (*European Commission*, 29 November 2017) <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=609786> accessed 17 October 2018

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

Duty	Timing	Exemption
Data processor to notify data controller	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	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 (in close cooperation with the supervisory authority)	Without undue delay.	Notification is not required if: 1. The breach is unlikely to result in a high risk for the rights and freedoms of individuals; or 2. Appropriate technical and organisational protection measures were in place at the time of the incident (e.g. data encryption); or 3. Measures have been taken, subsequent to the incident, ensuring that the risk to the right and freedoms of individuals is unlikely to materialise; or 4. It would trigger disproportionate efforts. However, in this case, a public communication or similar measure to inform the public is required.

Table 10: Breach notification requirements under the GDPR

On 6 February 2018, the Article 29 Working Party finalised its guidelines on personal data breach reporting (issued in draft in October 2017).[241] In such guidance, the EU authority clarifies certain aspects such as for instance:

[241] Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679' (2018) wp250rev.01

- When does the clock start? Although breaches generally need to be notified "without undue delay", the GDPR includes a 72 hours deadline for notification by the data controller to the supervisory authority, as from the moment the controller becomes *'aware'* of the breach. The WP29 considers that this requires a *'reasonable degree of certainty*' that a security incident has occurred which has led to personal data being compromised. It further notes that, whilst in some cases this may be clear at the outset, in other case it may take some time to establish that personal data has been compromised. The WP29 emphasises that any investigation by the controller should begin as soon as possible: the aim should be to establish quickly whether a breach has taken place. A more detailed investigation can then follow later.

  In such context, the WP29 gives examples to assist with determining when the clock starts to run:

Loss of unencrypted media	⇨	Controller aware as soon as realises the media has been lost. Fact that controller does not know whether unauthorised persons have gained access not relevant
Third party advises controller they have been sent personal data by the controller in error	⇨	Controller immediately aware of the breach
Controller detects possible network intrusion	⇨	Controller should establish if there is personal data on the system that has been compromised. Aware once controller confirms this is the case
Cybercriminal contacts controller with ransom demand after hacking system	⇨	Controller immediately aware

Table 11: WP29 examples of start 72-hour period for notification

- When does a breach not need to be reported? There is no need to report breaches to data protection authorities if the breach is 'unlikely to result in a risk to the rights and freedoms of natural persons'. Furthermore, controllers should only report to individuals when the breach poses a 'high risk' – to avoid unnecessary notification fatigue. There are examples throughout the guidelines of the WP29 of what may, or may not, need to be reported and there is an Annex of worked examples, which do, or do not, need to be reported.

3.3.2.2 Notifications in practice

The breach notification obligation under the GDPR evidently only applies in case of a breach of personal data. It is therefore essential to carefully assess, in the event of an incident, the nature of the data exposed. If such assessment shows that no personal data has been affected, in principle no data breach notification is required under the GDPR. In this respect, it

could reasonably be advocated that a breach of anonymised data or encrypted data, the key to which cannot be retrieved by a third party, does not need to be notified under the GDPR.

Therefore, appropriate technical and organisational measures should be implemented to be able to detect promptly whether a personal data breach has taken place and to immediately inform the supervisory authority and the individual, if needed.[242] Such measures include the keeping of good logs, which facilitates a swift and efficient forensic investigation in case of an incident.

The personal data breach notification by the data controller to the supervisory authority must at least mention the following information:[243]

(i) The nature of the breach, including the categories and approximate number of individuals as well as personal data records affected;

(ii) The name and contact details of the data protection officer or any other contact point that could provide more information;

(iii) The likely consequences of the breach; and

(iv) The measures (proposed to be) taken by the data controller to address the breach, including any measures to mitigate its negative effects.

The WP29 emphasizes that the focus is on assessment of risk – so precise numbers are not needed, but factors relevant to risk should be highlighted (i.e. special categories of data, vulnerable groups). It also suggests that if the breach is caused by a processor – and if the processor has caused a breach for multiple controllers – that the controller 'may find it useful to name its processor [in the report] if it is at the root cause'.[244]

In case it proves impossible to provide such information simultaneously within 72 hours, the GDPR allows providing such information in phases.[245] However, the notification should indicate the reasons for the deferment, and the missing information should be provided without further undue delay.[246]

The communication to the affected individuals must detail in clear and plain language the nature of the personal data breach, recommendations to mitigate possible adverse effects, as well as the information listed under (ii), (iii) and (iv) above.[247]

In line with the principle of accountability, further elaborated in the Section dedicated to the GDPR (see above), the data controller must document any personal data breach as well as the corrective measures taken in order to allow the supervisory authority to assess compliance with the data breach notification obligations.[248]

[242] GDPR, Recital 87
[243] GDPR, art 33(3)
[244] Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679' (2018) WP250rev.01, 15
[245] GDPR, art 33(4)
[246] GDPR, Recital 85
[247] GDPR, art 34(2) and Recital 86
[248] GDPR, art 33(5)

3.3.3 Notification obligation under the NISD

Under the NIS Directive, operators of essential services and digital services providers must notify without undue delay to the NCA or the CSIRT incidents having a significant impact on the continuity or provision of the services.[249]

The NIS Directive – given its nature as a Directive – is not directly applicable in the EU Member States but needs to be implemented in each national Member State law. 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.

3.3.3.1 Notification in practice

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

Operators of essential services	Digital service providers
• the number of users affected by the incident; • the duration of the incident; and • the geographical spread of the incident.[250]	• 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.[251]

Table 12: Factors to determine the significance of an incident

In addition to the above general rules included under the NISD, the following clarification documents have been published:

[249] NIS Directive, arts 14(3) and 16(3)
[250] NIS Directive, art 14(4)
[251] NIS Directive, art 16(4)

Operators of essential services	Digital service providers
• "Reference document on Incident Notification for Operators of Essential Services – Circumstances of notification"[252], published by the NIS Cooperation Group in February 2018.[253] Such document details the incident notification scheme for OES but also the parameters used to measure the impact of incidents. It also examines the intricacies of cross-border situations and the interplay of the NISD with notification requirements in other legislations (including the GDPR). • "Reference document on Incident Notification for Operators of Essential Services – Formats and procedures"[254], published by the NIS Cooperation Group in May 2018.[255] 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	• 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 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.[256] Such document notably clarifies four situations in which digital service providers are required to notify the relevant national competent authority 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

[252] 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

[253] The NIS Cooperation Group is established by the NISD 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.

[254] 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

[255] The NIS Cooperation Group is established by the NISD 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.

[256] 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

alignment in the implementation of the NIS Directive across the EU.	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 DSP, to facilitate alignment in the implementation of the NIS directive across the EU. • "Incident notification for DSPs in the context of the NIS Directive"[257] report published by ENISA on 27 February 2017. Such report includes a comprehensive guideline on how to implement incident notification for DSPs.

Table 13: Overview of EU guidelines related to NISD notification requirements

In case an operator of essential services depends on a digital service provider for the provision of such essential services, any significant impact on the continuity of those services due to an incident affecting the digital service provider must be notified by that operator.[258] The NIS Directive remains silent as to whether, in such circumstances, the digital service provider 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 digital service provider contractually.

The notified NCA or CSIRT shall inform other Member States affected.[259] 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.[260]

[257] 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

[258] NIS Directive, art 16(5)

[259] NIS Directive, arts 14(5) and 16(6)

[260] NIS Directive, arts 14(6) and 16(7)

Pursuant to the NIS Directive, the EU Member States may not impose any further notification requirements on digital service providers, unless for the protection of essential State functions and for the preservation of law and order.[261]

3.3.3.2 Sanctions

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

3.3.4 Incident Response Plan

All actors involved in the data value chain may be prone to security incidents, regardless of whether they derive from malicious attacks or inadvertent leaks.

Such security incidents should be anticipated and prevented, or at least mitigated, before any severe harm is done. An internal incident response policy should therefore be considered. Although no incident response policy is fool proof, adopting one may contribute to the development of the right reflexes in the response to security incidents. It shall also be noted that an incident response plan cannot be rigid, but needs to be (re-)assessed on a regular basis and adapted to the changing technological circumstances as well as the (legal) environment in which it is implemented.

3.3.5 Summary

In recent years the EU has made significant progress in terms of cyber security 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 Cyber Security Strategy and the NISD.

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 guiding documents published at EU and national levels. Accordingly, the various actors of the data value chain need to implement measures, procedures and policies to abide by the strict notification requirements and be prepared to provide the necessary information to the authorities, 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.

The opportunities and challenges in relation to breach-related obligations in the context of big data in the transport sector are covered in the previous Section dedicated to (cyber-)security.

[261] NIS Directive, art 16(10) *juncto* art 1(6)

3.4 *Anonymisation and Pseudonymisation*

As mentioned in sub-Section 3.1.5.3 above, an example of privacy-enhancing technologies ("**PET**") that are already being widely used are anonymisation and pseudonymisation techniques.

3.4.1 Introduction to the Key Concepts of Anonymisation and Pseudonymisation

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.[262] The Oxford English Dictionary defines it as the act of removing identifying particulars or details for statistical or other purposes. Common ways to achieve anonymisation are deletion or omission of 'identifying details', or aggregation of information.[263]

In its Opinion 05/2014 on Anonymisation Techniques, the Article 29 Working Party (the predecessor of the European Data Protection Board) discusses several anonymisation techniques, displayed in Table 14 above.[264]

[262] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701, 1707
[263] Kuan Hon and others, 'The Problem of 'Personal Data' in Cloud Computing - What Information is Regulated? The Cloud of Unknowing, Part 1' (2011) 1(4) IDPL 211, 214
[264] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 11-20

Randomisation	A family of anonymisation techniques that alter the veracity of the data in order to remove the strong link between the data and the individual.
Noise addition	A randomisation technique that consists of modifying attributes in the dataset such that they are less accurate whilst retaining the overall distribution.
Permutation	A randomisation technique that consists of shuffling the values of attributes in a table so that some of them are artificially linked to different data subjects.
Differential privacy	A randomisation technique that intervenes when generating anonymised views of a dataset whilst retaining a copy of the original data and that indicates to the data controller how much noise he needs to add, and in which form, to get the necessary privacy guarantees.
Generalisation	A family of anonymisation techniques that generalise, or dilute, the attributes of data subjects by modifying the respective scale or order of magnitude.
Aggregation / K-anonymity	Generalisation techniques that consist of grouping data subjects with, at least, *k* other individuals by generalising the attribute values to an extent such that each individual shares the same value.
L-diversity	A generalisation technique that extends k-anonymity by making sure that every attribute in each equivalence class has at least *l* different values.
T-closeness	A generalisation technique that refines l-diversity by creating equivalent classes that resemble the initial distribution of attributes in the table.

Table 14: Anonymisation techniques

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.[265]

[265] GDPR, art 4(5)

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:[266]

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 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.

Table 15: Pseudonymisation techniques

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.

The present Section will therefore discuss, on the one hand, the impact of anonymisation and pseudonymisation in a personal data protection context, and on the other hand, the use of anonymisation and pseudonymisation techniques as a way to protect non-personal data. It shall be noted that a discrepancy may exist between the legal and technical definitions of certain anonymisation and pseudonymisation techniques discussed in this Section. For the purpose of our legal analysis, this Section will rely on the legal definitions.

In short, the main concepts discussed above shall be used throughout this Section with the following meanings:

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

Anonymisation	Pseudonymisation	Encryption
A technique of processing personal data to reduce the likelihood of identifiability of individuals.	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.	A technique whereby plain text is changed into unintelligible code.

Table 16: Key legal notions of anonymisation, pseudonymisation and encryption

3.4.2 Anonymisation and pseudonymisation of personal data

The present Section aims to provide an overview of the legal consequences of anonymisation and pseudonymisation in a personal data protection context.

Despite the relatively recent attention for the legal issues related to anonymisation and pseudonymisation, the EU Data Protection Directive (95/46/EC) already addressed the subject of anonymisation in 1995, putting forth the following rationale under Recital 26:[267]

- The principles of data protection must apply to any information concerning an identified or identifiable person;
- To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person;
- The principles of protection shall not apply to data rendered *anonymous* in such a way that the data subject is no longer identifiable.

By its very nature, anonymisation and pseudonymisation (including encryption) 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.[268] Pseudonymisation and encryption, however, are 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. Such functions are discussed further in the sub-Sections below.

3.4.2.1 Anonymisation and pseudonymisation as a processing subject to data protection law

The Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques emphasises that, in case of such manipulation of personal data, "*anonymisation constitutes a further processing*

[267] Julien Debussche and Benoit Van Asbroeck, 'Cloud Computing and Privacy Series: a Legal Perspective on Data Anonymisation (part 4 of 6)' (2015) 20(2) Cyberspace Lawyer 7

[268] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data' (2007) WP 136, 29

of personal data."[269] The same reasoning can be applied to pseudonymisation, which is apparent from the definition of pseudonymisation included in the GDPR.[270]

This entails that, when applying an anonymisation or pseudonymisation technique to personal data, one must comply with the principle of purpose limitation, and notably with the requirement of compatibility with the purpose for which the data were initially collected (see also sub-Section 3.1.3.2 of this Deliverable). [271] 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.[272]

Such strict application is criticisable as it may discourage data controllers from applying such techniques in the first place. Furthermore, as will be demonstrated in sub-Section 3.4.2.4 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.

3.4.2.2 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.[273]

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.[274]

The present sub-Section 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.

The Article 29 Working Party approach

In the context of the Data Protection Directive, the Article 29 Working Party highlighted in its Opinion on Anonymisation Techniques 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

[269] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 3
[270] GDPR, art. 4(5): 'Pseudonymisation' means the processing of personal data […].
[271] 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 17 October 2018
[272] 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
[273] GDPR, Recital 26
[274] GDPR, Recital 26

personal data.[275] In the opinion of the Working Party, this would require anonymisation as permanent as erasure, i.e. irreversible anonymisation.[276] 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.[277]

An important factor that needs to be considered in this respect, is whether there is any kind of data in the hands of the controller or any other person that could be used to identify the individual. According to the Working Party, if a data controller keeps the original (identifiable) data, and hands over part of this dataset to another party by removing or masking the identifiable data; the resulting dataset will still constitute personal data because of the data in the hands of the controller that could be used to link back to the individual.[278]

The Working Party examines, in the third and substantial section of Opinion 05/2014, various anonymisation practices and techniques, including pseudonymisation, elaborating on the robustness of each technique based on three cumulative questions:[279]

- Is it still possible to single out an individual?
- Is it still possible to link records relating to an individual?
- Can information be inferred concerning an individual?

Using these three questions, the Working Party produced the table below that shows the strengths and weaknesses of different anonymisation techniques.[280]

[275] 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).
[276] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 6
[277] Ibid 9; see also Commission de la protection de la vie privée, 'Big Data Rapport' (CPVP 2017) 20 <https://www.gegevensbeschermingsautoriteit.be/sites/privacycommission/files/documents/Big_Data_Rapport_2017.pdf> accessed 16 October 2018 on the concept of "singling out".
[278] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 9
[279] Julien Debussche and Benoit Van Asbroeck, 'Cloud Computing and Privacy Series: a Legal Perspective on Data Anonymisation (part 4 of 6)' (2015) 20(2) Cyberspace Lawyer 7
[280] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 24

Is still a risk:	Singling out	Linkability	Inference
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation / K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing / Tokenization	Yes	Yes	May not

Table 17: Strengths and weaknesses of anonymisation techniques

According to the Working Party, some techniques show inherent limitations and each technique examined fails to meet with certainty the criteria of effective anonymisation in light of the three questions above. 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.[281]

As shown in Table 17 below and as recognised in the GDPR definition of the concept, pseudonymisation in itself does not fulfil the Article 29 Working Party's criteria of effective anonymisation, given that pseudonymised data still enables the singling out and linking of a data subject across different data sets. In a big data context, this is not necessarily a bad thing given that some level of identifiability may be needed, notably to achieve predictability in the analytics. It does imply however that pseudonymised data remains subject to data protection rules.[282]

[281] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 24

[282] Ibid 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.

Anonymisation and pseudonymisation in the transport sector – Example 1

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. Furthermore, Paul Francis, a research scientist at the Max Planck Institute for Software Systems, demonstrated remaining data utility when anonymising data of the NYC taxi dataset according to the *Diffix method*, i.e. a method whereby SQL queries made to the database are intercepted and the answers are anonymised by placing limitation on the SQL and adding noise.[283]

Encryption is subject to the same reasoning. The Article 29 Working Party has held in this respect that "*neither encryption nor key-coding per se lends itself to the goal of making a data subject unidentifiable.*"[284] Although the Article 29 Working Party acknowledges that state-of-the-art encryption may lead to a more robust protection of personal data, it stresses the remaining possibility to identify the data subject.

Critique

It shall be noted that 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.[285] 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.[286] 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, instead of taking a strict approach, one could

[283] Estelle Hary, '[Cabanon] Can Anonymised Data still Be Useful?' (*LINC*, 14 November 2017) <https://linc.cnil.fr/fr/cabanon-can-anonymised-data-still-be-useful> accessed 17 October 2018; Laboratoire d'Innovation Numérique de la CNIL, 'CabAnon: Exploring and Visualizing Anonymized Datasets' (*LINC*, 22 February 2017) <https://linc.cnil.fr/fr/cabanon-exploring-and-visualizing-anonymized-datasets> accessed 17 October 2018; Félicien Vallet, '[CabAnon] Anonymity vs Usability, Another Shot at Anonymizing the NYC Taxi Dataset' (*LINC*, 24 September 2018) <https://linc.cnil.fr/fr/cabanon-anonymity-vs-usability-another-shot-anonymizing-nyc-taxi-dataset> accessed 17 October 2018

[284] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 29

[285] 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

[286] Ibid

also 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.[287] Accordingly, if it is not reasonably possible – given the time, resources, technology, and labour required – to associate the data to a particular individual, then the data would remain non-personal if one were to follow such legal reasoning. This would allow recipients to process anonymised data without the need to comply with data protection laws.

The Court of Justice's approach

A judgment from the Court of Justice of the European Union ("**CJEU**") of 19 October 2016 in the *Breyer* case, though still rendered under the Data Protection Directive, might indicate another, 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*."[288] (Emphasis added).

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.

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.[289] The Article 29 Working Party did however indicate that different techniques could be combined to obtain the required level of 'anonymisation', in which case the data protection rules would not apply.[290] Furthermore, pursuant to the GDPR, associations and other bodies representing categories of data controllers or processors may prepare codes of conduct regarding the pseudonymisation of personal data.[291] We believe such codes of conduct are indispensable to the uptake of pseudonymisation techniques in a big data context, including in the transport sector.

Any company should therefore, when contemplating to apply anonymisation techniques, adopt a risk-based approach, in line with the CJEU's *Breyer* judgment, whereby it considers the possible re-identification risks that may remain after application of the technique. Such risks (if any) should be the subject of a balancing exercise to determine whether data protection law is applicable (see Figure 7 above). An occurrence of zero risk will be rather unlikely, especially in view of emerging technologies (including big data), which guarantee

[287] GDPR, Recital 26
[288] Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:339, paras 45-46
[289] 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
[290] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' WP 216, 23-24
[291] GDPR, art 40(2)(d)

with less and less certainty the existence of anonymisation techniques that would allow falling outside of the scope of data protection legislation. In this respect, it shall be noted that the definition of personal data itself is constantly evolving.[292] Certain types of information (e.g. dynamic 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.

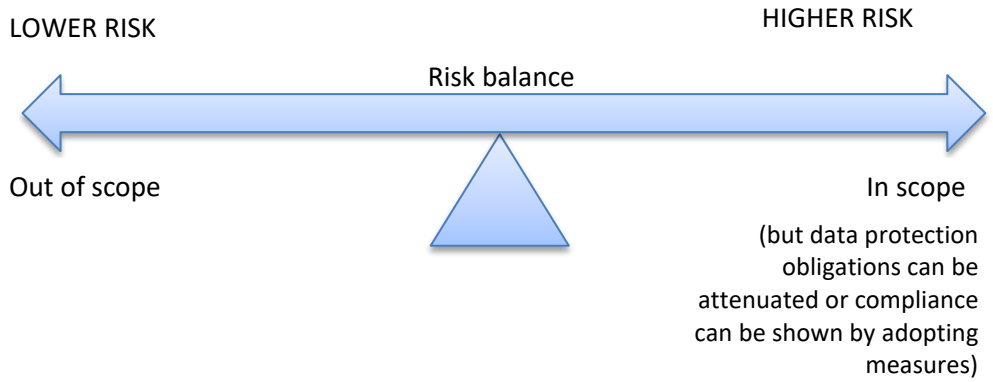


Figure 7: Risk-based approach

By way of illustration, the Belgian DPA recommends in its Big Data Report to aggregate data processed in the framework of a big data project, in combination with a frequently repeated quantitative assessment that calculates the probability of singling out or re-identification from the aggregated datasets, such as "Small Cells Risk Analysis" (hereinafter referred to as "SCRA").[293] SCRA is a theoretical analysis of the number of unique combinations of the reported values of quasi-identifiers[294] relative to the number of points in the data at personal level.[295] Groups in the aggregated data with too few points at personal level are called "small cells", which engender a risk of indirect identification. In such event, it may be advisable to apply a number of restrictions, such as deleting one or more quasi-identifiers or aggregating a quasi-identifier.[296] Only if the SCRA indicates that the possibility of "singling out" is sufficiently excluded, the data are adequately aggregated or anonymised.[297]

[292] Václav Janecek, 'Ownership of Personal Data in the Internet of Things' (2018) 34(5) Computer Law & Security Review 1039
[293] Commission de la protection de la vie privée, 'Big Data Rapport' (CPVP 2017) 22 <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 16 October 2018
[294] Ibid: Quasi-identifiers are attributes that are used to delineate the groups in aggregated data, but which – combined – could also be used to identify an individual
[295] Ibid
[296] Ibid
[297] Ibid 23

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

Anonymisation and pseudonymisation may serve as mechanisms to release data controllers or processors from certain specific data protection obligations related to personal data breach. That way, while the general data protection rules remain applicable, data controllers or processors may escape the application as such of the specific personal data breach-related obligations (see also Section 3.3 below).

Notification of a personal data breach to the supervisory authority

As further assessed in Section 3.3 of this Deliverable, the GDPR requires the notification 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*."[298]

Such requirement however does not apply when the data controller is able to demonstrate, in accordance with the accountability principle, that the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons.[299] 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.

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.[300] In 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.[301]

Communication of a personal data breach to the data subject

As mentioned in Section 3.3, when the breach is likely to result in a high risk to the rights and freedoms of natural persons, the data controller must communicate the personal data breach to the natural persons concerned.[302]

Such communication shall however 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.[303] The GDPR mentions in particular "*those [measures] that render the personal data unintelligible to any person who is not authorised to access it, such as encryption*."

---

[298] GDPR, arts 4(12) and 33
[299] GDPR, art 33(1) and Recital 85
[300] Article 29 Data Protection Working Party, 'Opinion 03/2014 on Personal data breach notification' (2014) WP 213, 3; Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679' (2018) WP250rev.01, 25
[301] Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679' (2018) WP250rev.01, 25
[302] GDPR, art 34(1)
[303] GDPR, art 34(3)(a)

Anonymisation and pseudonymisation techniques, and notably encryption, may therefore result in an exemption from the specific obligation to communicate a personal data breach to the data subject. The observations made under the previous title with respect to the Article 29 Working Party guidance on personal data breach further support this conclusion.

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

Lastly, anonymisation and pseudonymisation may constitute a means to comply with certain data protection rules. Thus, even when the application of data protection law cannot be bypassed, pseudonymisation 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.*"

We discuss below the several ways in which the anonymisation or pseudonymisation of personal data could contribute in complying with data protection law.

Accountability principle: data protection by design and data protection by default

As discussed in sub-Section 3.1.3.6 below, Recital 74 of the GDPR requires the data controller to implement appropriate and effective measures and to be able to demonstrate the compliance of processing activities with the GDPR, including the effectiveness of the measures.

Such measures refer notably to the principles of data protection by design and data protection by default, according to which data controllers are under a specific obligation to consider data protection issues throughout the entire lifecycle of a project or system, including as from the outset of the design stage.[304]

More specifically, having regard to the state of the art and costs of implementation when implementing such measures, and taking into account the nature, scope, context and purposes of the processing, 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.

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.

Security of processing

Controllers (and processors) are required to "*implement appropriate technical and organisational measures*".[305] Such measures shall take into account several factors such as (i)

[304] GDPR, art 25
[305] GDPR, art 32

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*". 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.[306]

Purpose limitation (further processing of personal data)

Anonymisation and pseudonymisation may be a means to comply with the purpose limitation principle enshrined in Article 5(1)(b) of the GDPR (see sub-Section 3.1.3.2 below). According to that principle, 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 the personal data were initially collected, the data controller shall take into account the existence of appropriate safeguards, including pseudonymisation and encryption. [307] The GDPR therefore explicitly acknowledges the important role pseudonymisation and encryption may play in complying with the purpose limitation principle.

Storage limitation

Anonymisation could also be used to comply with Article 5(1)(e) of the GDPR, which lays down the storage limitation principle.[308] Such principle 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.

In this context, anonymisation might constitute a compulsory processing activity that enables one to comply with its data protection obligations.[309]

[306] 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)
[307] GDPR, art 6(4)(e)
[308] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP 216, 7
[309] Khaled El Emam and Cecilia Alvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymisation Techniques' (2015) 5(1) IDPL 73

3.4.2.5 *Examples of anonymisation and pseudonymisation of personal data in the transport sector*

Anonymisation and pseudonymisation in the transport sector – Example 2

In its Code of Practice on Anonymisation[310], the 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 (containing mobile phone number, approximate location, and date and time) 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. However, in order to ensure that related records can still be linked for the purposes of the research, the same real phone number should always be replaced by the same dummy value. 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. Randomisation without guaranteeing uniqueness may also be a possible solution.

[310] 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 17 October 2018

Anonymisation and pseudonymisation in the transport sector – Example 3

The ICO also examined in its Code of Practice on Anonymisation a case study on analysing passengers' journey times.[311]

In such hypothesis, a public transport company would use its Go-Card data to carry out a study showing the amount of time commuters of particular age groups take to make various journeys. The study results would be used to improve accessibility planning. Inevitably, such study would involve the processing of personal data, such as the passengers' Go-Card numbers and their dates of birth. The ICO suggests techniques such as pseudonymisation, hashing, and data banding[312] to anonymise the data. The results of such manipulation are shown in Figure 8 above.

Go-card no.	Passenger DoB	Start point	End point	Journey time
WT98765G	01/09/1973	Brooks End	Tree Street	17m 45s
WT45678B	18/09/1933	Brooks End	Tree Street	15m 05s

and this:

Hashed* passenger ref. no.	Age band	Start Point	End Point	Journey time
14793X...	35 - 45	Brooks End	Tree Street	18m
23955P...	75 – 80	Brooks End	Tree Street	15m

* a keyed cryptographic hash function such as SHA356

Figure 8: ICO Anonymisation Code of Practice – Case study 3

[311] Ibid 69

[312] The ICO defines data banding as "*a technique to produce coarser-grained descriptions of values than in the source dataset*". The result of such technique is derived data, i.e. a set of values that reflect the character of the source data but which hide the exact original values (Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (ICO 2012) 53 <https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 17 October 2018)

Anonymisation and pseudonymisation in the transport sector – Example 4

Intelligent Transportation Systems ("**ITS**") aim to increase the efficiency of both personal and commercial travel. An illustration of ITS already being deployed is that of parking information services, whereby an application or system provides drivers updates on the location of available parking spaces through sensors embedded in parking spaces connected to GPS-enabled devices. Such service may be coupled with the possibility to reserve a spot.[313] Personal data may be collected from the drivers accessing the system and/or making a reservation. In such context, a risk of individual location data processing exists. A potential countermeasure would be to provide updates on free spots in a given street segment. This would achieve k-anonymity of monitored vehicles/users, with k being the number of vehicles parked on the same segment.[314]

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

The previous sub-Section examined how anonymisation and pseudonymisation techniques may be used to protect personal data. Nevertheless, certain non-personal information, such as trade secrets or confidential information, may constitute an essential asset of a company and therefore merits protection throughout its lifecycle in the big data analytics process, both on a technical and legal level.

In order to protect such non-personal data on a legal level, inspiration can be drawn from what is known about the impact of anonymisation techniques on personal data in a data protection law context, as discussed in sub-Section 3.4.2 above.

The results of our research and, in particular, the conclusions above on anonymisation techniques in a data protection context have led us to propose an innovative way to extrapolate to some extent such conclusions to other fields of law, such as trade secrets, which may be essential assets of the actors involved in the big data lifecycle.

3.4.3.1 Trade secrets

Any company has an interest in safeguarding its trade secrets (if any), given that those are usually part of the foundations the whole business rests on. Moreover, a specific legal ground for the protection of trade secrets now exists at EU level, namely the EU Trade Secrets Directive, which entered into force on 5 July 2016.[315] The EU Member States had to transpose the Directive into their national laws by 9 June 2018.

Like Article 39 of the TRIPs Agreement, the EU Trade Secrets Directive defines a 'trade secret' as information which meets all of the following requirements:

[313] Jaimee Lederman, Bran D. Taylor and Mark Garett, 'A Private Matter: The Implications of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

[314] Juan Hernandez-Serrano and others, 'On the Road to Secure and Privacy-Preserving IoT Ecosystems' in Ivana Podnar Žarko, Arne Broering, Sergios Soursos and Martin Serrano (eds), *Interoperability and Open-Source Solutions for the Internet of Things* (InterOSS-IoT 2016, Lecture Notes in Computer Science, volume 10218, Springer 2017)

[315] 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

- 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;
- it has commercial value because it is secret; and
- it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.[316]

It cannot be excluded that certain stakeholders participating in big data analytics, including in the transport sector, engage in the disclosure of their trade secrets. Such information should therefore be duly protected.

3.4.3.2 Confidential information

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. It is therefore necessary to obtain assurance from the provider of the big data analytics solution that it will keep confidential information confidential.

Confidentiality can be (at least partially) achieved contractually, but should also be aimed to be obtained on a technical level in order to avoid that data the publication of which would potentially be harmful to the commercial interest of any party concerned, is made publicly available.

3.4.3.3 Anonymisation of trade secrets and confidential information

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. Indeed, due to the lack of a legal framework for the ownership of data (as discussed in Section 3.10 of this Deliverable), the issue is currently regulated (insufficiently) through contracts. However, 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. 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

[316] Trade Secrets Directive, art 2(1)

legal perspective[317], 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.

Table 18 above aims to provide an overview of the benefits of using a technical solution, such as anonymisation, for the protection of non-personal information like trade secrets and confidential information.

Advantages of anonymisation in protecting non-personal information
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 (see definition above) 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*[318] in protecting one's own or another's assets.
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.
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.
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.[319] 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.

Table 18: Advantages of the use of anonymisation to protect non-personal information

[317] But also from a technical perspective.
[318] 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").
[319] Directive 2009/24/EC of the European Parliament and of the Council on the legal protection of computer programs [2009] OJ L 111/16

It cannot be excluded that the participation of companies in big data analytics involves the sharing of trade secrets or confidential information with the other stakeholders involved in the process.

Taking into account the potential commercial value of such non-personal information, a comprehensive protection should be put in place. This entails that, aside from practical (disclosure on a strict need-to-know basis) and contractual (non-disclosure agreements) protection mechanisms, the information should be protected in a technical manner.

From a legal perspective, anonymisation and pseudonymisation techniques may prove to be good instruments thereto, as their legal significance has already been acknowledged in the context of data protection legislation. 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.

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.

3.4.3.4 Example of Anonymisation and pseudonymisation of non-personal data in the transport sector

Anonymisation and pseudonymisation in the transport sector – Example 5
In their paper on Anonymization of Data from Field Operational Tests, 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 ITS, including 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 anonymisation technique, preserving however research essential information, would facilitate the access to and re-use of valuable data.

3.4.4 Summary

As demonstrated by this Section, 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 Directive[320] as "*one of the principles of enhancing*

[320] 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

individuals' privacy". In addition, this Section 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.

Opportunities in relation to anonymisation / pseudonymisation in the context of big data in the transport sector	Challenges in relation to anonymisation / pseudonymisation in the context of big data in the transport sector
Irreversibly anonymised personal data may be processed without the need to comply with data protection laws.	As the definition of personal data is constantly evolving, anonymisation techniques should also continuously evolve and become increasingly robust in order to achieve irreversible anonymisation (where desired).
The application of anonymisation techniques may engender an exemption from the notification obligations related to personal data breach.	The GDPR will remain applicable if, in spite of the anonymisation techniques used, the data subject can still be identified. In such event, all data protection principles and obligations must be respected by the data controller as well as the data processor when processing the personal data.
Anonymisation techniques may serve as a means to comply with data protection law, and specifically with the following obligations: (i) data protection by design and by default; (ii) security of processing; (iii) purpose limitation; and (iv) storage limitation.	A too far-reaching anonymisation of data may limit predictability in the big data analytics.
Anonymisation techniques may prove to be apt instruments to protect non-personal information in a technical manner. If successful, this may encourage stakeholders involved in the big data value cycle to engage in data sharing.	

Table 19: Summary table of opportunities and challenges in relation to anonymisation / pseudonymisation in the context of big data in the transport sector

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, such as the ICO Anonymisation Code of Practice[321], is indispensable to support data controllers and processors 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 Section. 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.

[321] Which however dates from before adoption of the GDPR (Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (ICO 2012) 69 <https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 17 October 2018)

3.5 Supply of digital content and services (personal data as counter performance)

3.5.1 Introduction to personal data as counter-performance

Digital content, in short, means data which are produced and supplied in digital form. Forms of digital content may include computer programs, applications, games, music, videos or texts, cloud storage and potentially social media. In particular, the transport sector accounted for 5% of companies providing digital content in 2016, representing in this way a significant part of the sectoral distribution.[322]

Free digital content represents an extremely important component of digital content. The latter is particularly popular with consumers who have shown a strong appetite for free digital content: only a small minority of consumers pay for digital content on regular basis. Thanks to this free model, many more consumers access digital content than if they had to pay for it. Companies are henceforth able to reach a larger pool of consumers. It further allows them testing quickly new ideas and offering more innovative services. Finally, the free model plays an important role for small companies in the digital market as it reduces the barriers to entry and provides them with cost-saving opportunities such as marketing and advertising.

Digital companies foster the common perception that this digital content is provided for free; in reality it requires users to surrender valuable personal information in exchange to enjoy them. Since the Cambridge Analytica Files in March 2018, personal data for "free" digital content in addition to being an emerging reality, has gained public visibility. The extent to which personal data can be monetized by companies gives currently rise to intense discussions.

The present Section seeks to appreciate whether or not it is appropriate to legislate, and hereby legalise, personal data as an economic asset in the supply of digital content. More particularly, this Section will rely on the 2015 proposal of the European Commission for a Directive on contracts for supply of digital content[323] (hereafter the "Proposal")[324]. The scope of such Proposal is to "*apply to any contract where the supplier supplies digital content to the 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*."[325]

3.5.2 Economic dimension

The present sub-Section aims to provide an overview of the strictly economic dimension of the monetary value of personal data.

---

[322] 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.org/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf> accessed 17 October 2018

[323] 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

[324] It shall be noted that along the lengthy legislative process, the text of the Proposal is subject to, possibly substantial, modifications.

[325] COM (2015) 634 final, art 3(1)

3.5.2.1 Business model of free digital content

Personal data of individuals representing monetary value is already a reality in the digital market. Indeed, as affirmed by the European Commission, business models based on the monetisation of data become predominant.[326] Two types of business models trading (personal) data – categorised according to their uses – can be distinguished:

- The "free" provision of online content;
- The "free" provision of online services.

Supply of digital content and services in the transport sector – Example 1
A relevant example of "free" online services are "free" Wi-Fi services in airport or train station, where users need to accept cookies and provide their email address if they want to navigate on the Internet. That is to say, if a user wants a free provision of Internet data, they must disclose to the suppliers – and often shared or sold to third-partners – mailing address, location data, a chronology of websites visited, etc.

In those business models, personal data operates as a currency, and generally as the sole currency, in the exchange of digital content/services in the digital market.

3.5.2.2 Quantifying personal data

Unlike money, data does not have a standardised value. On the contrary, data is characterized by its fluidity and intangibility.[327] Moreover data is risk- and context-dependent. In short, data is a dynamic product. Can personal data consequently be treated as a form of payment for digital content?

As argued by some commentators, attaching a monetary value to personal data is not impossible. Proof of this can be found in the different existing initiatives allowing the monetization of individual's data, such as CitizenMe or Brave (see Chapter 7 of the LeMO Report on Economic and Political issues D.2.1). 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").[328]

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 reasons, 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

[326] Ibid
[327] Rebecca Kelly and Gerald Swaby, 'Consumer Protection Rights and "Free" Digital Content' (2017) 23(7) Computer and Telecommunications Law Review 165, 168
[328] 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

sold several times. By giving its 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 either 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.[329] Some of them are based on market-evaluation whereas some are based on individual valuation. The summary of measures of value of personal data is displayed in the Table 20 below.

Indicator	Description
Financial results per data record	Aggregated market cap (i.e. revenue or net income) of a company divided by the total number of personal data records used by the company.
Market prices for data	Price per personal data entry offered on the market by data brokers.
Cost of data breach	Economic cost of a data breach (for firms and individuals) per data entry.
Data prices in illegal markets	Estimation of prices of personal data (per data entry) in illegal markets.
Surveys and economic experiments	Valuation of personal data in monetary terms is reported/revealed by individuals in surveys/economic experiments.
Data on willingness of users to protect their data	Amounts that individuals are ready to spend to protect their data.

Table 20: Key measures to valuate personal data

[329] 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=9725A618211DF41C00207963B84C43F0> accessed 17 October 2018

All of these elements are incomplete: while individual-based methodologies are not incentive compatible, other priced-in factors and externalities are ruled out.[330] Although presenting some drawbacks, the wide range of methods briefly mentioned in this Section indicates at least that monetary value of personal can be quantified.

3.5.3 Educational dimension

The previous sub-Section examined from an economic perspective how personal data can be considered as a monetary value. Beyond this economic dimension, the legislator's approach also entails a so-called "educational" dimension.

In practice, individuals attempt to avoid the disclosure of their personal data. A Eurobarometer research found indeed that 89% of the people agreed that they avoid disclosing their personal data online.[331] While avoiding such disclosure, individuals are however usually not fully aware of the value of their personal data and cannot evaluate the value that will be created with their data.

The recognition of personal data as a legal and legitimate counter-performance would contribute to the protection and the empowerment of individuals, i.e. the "educational" dimension of personal data as a form of payment for digital content.

In this respect, some commentators suggest that the real problem is that "irresponsible" individuals do not pay enough attention to what they do (or agree to let to do) with their personal data, and that this problem could be resolved by making them understand that the data may be worth money.[332] So it would be a form of accountability through commercialisation and commodification. They further state that this argument is at the very least twisted, if not absurd: as if economic value was a source of responsibility. This problematic is part of the more general debate on how far the legal system should “protect consumers/users from themselves” without risking becoming overly paternalistic.

In line with this, some scholars raise the question whether a right to know the value of their personal data should be granted to consumers/users in order to increase their awareness on their own personal data as well as their power in the digital market.[333] Quantification of personal data is nonetheless a prerequisite for such a right. As previously demonstrated in sub-Section 3.5.2 below, different methods for quantifying the value of personal data indicates that the monetary value of personal data can be quantified. Nonetheless a set of practical problems arises from this proposed right regarding, among others, the monitoring and the enforcement.

---

[330] Nicola Jentzsch, 'State-of-the-Art of the Economics of Cyber-Security and Privacy, IPACSCO – Innovation Framework for ICT Security Deliverable, No. 4.1' (Waterford Institute of Technology 2016) para 3.8.1 <https://www.econstor.eu/bitstream/10419/126223/1/Jentzsch_2016_State-Art-Economics.pdf> accessed 17 October 2018

[331] TNS Opinion & Social, 'Special Eurobarometer 423 – Cyber Security Report' (European Commission 2015) <http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_423_en.pdf> accessed 17 October 2018

[332] Serge Gutwirth and Gloria Gonzales Fusters, 'L'éternel retour de la propriété des données : de l'insistance d'un mot d'ordre' in Elise Degrave, Cécile de Terwangne, Séverine Dusollier and Robert Queck (eds), *Law, Norms and Freedoms in Cyberspace / Droit, normes et libertés dans le cybermonde (*Larcier 2018) 138

[333] 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

More significantly, there already exist under EU data protection law several rights recognised to data subjects (see sub-Section 3.1.6 below), notably requiring to inform users when their personal data is processed, including when processioned as "counter-performance other than money". It is therefore difficult to see the benefit of introducing a right to know the value of one's personal data.

Finally, another argument against the so-called educational advantage of the Proposal is that, in most cases, it is not the personal data in raw form that is at the heart of the transaction, but their time, their attention, their life. Digital businesses have indeed a very instrumental interest in the personal data of consumers/users, which represents for them, in the first place and almost systematically, not an end in itself, but above all an open door to the possibility of using their personal data for marketing purposes, itself a source of profits.[334]

3.5.4 The desirability of legislating

Digital content supplied against personal data provided by consumers as a counter-performance for the supply of digital content is currently considered in the Proposal. It is the subject of intense discussions and remains a long way from being implemented into the national law of Member States. This clearly demonstrates that data as counter-performance can either be seen as a step back or as new way forward, including for big data in the transport sector.

With this proposal the European Commission decides to solve the problem by legalising the practice of individuals "paying" for "free" digital content with their personal data through a contract. In other words, it seeks to reconcile the legal reality with the requirements of the economic reality by means of consumer and contractual law.

In a nutshell, the Proposal provides the same right for the data subject providing his/her personal data to the supplier as a paying customer to gain access to digital content or services. Digital content contract rights and remedies are thus extended to data-driven transactions.

3.5.4.1 Practical challenges

Apart from extending the scope to digital content supplied against a counter-performance other than money (i.e. personal data), the Proposal does not foresee specific rules explaining its practical application.

Firstly, treating data as counter-performance presents practical challenges, including the fact that companies do not always directly monetize data, as discussed above in sub-Section 3.5.2 below. Data is indeed often used for a wide range of commercial purposes which implicate *indirect* rather than direct monetization, such as improving security or improving consumer's experience. In other words, some investment may sometimes need to be made to generate value from the data. The idea of data as a counter-performance therefore designates a catch-

[334] Serge Gutwirth and Gloria Gonzales Fusters, 'L'éternel retour de la propriété des données : de l'insistance d'un mot d'ordre' in Elise Degrave, Cécile de Terwangne, Séverine Dusollier and Robert Queck (eds), *Law, Norms and Freedoms in Cyberspace / Droit, normes et libertés dans le cybermonde (*Larcier 2018) 138-139

all term and does not address the variety and the specificity of data uses, business models as well as the relationship between the users and suppliers. The question remains of when the user should be regarded as having given counter-performance by providing data.

Returning data to consumers also presents challenges for big data. As if isolation of data was not enough, anonymisation and pseudonymisation (see Section 3.4 below) could further make it impossible to return the data to the user without sometimes collecting additional data than is currently collected. Some speak of inconsistency in the principles of data retrieval and data anonymisation.[335]

Another major concern includes the potential impact of retrieving data and returning it to the user on other user's experience since the data provided or generated by those accessing the digital content enable the product to function. In some cases, this could go as far as making some current services in the transport sector inoperable.

Supply of digital content and services in the transport sector – Example 2
A carpooling service, such as BlaBlaCar, may have to delete reviews users have uploaded to comply with the data retrieval obligation. Returning data would negatively alter the experience of all other consumers of the service by affecting the featuring reviews and star ratings of drivers.

Supply of digital content and services in the transport sector – Example 3
Automatic Vehicle location enables to acquire the precise location of public transportation in real time via passenger generated vehicle location sent from smartphones. Passenger generated vehicle location track the route buses that the passenger are riding with precise location in order to support public transportation users. Retrieving data may affect the frequency and punctuality of public transportation.

Data as counter-performance can therefore be seen as a step back for big data in the transport sector if *ex ante* guidelines are not released on when data should be regarded as counter-performance and how to technically enable individuals to retrieve data provided in event of termination.

3.5.4.2 Legal challenges

The Proposal is not exempt of criticism from a legal perspective. Indeed, some commentators have examined the difficulties that the proposed text is likely to pose. Some of the critiques presented by Prof. Dr. Axel Metzger can be summarised as follows[336]:

[335] 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.org/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf> accessed 17 October 2018

[336] Axel Metzger, 'Data as Counter-Performance: What Rights and Duties do Parties Have?' (2017) 8(1) JIPITEC 2 <https://www.jipitec.eu/issues/jipitec-8-1-2017/4528> accessed 23 October 2018

- By acknowledging personal data as counter-performance, the proposal merely codifies a social practice. The legal recognition of a common social practice is likely to have legal consequences for both parties to the contract.
- The combination of European law for the rights of one party and national law for the rights of the other party raises a number of fundamental challenges, especially in light of the full harmonisation approach of the Proposal and the principle of effectiveness of European law.
- Accepting personal data as counter-performance in bilateral contracts intensifies the rights and duties of both parties. For the consumer, the Proposal makes clear that the data subject providing his/her personal data to the supplier shall have the same rights as in the case of a money consideration paid to the supplier. However, the Proposal says nothing about the duties of the consumer and the rights of the supplier?
- The service provider should have the right to claim for the counter-performance within the limits of data protection law. It follows that the consumer is under an obligation to submit his 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 privacy and personal data protection legislation, and in particular the GDPR.
- Whether the Proposal will finally improve the legal situation of consumers on the digital markets will also depend on the protection given to the supplier on the national level. On the one hand, it will hardly be acceptable to give full protection to the consumer "paying with its personal data" without looking at the same time at the suppliers rights in such contract settings. On the other hand, the rights of the supplier in application of the national contract law may also not undermine the legislative purpose of the Proposal.

In addition, the Proposal 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 remain in the realm of autonomous national contract law. Finally, The Proposal is mainly focused on the consumer's rights and the supplier's obligations, leaving outside its scope the aspects related to the consumer's duties. In the same vein, while it provides detailed rules for the rights of the consumer to terminate the contract, it remains silent on the termination right of the supplier.

3.5.4.3 Is there a need to monetize data?

With regard to the above, some commentators, such as the European Data Protection Supervisor,[337] have been critical vis-à-vis the introduction of the explicit possibility to use personal data as a counter-performance. They argue that personal data cannot be monetized and the Proposal, covering the field of contract law, is not 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 do

[337] 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

essentially not see the need to attach legal consequences to a practice which may be observed everywhere in the digital environment.

3.5.5 Summary

As demonstrated by this Section, personal data as a form of payment for the supply of digital content is an emerging reality. The Proposal recognises for the first time the provision of data by users as counter-performance. Hereby it provides an indication of the legislature's desire to take into account the underlying economic reality of transactions and admit, once again, its concern for individuals.

Nevertheless, various challenges – summarised in the table below – with regard to big data applications, including in the transport sector, may arise and must be taken into account.

Opportunities in relation to the supply of digital content and services in the context of big data in the transport sector	Challenges in relation to the supply of digital content and services in the context of big data in the transport sector
The legal framework on supply of digital content and services will ensure an adequate level of protection for the consumer.	Data retrieval obligations will increase costs disproportionately and lead to administrative burdens on the industries supplying digital content in the transport sector.
	The obligations concerning data may make some current services inoperable. Some companies may also start to charge for digital content services that are currently free. On a wider scale the ecosystem of innovative services in the field of transport could be jeopardized. Many start-ups and small companies do indeed rely on free digital content for their business model.
	Quality of digital services/content may be affected and hereby the experience of other users.

Table 21: Summary table of opportunities and challenges in relation to the supply of digital content and services in the context of big data in the transport sector

It follows from the foregoing that personal data as counter-performance for the supply of digital content is *per se* not an undesirable concept in the context of big data, including in the transport sector. However, the legalisation of the practice and its inclusion in the Proposal generate various practical concerns around the obligations concerning data and require further clarifications. The subject calls for the establishment of guidelines, or similar initiatives, to assist the suppliers of digital content and provide greater certainty.

3.6 *Free flow of data*

The term "free flow of data" is usually mentioned in the debate on restrictions to the cross-border flow of data, where the "free flow of data" illustrates an ideal scenario in which no (legal) barriers to cross-border data flows remain. Today, that scenario has yet to materialise due to the continued existence of so-called "data localisation requirements".

Data localisation requirements are a worldwide phenomenon and come in many different shapes and forms. They can apply to personal data, non-personal data, or all data no matter their qualification. They range from the Russian law that requires all processing of Russian citizens' personal data to be carried out using servers located in the Russian Federation to the French Ministerial Circular which makes it illegal to use a non-"sovereign" cloud for data produced by public (national and local) administration.[338]

While data localisation requirements vary widely, they have one feature in common: they raise the cost of conducting business across borders.[339] In the EU, over 60 restrictions were identified in 25 jurisdictions.[340] Security, surveillance and economic protectionism, among others, are some of the reasons prompting countries to adopt such restrictions to the free flow of data.[341]

3.6.1 Types of restrictions to the free flow of data

FERRACANE (2017) has identified five types of restrictions, which can be divided into two main categories of, on the one hand, strict restrictions to cross-border data flows and, on the other hand, conditional restrictions to such data flows.

3.6.1.1 Strict restrictions

Three categories of strict restrictions can be distinguished: (i) local storage requirements; (ii) local processing requirements; and (iii) bans on transfers.[342] A local storage requirement entails that data can only be transferred to another country on condition that a copy is also stored locally, within the territory of that country or jurisdiction. A local processing requirement goes one step further, and requires the data not only to be stored, but also to be, for the most part, processed within a given country's territory. A company will in such case be required to use a data centre located in that territory for its main data processing activities. The most far-reaching type of restriction is an outright ban on cross-border data transfers, in which case data must be stored, processed, and also accessed from within the territory concerned. Such restriction is typically imposed for highly sensitive categories of data.[343]

---

[338] Martina F. Ferracane, 'Restrictions on Cross-Border Data Flows: A Taxonomy' (ECIPE Working Paper, No. 1/2017) 14; 23 <http://ecipe.org//app/uploads/2017/11/Restrictions-on-cross-border-data-flows-a-taxonomy-final1.pdf> accessed 17 October 2018

[339] Ibid 2

[340] 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)

[341] Cathal Flynn, 'Shortcomings of the EU Proposal for Free Flow of Data' (2018) 45(4) InterMEDIA 30, 31

[342] Martina F. Ferracane, 'Restrictions on Cross-Border Data Flows: A Taxonomy' (ECIPE Working Paper, No. 1/2017) 3 <http://ecipe.org//app/uploads/2017/11/Restrictions-on-cross-border-data-flows-a-taxonomy-final1.pdf> accessed 17 October 2018

[343] Ibid 4

3.6.1.2 Conditional restrictions

Conditional restrictions on cross-border data flows may be categorised into those where conditions apply to the recipient country and those where they apply to the data controller or the data processor.[344] If the conditions are so strict that they cannot actually be fulfilled, the measure must be reclassified as a ban on cross-border data transfers.

The GDPR has for instance introduced a conditional regime for the processing of personal data outside the EU/EEA by allowing such processing on condition that appropriate safeguards are in place. As mentioned in sub-Section 3.1.7 above, these safeguards can take a variety of forms. Another example is the imposition by the recipient country of an infrastructure requirement, requiring all companies aspiring to operate in that country to build one or more local servers.[345]

3.6.2 Rationale and impact of restrictions

Data localisation requirements are often prompted by legislators' or policy makers' perception that data are more secure when stored within a country's border. That perception is however 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.[346] 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 is no different than a secure server in Belgium. Another important consideration for lawmakers however is the fact that law enforcement is able to access data that are stored domestically much easier than data stored outside the home jurisdiction.[347]

In 2017, the European Commission conducted a public consultation in the framework of its communication "Building a European Data Economy". One of the consultation's objectives was to gather information from stakeholders on whether and how data localisation requirements hinder the free flow of data in the EU. The most cited impact was very clear: "costs".[348] Many answers were received from cloud service providers ("**CSPs**") which are particularly affected by data localisation requirements. The CSPs argued that these restrictions undermine the cloud business model. In some cases, this is accomplished by preventing providers from accessing markets where they do not have a data centre. In other cases, the users themselves are prevented from using cloud services provided from another EU Member State.[349]

Data localisation requirements 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 this

[344] Ibid 3
[345] Ibid 5
[346] Daniel Castro, 'The False Promise of Data Nationalism' (ITIF 2013) <http://www2.itif.org/2013-false-promise-data-nationalism.pdf> accessed 17 October 2018
[347] Sam Pfeifle, 'Is the GDPR a Data Localization Law?' (*IAPP*, 29 September 2017) <https://iapp.org/news/a/is-the-gdpr-a-data-localization-law/> accessed 17 October 2018
[348] European Commission, 'Annex to the Synopsis Report. Detailed Analysis of the Public Online Consultation Results on 'Building a European Data Economy" (European Commission 2017) 3 <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 17 October 2018
[349] Ibid 3-4

constitutes a serious obstacle to growth, to entering new markets, and to the development of new products and services.[350]

3.6.3 Existing legal instruments to assess the validity of data localisation requirements

A number of legal instruments at EU level already prevent data localisation requirements to some extent. This sub-Section offers a brief overview of the main instruments in this respect.

The e-Commerce Directive[351] (2000/31/EC) eliminates restrictions to the freedom to offer information society services to individuals located in another Member State. It prohibits EU Member States to require prior authorisation or conditions having equivalent effect, such as for instance a requirement to build one or more servers in that Member State, before allowing the provision of information society services in that Member State.

The Services Directive [352] (2006/123/EC) deals with authorisation schemes and other requirements regulating access to, or the exercise of, a service activity. It ensures the right of a service provider to offer a service in a Member State other than that in which it is established and also prevents a Member State from introducing measures that prohibit or restrict a recipient in that Member State to use such a service supplied by a provider established in another Member State. The Service Directive's impact on the data economy is however limited, as electronic communications services and networks, and associated facilities and services, are excluded from its scope. Additionally, it should be noted that "services in the field of transport, including port services, falling within the scope of Title V of the Treaty" are also excluded.

Finally, the Transparency Directive[353] (2015/1535) creates a mechanism to prevent Member States from establishing rules on information society services that could create barriers to the free movement of services in the internal market.[354] This includes cloud storage and other cloud processing services.

3.6.4 Proposed Regulation on a framework for the free flow of non-personal data

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[355] (hereinafter referred to as the "**Free Flow**

[350] Commission, 'Building a European Data Economy' (Communication) COM (2017) 9 final, 6-7
[351] 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 [2000] OJ L 178/1
[352] Directive 2006/123/EC of the European Parliament and of the Council on services in the internal market [2006] OJ L 376/36
[353] Directive 2013/50/EU of the European Parliament and of the Council amending Directive 2004/109/EC of the European Parliament and of the Council on the harmonisation of transparency requirements in relation to information about issuers whose securities are admitted to trading on a regulated market, Directive 2003/71/EC of the European Parliament and of the Council on the prospectus to be published when securities are offered to the public or admitted to trading and Commission Directive 2007/14/EC laying down detailed rules for the implementation of certain provisions of Directive 2004/109/EC [2013] OJ L 294/13
[354] Commission, 'Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, accompanying the document Communication Building a European Data Economy' SWD (2017) 2 final, 10
[355] 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

Regulation"). On 19 June 2018, negotiators achieved a political agreement on the content of the future Regulation.[356] On 4 October 2018, the Regulation was adopted by the European Parliament.[357] The Council of the EU will adopt the Regulation shortly before or after the due date of this Deliverable, and the Regulation will enter into force by the end of 2018. This sub-Section examines the main features and challenges of the text as adopted by the European Parliament on 4 October.

3.6.4.1 Main features of the proposed Regulation

The Free Flow Regulation will apply to all processing of electronic data other than personal data within the meaning of the GDPR. The rationale behind this is to complement the GDPR which already makes up the legal framework applicable to personal data and which entered into application on 25 May 2018. The current approach is however subject to criticism, which will be addressed in sub-Section 3.6.4.2 below.

The Regulation's key provision prohibits data localisation requirements in the EU.[358] EU Member States will no longer be allowed to restrict the location of data processing activities to a particular Member State's territory, nor will they be able to achieve the same result by imposing any restrictions on the processing of data in other Member States. Not only hard core localisation requirements will be caught by this prohibition, but also other types of restrictions such as conditions imposed on the free flow of data. 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.

The term "public security" is not defined in the Regulation itself, but is clarified in Recital 19. It must be understood within the meaning of Article 52 TFEU as interpreted by the Court of Justice and covers both the internal and external security of a Member State, on condition that a "*genuine and sufficiently serious threat affecting one of the fundamental interests of society*" is at stake. The proportionality requirement finally entails that any data localisation requirement adopted on such ground is suitable to achieve the objective pursued, and does not go beyond what is necessary to achieve that objective.

As regards any existing data localisation requirements, a double obligation is imposed on the Member States. On the one hand, they must repeal any existing laws or regulations which are not compliant with the abovementioned rules and on the other hand, they will need to justify any instances where they consider a certain data localisation requirement to be compliant and therefore intend to retain such requirement. All remaining data localisation requirements must moreover be published via a national online single information point.

Article 5 addresses the availability of (non-personal) data for authorities in the performance of their duties. It establishes that authorities may not be refused access to data on the basis that

[356] European Commission, 'Digital Single Market: EU Negotiators Reach a Political Agreement on Free Flow of Non-personal Data' (*European Commission*, 19 June 2018) <http://europa.eu/rapid/press-release_IP-18-4227_en.htm> accessed 20 September 2018

[357] European Commission, 'Joint statement by Vice-President Ansip and Commissioner Gabriel on the European Parliament's Vote on the New EU Rules Facilitating the Free Flow of Non-personal Data' (*European Commission*, 4 October 2018) <http://europa.eu/rapid/press-release_STATEMENT-18-6001_en.htm> accessed 25 October 2018

[358] Free Flow of Data Regulation, art 4

it is processed outside the authority's Member State. If the latter is the case, and the competent authority cannot get access, it may request assistance from a competent authority in the relevant Member State. The Regulation sets out a procedure for dealing with such requests for assistance, which involves the designation by each Member State of a "single point of contact" for this purpose.[359]

While the porting of data is addressed, no hard obligations are imposed. Instead, Article 6 states that the Commission will encourage and facilitate the development of self-regulatory codes of conduct at EU level. These codes of conduct should notably offer guidance on best practices in assisting end-users that wish to switch providers. They should also address the provision to professional users of detailed, clear and transparent information on the specificities of switching providers and porting data, before a contract for data processing is concluded.

3.6.4.2 Challenges encountered in the proposed Regulation

3.6.4.2.1 Scope of application: non-personal data

The Free Flow Regulation will apply to electronic data, with “data” being defined as all "*data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679*". The underlying rationale is that the new Regulation should not affect the existing framework for personal data protection. Instead, it should complement the GDPR and the e-Privacy Directive (2002/58/EC) to create a comprehensive and coherent EU framework for the free movement of all data in the digital single market.[360]

Upon closer analysis 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. This sub-Section addresses the practical concerns that stem from the decision to determine the scope of the Free Flow Regulation entirely in terms of personal data within the meaning of the GDPR. Other concerns will be addressed in the sub-Sections below.

As discussed in Section 3.1 on privacy and data protection, the definition of personal data under Article 4(1) of the GDPR is very far-reaching. We repeat it here for ease of reference. Personal data is "*any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly*".[361] The fact that this term applies so broadly entails questions about the *de facto* scope of the Free Flow Regulation.

We briefly address the possible extent of the term “personal data”, as clarified by the CJEU in its judgment of 12 May 2016 in case C-582/12, commonly known as the *Breyer* case, which is also touched upon in Section 3.4 of this Deliverable in the context of anonymisation.[362] It should be noted that, while the *Breyer* judgment concerns the interpretation of personal data

[359] Free Flow of Data Regulation, art 7
[360] COM (2017) 495 final, 3
[361] GDPR, art 4(1)
[362] Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:339

under the Data Protection Directive (95/46/EC), this term remains unchanged under the GDPR and the CJEU's interpretation remains relevant.

The central question in *Patrick Breyer v Bundesrepublik Deutschland* 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 ("**ISP**")). In this regard, the Court referred to Recital 26 of the Data Protection Directive, which read that in order to "*determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person*". It thus had to be established whether the possibility to combine a dynamic IP address with additional data held by an ISP constitutes a means likely reasonably used for identifying the data subject.

The Court gave a conditional answer to the issue. It considered that a dynamic IP address would constitute personal data in the hands of any party that "*has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person*". On the other hand, that same piece of information is not considered personal data in the hands of a party that cannot lawfully obtain sufficient additional data allowing it to identify the person in question. In essence, this judgment means that a piece of information can be considered personal data where additional information can be sought from third parties to identify the subjects.

When applying the principles of *Breyer* in practice, it is not unlikely that many individual pieces of data that *prima facie* seem to constitute non-personal data, still end up falling within the ambit of the GDPR's definition of personal data.

Recital 9 of the Free Flow Regulation offers examples of categories of such non-personal data. It mentions the Internet of Things, artificial intelligence, and machine learning as sources of non-personal data, for instance as used in automated industrial production processes. It goes on to list a few specific examples, notably "*aggregated and anonymised datasets used for big data analytics, data on precision farming that can help to monitor and optimise the use of pesticides and water, or data on maintenance needs for industrial machines*". While to a certain extent, this clarifies the Commission's intention, one may imagine situations of data (re-)combination and re-identification that would render even these types of data personal data. This gives rises to some uncertainty as to what data will actually fall within the scope of the Free Flow Regulation.

Free flow of data in the transport sector – Example 1

In its Opinion on processing personal data in the context of C-ITS, WP29 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 system 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. However, WP29 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.[363]

The above 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.

Our analysis above offers a glimpse of the uncertainty that could follow from tying the Free Flow Regulation's application entirely to the residual category of non-personal data. In its current wording, applicability of the Regulation is determined entirely based upon the character of the data. Only data localisation requirements for non-personal data are prohibited, and Member States could still impose data localisation requirements on personal data that would indirectly impact so-called non-personal data.

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 character of the data. This was based on the idea that the GDPR itself already eliminates a number of data localisation requirements. 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*".

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 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 the latter were the case, such restrictions would already be addressed by GDPR and the Free Flow Regulation would not (need to) apply.

The impact assessment report illustrates this through the requirement to store registers of shareholders domestically. As this data localisation requirement's objective is to allow

[363] 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

shareholders and interested third parties to access the corporate information, rather than to protect the shareholders' personal data, the requirement would fall within the scope of the Free Flow Regulation and therefore the GDPR would not apply.

It is unfortunate that this approach has been abandoned, as the current wording allows Member States to adopt data localisation requirements for personal data for purposes not related to data protection. These requirements could moreover cover data which seem to be non-personal at first sight, but which could still constitute personal data in light of the broad interpretation of the definition of personal data. Additionally, the current approach risks creating discrepancies in the regulation of the different types of data flows within the EU. Non-personal data would be subject to the broad localisation prohibition of the Free Flow Regulation (with the only exception allowed being for "reasons of public security"), while all personal data would be subject to the prohibition under Article 1(3) of the GDPR that applies only in respect of measures adopted for purposes of personal data protection.[364]

3.6.4.2.2 Mixed data sets

A second concern follows from the first, and involves mixed datasets of personal and non-personal data. 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 the dataset and the Free Flow Regulation to all non-personal data elements.

Article 2 of the Free Flow Regulation confirms that, in the event of a dataset composed of both personal and non-personal data, the Regulation shall only apply to the non-personal data part of that dataset. This entails that the applicable provisions of the GDPR must be fully complied with in respect of the personal data part of the set. Article 2 moreover clarifies that, in case personal and non-personal data in a dataset are "inextricably linked", the Free Flow Regulation should not prejudice the application of the GDPR. Recital 10 adds that the Free Flow Regulation does not "*impose an obligation to store the different types of data separately*".

In practice however, it will often not be possible to determine which parts of a dataset contain personal data and which contain non-personal data. Therefore, it will be impossible to apply each Regulation to the relevant part of the dataset. This could again create a loophole for Member States to still impose data localisation requirements on other grounds than public security, simply by making those requirements applicable to "personal data".[365]

The concern rises again when 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

[364] Cathal Flynn, 'Shortcomings of the EU Proposal for Free Flow of Data' (2018) 45(4) InterMEDIA 30, 34
[365] Ibid

becoming personal data. In such case, this dataset will fall entirely within the scope of the GDPR, and the Free Flow Regulation will no longer apply. [366]

3.6.4.2.3 Availability of data to competent authorities

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*". 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.

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

3.6.4.3 Opportunities of the proposed Regulation

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.[367]

Several respondents to the public consultation stressed the fact that start-ups act rationally when contemplating entry into new markets. Consequently, if scaling across the EU would prove more expensive than scaling up globally, start-ups will choose to access other global markets prior to entering any (other) European markets. The Free Flow Regulation may therefore make the EU market(s) more attractive to start-ups planning to scale up.[368]

[366] 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 17 October 2018

[367] 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-ADFF-3778-E8DD2021E5CC883B_46670.pdf> accessed 17 October 2018

[368] Ibid

Free flow of data in the transport sector – Example 2

While eliminating data localisation requirements should bring about the possibility of cost reductions for data services across the EU, the question remains to what extent companies will make use of these new opportunities. An instructive example in this respect is that of the Volvo Group.

The Swedish Volvo Group is a manufacturer of trucks and buses, among many other things, that sells its products in over 190 markets across the globe. These products are digitised, meaning that trucks in operation generate data which is used to help customers optimise cargo loading, fuel efficiency, etc. Real-time vehicle monitoring combined with big data analytics moreover enables performance analysis to be conducted and predictions to be offered to the end-users about repair requirements.

Data is clearly a key asset for the Volvo Group, reason for which they consider secure processing to be of paramount importance. As a consequence, Volvo handles most of its data locally, with servers located primarily in Sweden. While some outsourcing has been done, this was usually preceded by lengthy negotiations to sufficiently ensure security and minimise risks. Volvo thus considers the location of its servers to be important.[369]

This example shows that nationally or locally imposed data localisation requirements are not the only barriers to cross-border data flows. The perception of data being more secure when stored within a country's borders is not limited to national legislators, but is also present among businesses themselves. Volvo could choose to locate its servers in a jurisdiction of choice, potentially contracting data processing and storage space at much lower prices than those on the Swedish market. Still, security concerns prevent the Group from doing so.

This also means that any legislative intervention eliminating legal data localisation requirements will not necessarily eliminate such practices, and therefore will not impact the Union's status quo in terms of competition and innovation. This should be taken into account when estimating the projected impact of the Free Flow Regulation in increasing competition and innovation in the EU digital economy.

Other obvious benefits of reducing data localisation requirements stem from the increased competition across the EU it will produce. Currently, data services markets have widely varying characteristics. A respondent to the public consultation asserted that in Germany, a server for hosting health data costs EUR 3,000 annually, while its equivalent in France will cost EUR 13,000. Eliminating data localisation restrictions will put pressure on the highest of these price levels and level out market distortions. This would be another step towards the creation of an actual EU digital single market. It would moreover reduce administrative costs, complexity, and time loss. Start-ups increasingly rely on competitive cloud services for their own product or service. Prohibiting localisation restrictions would increase competitiveness of

[369] Kommerskollegium, 'No Transfer, No Trade – the Importance of Cross-Border Data Transfers for Companies Based in Sweden' (Swedish National Board of Trade 2014) 35 <https://www.kommers.se/Documents/In_English/Publications/PDF/No_Transfer_No_Trade.pdf> accessed 17 October 2018

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 achieving economies of scale. [370]

Free flow of data in the transport sector – Example 3

In 2014, Brussels Airport launched an idea to start developing cloud-based logistics applications. This resulted in the creation of BRUcloud, which uses the data sharing technology and cloud platform of Nallian to create an open environment where all partners can easily place and access the data needed to plan their day-to-day business, while still offering appropriate security. BRUcloud has already resulted in the creation of a new cargo community platform for BRUcargo.

BRUcloud's main priority is to make data sharing in a cloud environment possible. It enables the different stakeholders in the air cargo supply chain to work in a more integrated manner and increasingly act as a network. Data is stored only once in a central location. Once a company is connected to the cloud, it can start using the different existing applications and can start exchanging data very easily with other stakeholders instead of maintaining system-to-system connections with all different partners individually.

Applications are plugged into the BRUcloud and create quick and easy efficiency gains for the parties involved. Several applications have already been created to improve the cargo handling process.[371]

The increased competition in the EU's cloud services market that will result from eliminating data localisation requirements will make sure that more services such as BRUcloud will be created across the EU, which will create cost reductions and efficiency gains for all actors in the transport sector.

A similar example is offered by the Maritime Connectivity Platform which has been developed in the framework of the H2020 EfficienSea2 Project, which was led by the Danish Maritime Authority. The cloud-based platform offers a communication framework that enables efficient, secure and reliable information exchange in the maritime sector and has been used to develop over 15 end user services aimed at simplifying and streamlining information flows to maritime stakeholders and actors, including for instance route optimisation services based on big data analytics.[372]

[370] 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-ADFF-3778-E8DD2021E5CC883B_46670.pdf> accessed 17 October 2018

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

[372] EfficienSea2, 'Maritime Connectivity Platform' (*EfficienSea2*) <https://efficiensea2.org/solution/maritime-connectivity-platform/> accessed 17 October 2018

3.6.5 Summary

The free flow of data presents a scenario in which no legal barriers hinder the cross-border flow of data. Such cross-border data flows may be restricted by data localisation requirements, which come in many shapes and forms and be categorised as either strict or conditional requirements. Such data localisation requirements are often prompted by the perception that data are more secure when stored within a country's border. Data security however depends on the security measures that are implemented rather than on the location where the data is stored.[373]

The Free Flow Regulation should ensure the free flow of data across EU Member States, ensure data availability for regulatory control by EU authorities, and encourage the creation of codes of conduct for cloud services. Overall, the aim is for the Regulation to complement the GDPR and the e-Privacy Directive and thereby create a comprehensive and coherent EU framework for the free movement of all data in the digital single market.

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. Other obvious benefits of reducing data localisation requirements stem from the increased competition that should result from it.

The wording of the Regulation, as adopted by the European Parliament on 4 October 2018, presents some concerns however. First, the Regulation would apply only to non-personal data, meaning the residual category of data that do not constitute personal data under the GDPR. It is shown however that the latter category extends very broadly and one can easily imagine frequent situations of data (re-)combination and re-identification that would render data personal. This gives rise to some uncertainty as to what data will actually fall within the scope of the Free Flow Regulation.

This is however not the only issue that arises with respect to the Free Flow Regulation's scope, which is determined entirely based upon the character of the data. Since only data localisation requirements for non-personal data are prohibited, Member States could still impose data localisation requirements on personal data that would indirectly impact so-called non-personal data. Additionally, the current approach risks creating discrepancies in the regulation of the different types of data flows within the EU. Non-personal data would be subject to the broad localisation prohibition of the Free Flow Regulation, while all personal data would be subject to the prohibition under Article 1(3) of the GDPR that applies only in respect of measures adopted for purposes of personal data protection.

Finally, the Free Flow Regulation creates uncertainty for service providers, as it does not foresee the situation in which a disclosure of data is required under the Regulation but prohibited by the Member State in which the data is located. It also does not introduce any

[373] Daniel Castro, 'The False Promise of Data Nationalism' (ITIF 2013) 1 <http://www2.itif.org/2013-false-promise-data-nationalism.pdf> accessed 17 October 2018

safeguards to such access by competent authorities, to protect e.g. intellectual property rights of third parties or data protected by commercial confidentiality such as trade secrets.

Further guidance from the European Commission is required to help organisations assess their obligations under the Free Flow Regulation in light of the above-mentioned issues.

Opportunities in relation to free flow of data in the context of big data in the transport sector	Challenges in relation to free flow of data in the context of big data in the transport sector
Stakeholders expect significant cost reductions for cloud storage and processing, necessary for big data analytics services in the transport sector.	The Free Flow Regulation will only apply to non-personal data, which entails uncertainty as personal data is broadly defined.
Eliminating data localisation requirements will lead to increased competition. This in turn will create more innovation, which will positively impact big data analytics in the transport sector.	There is currently a lack of clarity about which legal instruments apply to mixed datasets composed of both personal and non-personal data, which will very often be the case for big datasets in the transport sector.
It will be easier for SMEs and start-ups to access new markets when data localisation requirements are eliminated.	No safeguards (e.g. for third parties' IP rights or to protect the commercial value of trade secrets) are established concerning access by competent authorities to non-personal data.

Table 22: Summary table of opportunities and challenges in relation to free flow of data in the context of big data in the transport sector

3.7 *Intellectual property in big data environment*

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.374 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.[375]

In light of this definition, 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 Section examines those intellectual property rights that may be relevant in a big data context. This Section will therefore look into the particular application in a big data environment of (i) copyright; (ii) database rights; and (iii) trade secrets and confidentiality.

3.7.1 Copyright

It seems to be widely agreed amongst legal scholars that copyright can be invoked to protect, to a certain extent, non-personal and commercial data.

Below, we describe the current legal framework for copyright and explain the scope of the protection available. We also examine the relevant copyright aspects from a transactional point of view.

3.7.1.1 Legal Framework

The rules governing copyright protection have been established at international, regional and national level. In order to understand the protection granted to literary and artistic works, one needs to become familiar with this entire legal framework.

3.7.1.1.1 International Legal Framework

There exists no single international copyright instrument that would automatically confer uniform protection on literary and artistic works worldwide. However, international treaties, conventions and trade agreements were established as from the 19th century in order to ensure a minimal level of legal protection to creators of original works.

The international legal framework for copyright is based on the following principles:

[374] 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

[375] Ibid

- The **territoriality principle** refers to the fact that copyright is of a territorial nature and that national laws can only rule on conducts occurring within national borders.[376]
- According to the **national treatment principle**, a country must provide the nationals of other countries, party to the same international instruments, with a treatment no less favourable than the one it accords to its own nationals with regard to such rights.[377] There are however certain exceptions to this principle.
- **Reciprocity** is the negation of the national treatment principle as it refers to making protection, or the extent of protection, in a given country (A) of copyright or related rights of nationals of another country (B) conditional on the existence of the same (or at least similar) extent of protection granted in that other country (B), to the nationals of the country concerned (A).[378]

This Report does not aim at analysing issues in relation to territoriality, national treatment, jurisdiction, and conflicts of laws. It is however important to understand that the above principles are necessary in the field of copyright because copyright laws are not identical between countries.

It is important to keep in mind that the international treaties provide for minimum standards only and individual countries may therefore provide for additional protection. Also, treaties do not cover some important issues like ownership and transfer of rights.

The main international instruments of copyright law are the following:

- The Berne Convention;[379]
- The Universal Copyright Convention;[380]
- The TRIPS Agreement;[381]
- The WIPO Copyright Treaty.[382]

3.7.1.1.2 European Union Legal Framework

In addition to the international treaties, to which the European Union and the 28 Member States are contracting parties, a series of EU Directives was also adopted to harmonise various discrepancies between the copyright laws of the Member States, notably between civil law and common law jurisdictions.

---

[376] This has been confirmed by the Court of Justice of the European Union in *Lagardère*, wherein it states that "*it must be emphasised that it is clear from its wording and scheme that [the Rental and Lending Directive] provides for minimal harmonisation regarding rights related to copyright. Thus, it does not purport to detract, in particular, from the principle of the territoriality of those rights, which is recognised in international law and also in the EC Treaty. Those rights are therefore of a territorial nature and, moreover, domestic law can only penalise conduct engaged in within national territory*" Case C-192/04 *Lagardère Active Broadcast v Société pour la perception de la rémunération équitable and others* [2005] ECLI:EU:C:2005:475, para 46

[377] World Intellectual Property Organization, 'Guide to the Copyright and Related Right Treaties Administered by WIPO and Glossary of Copyright and Related Rights Terms' (WIPO 2003) 297 <http://www.wipo.int/edocs/pubdocs/en/copyright/891/wipo_pub_891.pdf> accessed 17 October 2018

[378] Ibid 306

[379] The Berne Convention for the Protection of Literary and Artistic Works of 9 September 1886

[380] The Universal Copyright Convention adopted in Geneva on 6 September 1952, as revised in Paris on 24 July 1971

[381] The Agreement on Trade-Related Aspects of Intellectual Property Rights, adopted in Marrakech on 15 April 1994 and which corresponds to Annex 1C to the Agreement establishing the World Trade Organization

[382] The World Intellectual Property Organization Copyright Treaty, adopted in Geneva on 20 December 1996

In spite of these Directives, there currently is no common and fully harmonised legal framework for copyright within the EU. Since copyright laws are to a great extent territorial in each Member State, the international treaties and national legislations remain important sources of copyright law.

The most important EU instrument related to copyright is the Information Society Directive (the InfoSoc Directive)[383]. It aims at (i) adapting the legislations on copyright and related rights to reflect the technological developments; and (ii) transposing into EU law the main international obligations arising notably from the WIPO Copyright Treaty.

In addition to the InfoSoc Directive, the EU has adopted other instruments particularly important in the context of copyright, such as in particular:

- The Database Directive;[384] (see sub-Section 3.7.2 below for further details)
- The Rental and Lending Directive;[385]
- The Copyright Term Directive;[386]
- The Software Directive;[387]
- The Orphan Works Directive.[388]

The CJEU has played an important role in the harmonisation of copyright (and database rights) by interpreting the various Directives listed above, and in particular the InfoSoc Directive and the Database Directive. Although legal systems of most of the EU Member States are based on continental law, which entails that they do not directly attach legal consequences to case law like common law countries, the judgments of the CJEU play an important role in providing a binding interpretation of EU law.

3.7.1.1.3 The EU Copyright Reform

On 14 September 2016, the Commission published several legislative proposals aiming to modernise the existing EU copyright rules.[389] The so-called Copyright Package consists of two Directives and two Regulations.

What seems to be at the core of the reform, from the point of view of digital services providers, are the Proposals for a Directive on copyright in the Digital Single Market[390] (the "DSM Directive") and for a Regulation laying down rules on the exercise of copyright and

[383] Directive 2001/29/EC of the European Parliament and of the Council on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L 167/10 (InfoSoc Directive)
[384] Directive 96/9/EC of the European Parliament and of the Council on the legal protection of databases [1996] OJ L 77/ 20 (Database Directive)
[385] Directive 2006/115/EC of the European Parliament and of the Council on rental right and lending right and on certain rights related to copyright in the field of intellectual property [2006] OJ L 376/28 (codified version – replacing Directive 92/100/EEC) (Rental and Lending Directive)
[386] Directive 2006/116/EC of the European Parliament and of the Council on the term of protection of copyright and certain related rights [2006] OJ L 372/12 (codified version – replacing Directive 93/98/EEC; amended by Directive 2011/77/EU) (Term Directive)
[387] Directive 2009/24/EC of the European Parliament and of the Council on the legal protection of computer programs [2009] OJ L 111/16 (Software Directive)
[388] Directive 2012/28/EU of the European Parliament and of the Council on certain permitted uses of orphan works [2012] OJ L 299/5 (Orphan Works Directive)
[389] More information available at <https://ec.europa.eu/digital-single-market/en/modernisation-eu-copyright-rules>
[390] Commission, 'Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market' COM (2016) 593 final

related rights applicable to certain online transmissions of broadcasting organisations and retransmissions.[391] According to the Commission's announcement, these Proposals aim to ensure better choice and access to content online and across borders, improved copyright rules on research, education and inclusion of disabled people, as well as a fairer and sustainable marketplace for creators, the creative industries and the press.

More specifically the Proposal for a DSM Directive introduces a new related right for press publications, which is supposed to give the press industry a stronger bargaining position to protect their investments, explore new business models and eventually complete its transition to the digital environment.

The Proposal also further harmonises the copyright exceptions by introducing three new mandatory exceptions covering:

- text and data mining by research organisations, for the purposes of scientific research, of copyright protected content to which they have lawful access (this exception is limited to non-commercial purposes) (see sub-Section 3.7.1.4.2 below for further details);
- digital uses of works or other protected content for the purposes of illustration for teaching; and
- copying by cultural heritage institutions of works that are permanently in their collection for the purpose of preservation of cultural heritage (this exception will cover works that were created directly in digital form as well as the digitisation of works in analogue formats, and will help audiences to access them for longer).

This Proposal for a DSM Directive is consistent with the existing EU copyright legal framework, and is thus based upon, and aims to complement the rules laid down in various copyright directives, including the InfoSoc and Database Directives.

The legislative Proposals presented by the Commission were submitted to the European Parliament and to the Council for adoption.[392] The Council reached an agreement on a mandate to open negotiations with the European Parliament on the DSM Directive. After a positive vote in the European Parliament on 12 September 2018, the co-legislators are negotiating the text and intend to finalise the process by the end of 2018. At the time of publication of this Deliverable, the proposal for a Regulation laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions is being finalised between the European Parliament, the Council, and the Commission.

It shall nevertheless be noted that the Proposals do 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. They however include, as indicated

[391] Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes' COM (2016) 594 final

[392] Once the proposals are adopted, the Member States will have two years to implement the Directives into the national law.

above, a new – yet narrow – exception concerning text and data mining (see sub-Section 3.7.1.4.2 below for further details).

3.7.1.1.4 National Legal Framework

As indicated above, international treaties lay down the core principles of copyright protection. Various EU Directives, as interpreted by the CJEU, provide for a certain degree of further harmonisation in the EU.

However, 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.

3.7.1.2 Copyright Protection: General Overview

3.7.1.2.1 Scope of Copyright Protection

To understand to what extent copyright may be used to protect non-personal data, one must first understand what types of creations can be protected, and what are the terms and scope of such protection.

The Berne Convention presents a broad non-exhaustive list of works protected under copyright:[393]

> *"The expression "literary and artistic works" shall include every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression, such as books, pamphlets and other writings; lectures, addresses, sermons and other works of the same nature; dramatic or dramatico-musical works; choreographic works and entertainments in dumb show; musical compositions with or without words; cinematographic works to which are assimilated works expressed by a process analogous to cinematography; works of drawing, painting, architecture, sculpture, engraving and lithography; photographic works to which are assimilated works expressed by a process analogous to photography; works of applied art; illustrations, maps, plans, sketches and three-dimensional works relative to geography, topography, architecture or science.*"

It derives from that list that copyright protection has a broad scope but, at the same time, requires an intellectual human intervention and the consciousness of achieving a result. Therefore, raw data such as weather forecasts, stock quotations or sports scores would in principle be excluded from copyright protection.

The EU legal framework does not provide for a list of protected works like the Berne Convention does. Member States have however implemented Article 2(1) of the Berne Convention directly in their national legal frameworks. This approach implies that, in principle,

[393] Berne Convention, art 2(1)

any type of work can enjoy copyright protection as long as it meets the legal requirements for such protection.

Because they do not meet the fundamental requirements for copyright protection, copyright statutes and treaties (particularly the TRIPS Agreement[394] and the WIPO Copyright Treaty[395]) exclude mere ideas from copyright protection. However, the expression of such ideas may be protected.

In the same vein, because their subject matter is considered as being outside the scope of copyright protection, mathematical concepts, methods of operation, gambling procedures, and other intellectual tools are also excluded from copyright protection.

Furthermore, for a work to be protected, it must fulfil two cumulative conditions:

- it shall be fixed in some material (concrete) form;
- it shall be original, meaning that 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.

The originality criterion implies that some categories of data will not be protected by default. Having said that, the threshold for originality is rather low in the EU Member States, and even more in some of them (this is for instance the case in the Netherlands, in France and in Belgium).

United Kingdom

UK law places particular emphasis on the formal expression of an idea as being at the heart of copyright protection. Hence, certain forms may not be protected by copyright; e.g. technical features such as the functionality, programming language and interfaces (such as data file formats) of computer programs are not themselves protected by copyright although the software's source code which creates them is.[396]

The right owner (or the right holder in case of transfer of rights) of copyright protected works will enjoy various exclusive economic rights; i.e. to reproduce, communicate to the public and distribute the work. Accordingly, save where copyright exceptions apply, the author's consent is necessary to perform any of these activities (see sub-Section 3.7.1.4.1 below for further details).

[394] Art 9.2

[395] Art 2

[396] In *SAS Institute Inc v World Programming Limited* [2013] EWCA Civ 1482, Lewison LJ found that both the Software and the InfoSoc Directives incorporated the underlying principle from the Berne Convention that it was the form of expression rather than the underlying idea which was protected. The Court of Justice of the European Union found that whether it applied the Software Directive or the InfoSoc Directive, the functionality of the software in issue was not protected given that the functionality was the idea, but the source code was the expression in which that idea was embodied.

Moreover, authors are also granted so-called "moral rights". The concept of "moral rights" is the consequence of the predominant view in (continental) European copyright law that a work is not a mere staple commercial object, but also the expression of the personality of the author. Moral rights are recognised by the Berne Convention.[397] Its Article 6*bis* provides for minimum standards in this respect: the author has the right, even after the transfer of the economic rights, to 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. By contrast, the EU Directives explicitly exclude moral rights from their scope. More particularly, Recital 19 of the InfoSoc Directive stipulates that moral rights remain outside the scope of the Directive and that they should be exercised according to the legislation of the Member States and the provisions of the international treaties. It follows from such situation that moral rights suffer many discrepancies between Member States. Indeed, while some countries provide for a high level of protection of moral rights, others recognise such rights only within the minimum protection imposed by the Berne Convention. Some Member States even foresee additional moral rights.

France

One of such moral rights is the French *"droit de repentir"*, which can be particularly relevant in a big data context since it allows the author to take his work back from the commercial circuit, making further exploitation of such work impossible.[398] Even if the author needs to compensate the person who acquired the economic rights to the work for such a withdrawal, exercising the *"droit de repentir"* could prejudice an entire (big data) project based on the withdrawn work.

The general eligibility for copyright protection differs to a certain degree from one EU Member State to another. Indeed, the abovementioned abstract concepts have usually been specified in detail through the case law of each national state. Having said that, the numerous CJEU decisions in copyright-related matters lead to a gradual unification of the EU legal framework for copyright, with the notable exception of moral rights.

3.7.1.2.2 Ownership of Rights

In general, the copyright belongs to the (physical) author of the work. In case of works created by two or more persons, the copyright would be awarded to these persons jointly.

In the jurisdictions (such as France) recognising collective works (such as encyclopaedia or periodic reviews), the economic rights to such works will normally belong to their producer or publisher.

The question of ownership is usually more complex in case of works created by an employee. In some countries, the economic rights to works created by an employee in result of carrying

[397] At international level, moral rights are also recognised by article 5 of the WIPO Performances and Phonograms Treaty of 1996 and article 5 of the Beijing Treaty on Audiovisual Performances adopted in 2012.
[398] French Intellectual Property Code, art L.121-4

out his contractual employment duties will be automatically transferred to his employer (unless otherwise agreed in the contract), whereas in other countries such presumption does not exist.

3.7.1.2.3 Works in the Public Domain

Authors of protected works benefit from copyright during their entire life, and these rights are maintained for a period of 70 years after their death (or the death of the last author), before falling into the public domain. In the EU Member States, the initial length of protection was 50 years after the author's death (as it is still prescribed by Article 7(1) of the Berne Convention) but the Term Directive increased the protection term to 70 years in the EU.[399]

Once a work falls into the public domain, it means that it can be freely exploited, reproduced or executed. No authorisation is needed and no royalties must be paid for the use of the work.

3.7.1.3 Copyright Protection of Data

From a business perspective, the actors involved in the big data lifecycle may wish to claim protection over (big) data in order to secure the economic investment made in obtaining, verifying, storing, presenting and analysing the data.

Copyright enables protection of non-personal and commercial data to a certain extent. However, it is crucial to distinguish different elements used in the operations on big data that can benefit from copyright protection.

Individual data, understood as pieces of information, can be protected by copyright as long as they fulfil the conditions set out in the relevant legislation (in particular fixation in a tangible form and originality – see above for further details).

As the trend to modernise the existing legal framework confirms, the traditional copyright laws have struggled to deal with new technologies and digital content distribution methods. In practice, it means that the current laws do not contain provisions that would directly address the use of protected works when applying new technologies and in the new digital context, in particular for cloud computing, text and data mining[400], or big data projects. Nevertheless, the copyright rules will still apply, providing protection for those materials used that could be classified as works and would fulfil the protection requirements described in sub-Section 3.7.1.2.1 above.

In the context of big data projects, it is crucial to understand to what extent the data used can be copyright protected. 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.

Also, given that the copyright legal framework does not provide for a registration system (unlike trademarks and patents), copyright protection will only be confirmed *a posteriori* by a

[399] With the notable exception of France, where moral rights are perpetual, in most jurisdictions the 70-year term of protection after the death of the author applies to both economic and moral rights.
[400] See however sub-Section 3.7.1.4.2 below which discusses the proposed new exception to cover text and data mining.

court. Such characteristic of copyright is particularly problematic in a context of data, posing issues in terms of legal certainty.

Having said that, there are some objective criteria that can facilitate analysing whether or not specific data is protected. In particular, the data need to fulfil the two basic requirements for copyright protection – they need to be fixed in some material (concrete) form and they need to be original.

In this context, 'fixation' of data means 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.

The originality requirement can bring even more difficulties, since the evaluation of a work's originality leaves some room for discretion and requires, in any event, a human intervention in the creation process, whereby he/she can stamp the work with his/her personality.

In general, in order to be considered original, the data should represent a level of sophistication suggesting that no one else than the author could have created the same work even if based on the same raw data (e.g. summary texts).[401] Having said that, it is indeed so that the originality threshold for copyright protection in most of the EU Member States is rather low. Even a low level of creativity can therefore prove to be sufficient to claim protection. Such low threshold is however criticised by numerous scholars.

On that basis, we can attempt to identify what types of data will be more (and less) likely to benefit from copyright protection. However, such identification should only be seen as indicative. In practice, every piece of data would have to be evaluated on a case-by-case basis in order to determine whether it can be copyright protected in a particular country.

The following data, in our view, could more easily attract copyright protection:

- data in the form of free text;
- data presented in graphic form.

On the other hand, for the following types of data it would, in our view, be particularly difficult to demonstrate originality:

- raw numbers and other purely quantitative information;
- measurements results (e.g. measurements of temperature, pressure, other natural phenomena);
- financial results, prices of products and similar market data;

[401] If we were to ask ten photographers to take a picture of the same object on the same day and time, we would most probably still obtain ten different photographs, each of them embedding an individual view of the artist, and reflecting his/her artistic effort.

- sport results, competition results;
- demographic data;
- results of automated processes (e.g. video recording from security cameras, statistics on the use of electricity, water, use of the telephone (number of calls, use of data transfer), Internet (e.g. use of the browser)).

It follows that, in all likelihood, most of the data collected and processed in a big data analytics context will 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).

3.7.1.4 Exclusive Rights and Copyright Exceptions

In the event a particular piece of data is protected by copyright, the right owner (or right holder) will be granted several exclusive rights. Accordingly, when a work is protected, seeking authorisation is a requirement in order to reproduce, communicate or make available to the public, distribute, rent, lend, adapt, translate or alter such work. However, copyright laws include various exceptions (limitations) where, under specified conditions, such authorisation is not required.

3.7.1.4.1 Exclusive Rights

On the basis of the InfoSoc Directive, Member States are required to implement the following set of exclusive rights:

Reproduction	Communication to the public	Distribution
Exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part.[402]	Exclusive right to authorise or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access them from a place and at a time individually chosen by them.[403]	Exclusive right to authorise or prohibit any form of distribution to the public by sale or otherwise.[404]

Table 23: Overview of the author's exclusive rights

[402] InfoSoc Directive, art 2
[403] InfoSoc Directive, art 3
[404] InfoSoc Directive, art 4

From a technical perspective, (big) data analysis generally requires the **reproduction** of the data, in part or in whole. It can therefore not be excluded that a (big) data analysis process leads to the reproduction of copyright-protected data, requiring thus the authorisation of the right owner/holder (except if one may rely on one or more copyright exceptions – see below).

Indeed, the reproduction right under the InfoSoc Directive has an extensive meaning, as a broad definition is needed to ensure legal certainty within the internal market.[405] Such view was confirmed by the CJEU:

- Infopaq judgment (C-5/08) of 16 July 2009: "*an act occurring during a data capture process* [e.g., creation of TIFF file by scanning, transferred to an OCR server to be translated in order to be processed digitally]*, which consists of storing an extract of a protected work comprising 11 words and printing out that extract, is such as to come within the concept of reproduction in part within the meaning of Article 2*" (Infopaq, operative part).
- Premier League (C-403/08) of 4 October 2011: "*the reproduction right extends to transient fragments of the works within the memory of a satellite decoder and on a television screen*" (Premier League, operative part).

It follows that even technical reproductions or the transformation of files into other file formats will qualify as a "reproduction" within the meaning of EU copyright law. In our view, such conclusion is criticisable in the context of the digital era. Indeed, in many instances, the (partial) reproduction of protected works is performed for merely technical reasons (e.g. cache copies, temporary copies in the RAM, back-up reproductions to prevent data loss, copies made for availability and high performance purposes such as by CDNs, etc.).

As for the **communication to the public** right, it is less problematic in our view in a (big) data analysis context. Such conclusion applies despite the far-reaching concept of communication to the public, as interpreted by the CJEU, which notably concluded the following:

- Establishing whether a communication to the public took place requires an individual assessment (*Phonographic Performance*, C-162/10, judgment of 15 March 2012).
- For there to be a communication, the user's intervention needs to be of a deliberate character, and 'public' refers to an indeterminate number of potential viewers, implying a fairly large number of people, who may access the work simultaneously or cumulatively (*SCF*, C-135/10, judgment of 15 March 2012).
- The concept of 'communication to the public' includes two cumulative criteria, i.e. an 'act of communication' of a work and the communication of that work to a 'public' (*Svensson and Others*, C-466/12, judgment of 13 February 2014).
- To be categorised as a 'communication to the public', a protected work must be communicated using specific technical means, different from those previously used or, failing that, be directed to a 'new public', i.e. a public that was not already taken into

[405] InfoSoc Directive, Recital 21

account by copyright holders when they authorised the initial communication to the public of their work (*Svensson* and *BestWater International*, C-348/13, order of 21 October 2014).

- The making available of audio-visual content through a technology other than that previously authorised by right holders entails *ipso facto* a new communication to the public that requires a new authorisation of the right holders. (*ITV Broadcasting and Others*, C-607/11, judgment of 7 March 2013).
- In order for an entity to be qualified as the one carrying out the communication, its intervention must be essential in order for the public to be able to access the work, and therefore not a mere technical means of ensuring or improving reception of the work (*SBS Belgium*, C-325/14, judgment of 19 November 2015).
- Posting a hyperlink to a work that is freely available on a website without the right holder's consent is an act of communication to the public if the linker knew or should reasonably have known that the original publication was unauthorised. In circumstances where the link is posted for financial gain, such knowledge is to be presumed. (*GS Media*, C-160/15, judgment of 8 September 2016)

Finally, with respect to the **distribution** right, it may apply in the context where the recipients of the data are not only those involved in the (big) data analysis process. Hence, where the dataset is copied and distributed to the public (e.g. if it is sold), the exclusive distribution right may apply. This being said, the distribution right finds an important limit in the exhaustion (also known as "first sale") doctrine. It allows the resale of copies of works (or related subject matter) without authorisation once the protected work has been put in the market with the consent of the right owner/holder.

IP in a big data environment in the transport sector – Example 1

Real-time public transport data (i.e. data communicated from GPS units on public transport vehicles to a central hub) has increasingly been used for diverging purposes, such as fleet management, performance management, safety and security considerations, etc.[406] Such data however have customer-oriented uses as well, as they may provide useful information about predicted arrival times. In the United States, for example, public transport or transit authorities relied on third-party service providers for the installation of the necessary equipment and the data collection, processing, and formatting. Some of those service providers (e.g. NextBus against Routsey in San Francisco[407]) subsequently tried to claim copyright in the predictive data produced by them, which had as an undesired consequence that the data could not be made public. Such copyright claims may be even more problematic in a smart cities environment.

[406] Teresa Scassa, 'Public Transit Data Through an Intellectual Property Lens: Lessons About Open Data' (2014) 41(5) Fordham Urb. L.J. 1759

[407] Eve Batey and Matt Baume, 'Does a Private Company Own Your Muni Arrival Times?' *The San Francisco APPEAL* (25 June 2009) <http://sfappeal.com/2009/06/who-owns-sfmta-arrival-data/> accessed 17 October 2018

3.7.1.4.2 Copyright Exceptions and Limitations

The InfoSoc Directive also provides for certain exceptions (or limitations) to the exclusive rights. Its Article 5 contains one mandatory exception for temporary acts of reproduction, and an exhaustive list of optional exceptions that Member States may implement into their national law.

Foremost, a mandatory exception to the right of reproduction is introduced with respect to certain temporary acts of reproduction which are integral parts to a technological process.[408] Generally, such exception concerns transient copies with a merely technical function and without any independent economic significance, in order to cover issues related to caching and Internet browsing.[409]

The CJEU has had the opportunity to examine such mandatory exception in several cases - i.e. Infopaq I[410] and II[411], and Premier League[412]. Apart from stating that a copyright exception must be interpreted restrictively, it also indicated that several cumulative conditions must be met in order to benefit from such exception:[413]

- The temporary copy must be transient or incidental. This means that the copy may only be ephemeral or at least non-permanent.
- The copy must be an integral and essential part of a technological process. In such context, the CJEU confirmed that the concept of the 'integral and essential part of a technological process' requires the temporary acts of reproduction to be carried out entirely in the context of the implementation of the technological process and, therefore, not to be carried out, fully or partially, outside of such a process. It further stated that this concept also assumes that the completion of the temporary act of reproduction is necessary, such that the technological process concerned could not function correctly and efficiently without that act.[414]
- The sole purpose of the copy must be to either enable a transmission in a network between third parties and an intermediary, or a lawful use of a protected work.
- The temporary copy must have no independent economic significance. Accordingly, the temporary reproduction cannot enable the generation of an additional profit, going beyond that derived from lawful use of the protected work. Also, the acts of temporary reproduction cannot lead to a modification of the work.[415]

[408] InfoSoc Directive, art 5.1
[409] InfoSoc Directive, Recital 33
[410] Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465 ("Infopaq I")
[411] Case C-302/10 Infopaq *International A/S v Danske Dagblades Forening* [2012] ECLI:EU:C:2012:16 ("Infopaq II")
[412] Joined Cases C-403/08 and C-429/08 *Football Association Premier League Ltd and Others v QC Leisure and Others* (C-403/08) and *Karen Murphy v Media Protection Services Ltd* (C-429/08) [2011] ECLI:EU:C:2011:631 ("Premier League")
[413] Jean-Paul Triaille, 'Study on the Legal Framework of Text and Data Mining (TDM)' (De Wolf & Partners 2014) <https://publications.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en> accessed 17 September 2018
[414] Infopaq II, para 30
[415] Infopaq II, para 54

It results that the acts performed on data in the data value cycle, by various stakeholders, may hardly rely on the exception for temporary acts of reproduction. This is notably supported by the fact that the acts performed in a big data analytics process can have a great economic value contrary to the last condition set under Article 5.1 of the InfoSoc Directive.

Accordingly, in our view, there is little to no legal certainty with the current exception for temporary acts of reproduction. Hence, authorisation remains a key requirement without which any use of a copyright protected work would amount to copyright infringement and would give rise to possible enforcement proceedings.

In addition to the mandatory exception discussed above, the InfoSoc Directive provides for several optional exceptions. Discretion is left to the Member States with regard to the transposition of such exceptions (e.g. with regard to the conditions and practical arrangements of such exceptions); consequently, the scope of exceptions differs largely between the Member States.

As part of the optional exceptions, there is one related to the use of a copyright protected work for the sole purpose of illustration for teaching or scientific research, for non-commercial purposes. Such exception however also presents some limits and is not entirely satisfactory as to provide sufficient certainty in case a stakeholder of the data value cycle wishes to rely on it. Also, its non-mandatory nature has led to having discrepancies between Member States, which is certainly not desirable in a (big) data analysis context.[416]

Finally, it should be mentioned that the EU Commission has made a Proposal in the context of the EU Copyright reform (see above) to include a new mandatory exception to cover "text and data mining". The definition of such concept is however broader and appears to cover the more general idea of "data analysis".[417]

Such new exception would allow reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject matter to which they have lawful access for the purposes of scientific research. Unfortunately, such exception would be very narrow, similar to what has been adopted in the United Kingdom, where "text and data mining" is excluded for commercial purposes.

Although research organisations should also benefit from the exception when they engage into public-private partnerships[418], it is regrettable that the Proposal provides for such a narrow opening to permit the reproduction of data in an analysis context. It is indeed short-minded to restrict acts of reproduction for the mere purpose of data analysis to the scientific sector.

In our opinion, there is a missed opportunity to rethink the essence of copyright and introduce a new infringement test through the creation of a new requirement: the use of the

[416] See the study of Jean-Paul Triaille 'Study on the Legal Framework of Text and Data Mining (TDM)' (De Wolf & Partners 2014) <https://publications.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en> accessed 17 September 2018

[417] Proposal for a Directive in the DSM, art 2.2: "'*text and data mining' means any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations*".

[418] Proposal for a Directive in the DSM, Recital 10

copyright-protected work as a "work".[419] Indeed, in many instances, when a protected work is used for (big) data analysis, it is not reproduced with the same aim as the original purpose. Accordingly, the right owner (or right holder) should not be in a position to hinder the reproduction of (part of) its work which is to be used for a different objective, especially in case such other objective fulfils the main conditions of the three-step test; i.e. it does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the right holder.[420]

3.7.1.5 Contractual Aspects

What is crucial in the context of big data projects is for the data to be operational and transferable. In particular, one needs to have the possibility to aggregate, reproduce, filter, enrich, merge, partition, share, etc. the data, and use them as a source of information. To that end, the ownership of data needs to be identified and access rights secured (see also Section 3.10 on data ownership). It is also necessary to ensure a possibility to convey data to a third party (i.e. make the data available on the basis of transfer or licence agreements), without risking that the acquirer will have to face property claims from the (alleged) data owner.

Therefore, in case data benefits from copyright protection, it is important to understand the rules governing the conclusion of transfer and licence agreements.

3.7.1.5.1 Transfer of Copyright

In order to permanently assign the ownership of a right to another person, a so-called transfer agreement should be concluded. Although the rights can also be donated or inherited, the transfer agreement (which can be compared to a sale) is the most common tool used in a commercial context.

When concluding a transfer agreement, it is very important to precisely describe the subject matter of the rights that will be transferred (i.e. the work), and, if possible, attach a copy of such work (or its specification) as an annex to the agreement. One should however note that in certain Member States, it is not possible to transfer rights to "all works" of a specific author or even "all works of a specific type"; if such provisions are included in the copyright transfer agreement, they may be declared null and void.

For practical reasons, it is also crucial to indicate a precise date of the copyright transfer. That allows establishing as of when one can use the right, and is important in case of infringements, as well as determining tax obligations.

[419] Such kind of test already exists to a certain extent in the field of trademarks. See also Alain Strowel, 'Reconstructing the Economic Rights: Taking Copyright Seriously' (2016), presented in the context of the research project "Reconstructing rights: Rethinking copyright's economic rights in a time of highly dynamic technological economic change, led by IVIR & CREATe.

[420] The latter two conditions derive from the three step test enshrined under Article 5.5 of the InfoSoc Directive. The latter provides the exceptions and limitations permitted by the InfoSoc Directive are to be applied (i) in certain special cases, (ii) which do not conflict with the normal exploitation of the work or other subject matter and (iii) which do not unreasonably prejudice the legitimate interests of the right holder (or other right holders).

Distinction needs to be made between transferring the ownership of a physical copy of the work (e.g. a movie on DVD, piece of furniture, or a sculpture) and transferring the copyright in relation to that work (intangible rights to the movie, the design of the furniture or sculpture). Selling a physical copy of the work will in principle not transfer the copyright to such work.

Since the moral rights are not attached to the economic rights, they will not be transferred along with the economic rights. Therefore, depending on whether the national jurisdiction allows for the transfer of moral rights, one should include in the copyright transfer agreement additional provisions, such as:

- provisions transferring the moral rights of the author to another party (if possibility to transfer such rights is recognised by the jurisdiction);
- provisions authorising the acquirer of the economic rights to exercise moral rights on behalf of the author (most common solution); or
- provisions imposing an obligation on the author not to exercise his moral rights against the acquirer of the economic rights.

In addition, inclusion of a separate clause is necessary to allow the acquirer to be able to create derivative works. Otherwise the acquirer will only be able to use the original work 'as is', and will not be able, on the basis of the transfer agreement, to modify the works acquired. This is a very important element of the agreement, since even the slightest adaptation of the work, even in its technical sense, may be seen as a modification that leads to the creation of a new, derivative work.

This is particularly relevant when considering the transfer of copyright in a big data context (provided such protection applies). Indeed, datasets are usually combined, processed and altered in some way. Also, the ultimate goal of big data analytics is to perform an analysis and thus create "derivative" information on the basis of the initial dataset(s).

Most jurisdictions require concluding the copyright transfer agreement in writing. Where this is the case, a transfer concluded with the omission of the written form will be considered null and void.

Also, in some jurisdictions it is necessary to indicate the so-called fields of exploitation of the work (e.g. in Poland it is an obligatory element of every copyright transfer agreement – if the fields of exploitation are not mentioned, the agreement risks to be considered null and void[421]).

3.7.1.5.2 Licence Agreements

By concluding a licence agreement, the right holder authorises a third party to use the work within the limits indicated in the licence, usually in return for remuneration (licence fees). The ownership of the copyright remains with the right holder.

There exist different types of licence agreements, depending on the territory, for how long and with what level of autonomy the licensee will be able to use the work.

[421] Polish law on copyright and related rights of 1994, art 41

Since the copyright protection remains territorial due to differences in the national law of the Member States, the licence agreement will often not cover a territory larger than one country. In practice, this is quite problematic – in order to be able to use a given work in the entire territory of the EU one will need to conclude 28 separate agreements.

A licence can be exclusive, guaranteeing the licensee to be the only entity able to use the licensed work in a given territory. This provides the licensee with an obvious competitive advantage, but also impacts the (higher) level of the licensor's remuneration.

In most countries licence agreements, contrary to transfer agreements, do not have to be concluded in writing.

3.7.1.5.3 Free and Open Licences

There are many initiatives aiming at facilitating the use and re-use of works to stimulate creativity, facilitate expression and enable fast sharing of information. These initiatives usually rely on what is popularly called "open" or "free" licences. Such licences can be characterised by two common features – they have predefined terms and conditions and are available for all the interested parties (everyone can enter into such agreement).

Also, since in most jurisdictions it is impossible for the author to effectively waive his copyright, the authors who want to freely spread the use of their work will rely on open licences.

The most popular types of open licences are the licences designed by Creative Commons. Creative Commons is a non-profit organisation that enables the sharing and use of creativity and knowledge through free legal tools.[422] Creative Commons licences are based on standardised licensing terms embedded in open licence agreements. The popularity of these licences mainly results from the fact that their creators were able to describe the terms and conditions of the licences in a graphic, easily accessible form. Because this system is simple and easy to use, it is frequently recommended, especially in relations between the public and private sector.

The importance of open licences in the context of big data projects comes from the fact that the use of such licences significantly improves the possibility to allow access to data. Not only does it facilitate conducting the clearance of rights, but it also removes the necessity of negotiating the conditions of the agreement with the copyright holder. This being said, the associated licences do not address the issue of data ownership (see also Section 3.10 on data ownership).

3.7.1.6 Conclusion on copyright in a big data environment

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

[422] More information on the Creative Commons licenses can be found on the following website: <https://creativecommons.org/> accessed 22 October 2018

understanding of these protection requirements facilitates extending protection to different types of data.

The long duration of copyright protection secures the possibility for the author to compensate the investment and effort put in the creation of the work.

The copyright holder is granted several exclusive rights that allow controlling the protected work's use and facilitate enforcement in case a third party uses the work without an authorisation. The reproduction, communication to the public and distribution right are indeed a useful toolkit which, balanced by the copyright exceptions, allows for an optimal protection of right holder's interests.

Copyright law 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.

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, this may mean identifying authors of hundreds (if not hundreds of thousands) 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.

Even if the originality threshold of works is relatively low, some of the data used in the context of big data projects will not be considered original. It can thus not be assumed that all of the data will benefit from copyright protection.

Moving to more general characteristics of copyright, it is important to stress that since the legal framework for copyright does not provide for a registration system, 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.

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.

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.

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

3.7.2 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. In this Section, we will therefore discuss the protection awarded to databases.

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. The protection of the latter elements is analysed in the sub-Sections below.

Finally, it shall be borne in mind that 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 the Software Directive.[423] The Directive also guarantees the right to create interoperable products[424], which is particularly important in the context of big data projects. Having said that, this Deliverable does not aim to discuss the issues related to the protection of computer programs.

3.7.2.1 Legal Framework

Similarly to copyright, the rules governing database protection have been established at international, regional, and national level. While international law provides only some underlying principles for database protection, the actual measures have been harmonised at EU level and implemented into national laws.

3.7.2.1.1 International Legal Framework

The protection of databases is anchored in international agreements.

Firstly, the Berne Convention explicitly provides in Article 5(2) related to protected works that:

> *"collections of literary or artistic works such as encyclopaedias and anthologies which, by reason of the selection and arrangement of their contents, constitute intellectual creations, shall be protected as such, without prejudice to the copyright in each of the works forming part of such collections".*[425]

Secondly, the TRIPS Agreement and the WIPO Copyright Treaty extend the database protection to compilations of data or other material which by reason of the selection

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

[424] According to Article 6 of the Software Directive on decompilation and under conditions specified therein, the authorisation of the right holder is not required where reproduction of the code and translation of its form are indispensable to obtain the information necessary to achieve the interoperability of an independently created computer program with other programs.

[425] The 1908 Berlin Act of the Convention introduced the protection of collections, which were protected as a category of "derivative works", and were mentioned (in Article 2(2)) along with translations, adaptations, etc. The 1948 Brussels revision conference transferred such protection to a separate paragraph.

or arrangement of their contents constitute intellectual creations.[426] Since the wording refers to "compilations of data or other material", protection of databases which do not contain copyrightable elements is also allowed. Consequently, both the TRIPS Agreement and the WIPO Copyright Treaty stipulate that database protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation.[427]

3.7.2.1.2 European Union Legal Framework

EU law provides for a specific protection of databases, which goes beyond other international legal instruments. The already mentioned Database Directive[428] was adopted with the objective of harmonising the protection of databases in all Member States:

- Its adoption was driven by the need to secure the investment made in the creation of databases and to create a level playing field between the creators and the makers of databases.
- It applies to both electronic and non-electronic databases, while it however excludes computer programs and moral rights from its scope.
- It establishes in substance a dual system of protection of databases (see sub-Sections 3.7.2.3 and 3.7.2.4 below for further details).

The European Commission has published two evaluations of the protection granted by the Database Directive. The first of these evaluations was carried out in 2005. The second was published on 25 April 2018.[429] From 24 May until 30 August 2017, the European Commission had held a public consultation in light of this second evaluation. The main purpose of the second evaluation was to assess the Database Directive, and in particular the *sui generis* right, in terms of (i) effectiveness; (ii) efficiency; (iii) relevance; (iv) coherence; and (v) EU-added value; notably in order to determine whether the Directive remains fit for purpose in the new legal, economic, and technological environment.[430]

The main findings of the second evaluation can be found in the table below.[431]

[426] TRIPS Agreement, art 10(2); WIPO Copyright Treaty, art 5

[427] Also note that the WIPO Diplomatic Conference on certain Copyright and Neighboring Rights Questions held in December 1996 had among its document a *Basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in Respect of Databases* (available at www.wipo.int/edocs/mdocs/diplconf/en/crnr_dc/crnr_dc_6.pdf) to be considered by the Diplomatic Conference. Although agreement was not reached, the Conference adopted a *Recommendation Concerning Databases* (available at www.wipo.int/edocs/mdocs/diplconf/en/crnr_dc/crnr_dc_100.pdf).

[428] 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)

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

[430] Ibid

[431] Commission, 'Executive Summary of the Evaluation of Directive 96/9/EC on the legal protection of databases' (Staff Working Document) SWD (2018) 147 final

Effectiveness	• The Database Directive has effectively harmonised the database protection regimes of the Member States. • The *sui generis* right has no proven impact on the overall production of databases or the competitiveness of the database industry in the EU. • The limited scope of protection allows striking an appropriate balance between the rights and interests of database producers and users.
Efficiency	Both costs and benefits of the *sui generis* right are moderate, though benefits seem to be higher.
Relevance	The Database Directive remains relevant in light of the Digital Single Market as it restricts regulatory fragmentation between Member States.
Coherence	No major inconsistencies exist between the Database Directive and other EU legislative instruments. Its interaction with the PSI Directive should however be clarified (see also Section 3.8 on Open data).
EU-added value	In light of the Digital Single Market, the Directive's main added value is the harmonisation of the main legal principles applicable to databases across the EU.

Table 24: Main findings of the second evaluation of the Database Directive

The Commission therefore concluded that a reform of the Database Directive, and notably the *sui generis* right, would be undesirable at this stage. Nevertheless, the Commission stressed the need to closely monitor the application of the *sui generis* right in the EU data economy.

3.7.2.2 General Principles of Database Protection in the EU

3.7.2.2.1 Definition of Database

The definition of database provided in Article 1(2) of the Database Directive is rather broad[432] and provides that a database should be understood as a collection of independent works, data or other materials which are:

- arranged in a systematic or methodical way; and
- individually accessible by electronic or other means.[433]

The CJEU further clarified the requirements of independence and individual accessibility providing that the term database, as defined in the Database Directive, refers to any collection of works, data or other materials, that are separable from one another without the

[432] According to the CJEU, a wide scope is offered to the concept of "database" for the purposes of the Database Directive (Case C-30/14 *Ryanair Ltd. v PR Aviation BV* [2015] ECLI:EU:C:2015:10, para 33 (Ryanair)).
[433] Database Directive, art 1(2). Also, Recital 17 of the Preamble stipulates that "*the term 'database' should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data (…)*."

value of their contents being affected, including a method or system of some sort for the retrieval of each of its constituent materials.[434] Having said that, a database does not need to be created for the purpose of retrieving individual elements of information in order to be protected.

Staying in line with the obligations deriving from international instruments, the database definition encompasses databases that include copyrighted and non-copyrighted elements, and databases in an electronic and non-electronic format.[435] Also, a database will be protected by itself (if it fulfils the conditions of protection) without affecting the rights of third parties to the individual pieces of information, which are contained in the database.[436]

The definition of "database" provided in the Database Directive is deliberately broad in scope, but it is not open-ended and in some instances, it has been subject to judicial scrutiny.[437] On the one hand, national courts have excluded from protection random collections of independent data[438], tourist bus routes[439], standard contract forms[440], computer programs used in the operation of a database[441], a system of indexation for pharmaceutical products[442], or an algorithm for sport betting and lottery games[443]. On the other hand, national courts recognised protection of *inter alia* telephone directories, collections of legal material, real estate information websites, radio and television guides, bibliographies, encyclopaedia, address lists, company registries, exhibition catalogues, tourism websites, collections of hyperlinks, and hit parades.[444] In conclusion, to verify the scope of the database definition for a particular dataset, one needs to examine national case law.

3.7.2.2.2 Types of Protection

Databases, within the meaning of the Database Directive, are protected in the EU by copyright (Chapter II of the Database Directive, see below), where such copyright protection echoes the one recognised in the international treaties, and a *sui generis* right (Chapter III of the Database Directive, see below). 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 are independent, and can be applied separately. They will however apply cumulatively if the conditions for both regimes are simultaneously met.

---

[434] Case C-444/02 *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE* [2004] ECLI:EU:C:2004:697, para 32

[435] Article 1(1) stipulates that the Database Directive "*concerns the legal protection of databases in any form*"; according to Recital 15 protection under the Database Directive should be extended to cover non-electronic databases.

[436] Article 3(2) of Database Directive provides that "*the copyright protection of databases provided for by this Directive shall not extend to their contents and shall be without prejudice to any rights subsisting in those contents themselves*".

[437] Maurizio Borghi and Stravoula Karapapa, 'Contractual Restrictions on Lawful Use of Information: Sole-source Databases Protected by the Back Door?' (2015) 37(8) EIPR 505

[438] *MIDI-Files,* Landgericht München I, Germany, March 30, 2000

[439] Court of Appeal of Brussels (8th Ch.), Belgium, June 5, 2007

[440] *Schutzfähigkeit von Musterverträger,* Landgericht Stuttgart, Germany, March 6, 2008

[441] Haju County Court, Estonia

[442] *Symposium Terapeutico,* Court of Appeal of Lisbon, Portugal, December 16, 2008

[443] *Z.S. v Sportbetting,* Supreme Court of Croatia, November 17, 2010

[444] Paul Goldstein and Bernt Hugenholtz, *International Copyright. Principles, Law, and Practice* (3rd edition, Oxford University Press 2013) 242

3.7.2.3 Copyright Protection for Databases

3.7.2.3.1 Protection Requirements

According to the Database Directive, 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"; no other criteria shall be applied to determine the eligibility of databases for that protection.[445]

The CJEU clarified that the concept of the "author's own intellectual creation" refers to the copyright law criterion of originality (described in sub-Section 3.7.1.2.1 above). Applied to databases, the criterion of originality is satisfied when, through the selection or arrangement of the data which it contains, its author expresses his creative ability in an original manner by making free and creative choices and thus adds his "personal touch".[446]

More specifically, the following principles apply according to the CJEU guidance:[447]

- the intellectual effort and skill of creating the underlying data are not relevant in order to assess the eligibility of the database for protection by copyright;
- the originality criterion is not satisfied when the setting up of the database is dictated by technical considerations, rules or constraints which leave no room for creative freedom;
- it is also irrelevant, for that purpose, whether or not the selection or arrangement of the underlying data includes the addition of important significance to that data; and
- the significant labour and skill required for setting up the database cannot as such justify copyright protection if in the end they have not conferred any originality in the selection or arrangement of the underlying data.

Consequently, 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 a protection of datasets is particularly relevant, the protection of the database structure has become less relevant and more difficult when confronted to new types of databases, unforeseen by the Database Directive.

---

[445] Database Directive, art 3(1)

[446] Case C-604/10 *Football Dataco Ltd and others v Yahoo! UK Ltd and others* [2012] ECLI:EU:C:2012:115, para 38; Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, para 45

[447] Case C-604/10 *Football Dataco Ltd and others v Yahoo! UK Ltd and others* [2012] ECLI:EU:C:2012:115, operative part

3.7.2.3.2 Ownership of Rights

The copyright database protection is generally granted to the creator (author) of the database. More precisely, Article 4 of the Database Directive ('database authorship') provides for the following explicit rules:

- the author of a database shall be the natural person or group of natural persons who created the database or, where the legislation of the Member States so permits, the legal person designated as the right holder by that legislation;
- where collective works are recognised by the legislation of a Member State, the economic rights shall be owned by the person holding the copyright;
- in respect of a database created by a group of natural persons jointly, the exclusive rights shall be owned jointly.

The above principles are generally reflected in the national laws on the copyright database protection in the Member States.

As with traditional copyright, the question is generally much more complex when considering works made within an employment relationship.

Belgium

Belgian copyright law contains a peculiarity with regard to works of employees. In contrast with other ordinary copyright works, the copyright on a database created by employees in the course of their employment contract will directly and exclusively belong to their employer, unless otherwise agreed upon. For databases created in the course of an employment (or service) contract, economic rights will be therefore directly held by the employer. Such presumption is however rebuttable, and concerns only the author's economic rights. It does not concern databases created in the cultural industry. Collective agreements (at the level of the enterprise or at the level of a sector, for instance) may determine the scope and practical arrangements of such presumption.[448]

By contrast, other Member States apply to databases the same rules as related to copyright in general.

France

In France, an employment contract does not have an incidence on the ownership of copyright on a database. The individual author remains the sole owner of his creation. Employers will thus need to conclude specific agreements granting rights except when the database was created under the regime of collective works, pursuant to Article L. 113-2 of the French Code of Intellectual Property.

[448] Alain Berenboom, *Le nouveau droit d'auteur et les droits voisins* (4th edition, Larcier, Brussels 2008) 297

Germany

Similarly, in Germany, when the database was created as part of the fulfilment of obligations resulting from an employment or service relationship, the provisions of the sub-section relating to the allocation of the exploitation rights of the German Copyright Act apply unless otherwise provided in accordance with the terms or nature of the employment or service relationship. If the database is established under a contract to produce a work, it must be ensured through contractual arrangements, that the necessary rights to use the database are granted.

United Kingdom

Finally, the usual rules in the United Kingdom regarding ownership also apply to databases, including the rule that an employer shall be deemed to be the owner of a database created by an employee during the course of his employment, and that a database can be jointly owned where the contribution of each author is indistinguishable from the contribution of co-authors.

3.7.2.3.3 Copyright on Database

The prerogatives awarded to the database's author echo the exclusive rights for "traditional" copyright. They are spelled out in Article 5 of the Database Directive, listing the following so-called 'restricted acts'[449] that require authorisation by the right holder:

- temporary or permanent reproduction of the database by any means and in any form, in whole or in part;
- any form of distribution to the public of the database or of copies thereof;[450]
- any communication, display or performance to the public;
- translation, adaptation, arrangement and any other alteration of a database, as well as any reproduction, distribution, communication, display or performance to the public of the results of these acts.

The moral rights of the natural person who created the database belong to the author and should be exercised according to the legislation of the Member States and the obligations

[449] Acts that cannot be performed without author's consent
[450] The first sale in the Community of a copy of the database by the right holder or with his consent shall exhaust the right to control resale of that copy within the Community.

resulting from the Berne Convention. However, they remain outside of the scope of the Database Directive.[451]

3.7.2.3.4 Exceptions

Exceptions to the exclusive rights of database authors are listed in Article 6 of the Database Directive.

Firstly, the Directive states that the authorisation of the right holder should not be required for a "lawful user" of a database to copy the database, or perform any other 'restricted act' which is necessary for the purposes of access to the contents of the databases and normal use of the contents.

Secondly, in addition to the exceptions traditionally authorised under general copyright law (see sub-Section 3.7.1.4.2 above), Member States have the possibility to provide for exceptions in case of:

- reproduction for private purposes of a non-electronic database;
- where there is use for the sole purpose of illustration for teaching or scientific research; and
- where there is use for the purposes of public security or for the purposes of an administrative or judicial procedure.

The three-step test also applies when relying on exceptions under the database legislation[452] (see sub-Section 3.7.1.4.2 above).

3.7.2.4 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* right[453], 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.

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.[454]

---

[451] Database Directive, Recital 28

[452] Database Directive, art 6(3)

[453] The term "*sui generis* right" is a generic one and means “the right of its own kind”.

[454] 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*".

In practice, should such protection be applied in a big data context, this could result in providing an indefinite protection, given that the databases are usually dynamic, hence, leading in all likelihood to "substantial changes to the contents of the database".

3.7.2.4.1 Protection Requirements

In order for a database maker to benefit from the *sui generis* protection it shall demonstrate that an investment was made to obtain, verify or present the contents of the database.[455] The CJEU has had the opportunity to provide guidance on the scope and conditions of the aforementioned terms in several cases (see in particular *British Horseracing Board*[456] and *Fixtures Marketing I* to *III* cases[457], all relating to databases of sport information). Said terms should therefore be interpreted as follows:

- **Obtaining**: "*The expression 'investment in…obtaining…of the contents' of a database in Article 7(1) of the directive must be understood to refer to the resources used to seek out existing independent materials and collect them in the database. It does not cover the resources used for the creation of materials which make up the contents of a database.*"[458]
- **Verification**: "*The expression 'investment in…the…verification…of contents' of a database in Article 7(1) of the directive must be understood to refer to the resources used, with a view to ensuring the reliability of the information, contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation.*"[459]
- **Presentation**: "*The expression 'investment in … the … presentation of the contents' of the database concerns, for its part, the resources used for the purpose of giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in that database and the organisation of their individual accessibility.*"[460]

In addition, such investment needs to be qualitatively and/or quantitatively '**substantial**'. It may consist in the deployment of financial resources and/or the expending of time, effort and energy.[461] Thus far, there are no specific CJEU judgments that would provide an indication of the threshold that should be met in order for an investment to be 'substantial'. The Member States courts however appear to grant protection to relatively low investment thresholds.[462]

---

[455] We note *en passant* that the French Code of Intellectual Property does not transpose the notion of "obtaining". Article L 341-1 indeed stipulates the following: *"Le producteur d'une base de données, entendu comme la personne qui prend l'initiative et le risque des investissements correspondants, bénéficie d'une protection du contenu de la base lorsque la constitution, la vérification ou la présentation de celui-ci atteste d'un investissement financier, matériel ou humain substantiel".*

[456] Case C-203/02 *British Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695

[457] 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

[458] Case C-203/02 *British Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 42

[459] Ibid

[460] Case C-444/02 *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE* [2004] ECLI:EU:C:2004:697, para 43

[461] Database Directive, Recital 40 of the Preamble

[462] SWD (2018) 146 final

Italy	France	Germany	Greece
According to the Court of Rome, inserting 4,000 records into a database and subsequently revising, adapting, and homogenising them satisfies the substantial investment threshold.[463]	The French Court of Cassation recognised the *sui generis* protection of an enriched telephone directory because of a substantial investment.[464]	The German Federal Court of Justice has accepted the *sui generis* protection of telephone directories.[465] Also, a 4,000 euro investment has been considered as sufficiently substantial.[466]	The Greek Supreme Court held that a map was protected by the *sui generis* right on the basis of a substantial investment in the collection of the information.[467]

Table 25: Database protection thresholds in selected EU Member States

3.7.2.4.2 Application to the Data Economy

As discussed in sub-Section 3.7.2.4.1 above, 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 *British Horseracing Board* and *Fixtures* 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.[468]

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 Internet of Things devices, big data, and artificial intelligence; as it becomes increasingly difficult to distinguish between the generation and the obtainment of data in the context of such processes.[469]

[463] Court of Rome, 10 December 2009 (see Annex 6 of the Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases carried out for the European Commission by JIIP, TNO and Technopolis Group).
[464] Cass. (Chambre commerciale, financière et économique), 23 March 2010, (2010) 225 RIDA373
[465] BGH, 6 May 1999, I ZR 199/96
[466] BGH, 1 December 2010, I ZR 196/08
[467] Supreme Court, Decision No. 1051/2015, HCO Website
[468] 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
[469] SWD (2018) 146 final

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/quantitative) in the obtaining, verification or presentation of the content, independent from the resources used to create the content.[470]

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.

IP in a big data environment in the transport sector – Example 2

In 2010, the German Federal Court of Justice held in its *Autobahnmaut* decision[471] 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.

3.7.2.4.3 Ownership of Rights

The *sui generis* right attaches to the maker (the producer) of a database, i.e., the person who takes the initiative and bears the risk of the investments that are at the origin of the database's creation.

[470] Case C-203/02 *Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 35
[471] BGH, 25 March 2010, I ZR 47/08

Such approach excludes the possibility to grant *sui generis* protection to subcontractors[472]: if the work is subcontracted, it is the commissioner of the sub-contracted work that will be granted the *sui generis* protection.

Such narrow view may be particularly problematic in a big data context which includes numerous actors. Indeed, this poses either the issue of assigning rights to a sole actor of the data value cycle, or granting exclusive rights to a multitude of persons.

3.7.2.4.4 Sui Generis Rights to Database

The maker of a database is granted in substance two (exclusive) economic rights in relation to the *sui generis* protection, i.e. the right to prevent extraction and reutilisation of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database. The contents of these rights are rather similar to the economic rights of an author in the copyright context (respectively to the reproduction right and to the right of communication to the public).

The extraction is defined as *"the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form*".[473] The CJEU has held that this concept needs to be interpreted broadly, as encompassing any unauthorised act of appropriation (via a physical copy or not) of the whole or part of the contents of a database.[474] Neither the purpose of this extraction (commercial or non-commercial) nor the technique of extraction (copying by hand or electronically) is of relevance in this regard.[475]

The right of reutilisation is defined as *"any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission"*.[476] This means, for instance, that incorporating the data from a database into a catalogue or a website without permission from the right holder amounts to a 'reutilisation' (this resembles the notion of communication to the public under the InfoSoc Directive).[477]

These two exclusive rights are limited to the extraction and reutilisation of '*substantial*' parts of databases. In this regard, *'substantial'* can mean both qualitatively substantial (a small part

[472] Database Directive, Recital 41 of the Preamble
[473] Database Directive, art 7(2)
[474] Case C-203/02 *Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 51
[475] Regarding the notion of "extraction", see particularly Case C-304/07 *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg* [2008] ECLI:EU:C:2008:552, para 36: "*The decisive criterion in this respect is to be found in the existence of an act of 'transfer' of all or part of the contents of the database concerned to another medium, whether of the same nature as the medium of that database or of a different nature. Such a transfer implies that all or part of the contents of a database is to be found in a medium other than that of the original database*".
[476] Database Directive, art 7(2)
[477] The CJEU has had the opportunity of clarifying such notion in Case C-203/02 *British Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695; Case C-173/11 *Football Dataco and others (*Football Dataco II) [2012] ECLI:EU:C:2012:642; and in Case C-202/12 *Innoweb BV v Wegener ICT Media BV and others* [2013] ECLI:EU:C:2013:850. See particularly *Innoweb*, para 37: *In the light of that purpose, the concept of 're-utilisation' as used in Article 7 of Directive 96/9 must be construed as referring to any act of making available to the public, without the consent of the database maker, the results of his investment, thus depriving him of revenue which should have enabled him to redeem the cost of the investment(…).*"

of the database that represents a substantial part of the investment[478]) or quantitatively substantial (a large part of the database).

According to the Database Directive, taking parts of the database that are *'insubstantial'* does not amount to an infringement, unless (i) it occurs repeatedly and systematically; and (ii) it conflicts with a normal exploitation of that database or unreasonably prejudices the legitimate interests of the maker of the database.[479]

In any event, if the database maker renders the contents of its database (or a part of it) accessible to the public, its *sui generis* protection does not allow it to prevent third parties from consulting that database.[480]

3.7.2.4.5 Rights and Obligations of "Lawful Users"

In the context of the *sui generis* protection, Article 8 of the Database Directive introduces the concept of *'lawful users'*. Such users are granted specific privileges (rights); in particular, the database's producer may not prevent them from extracting and/or re-utilising insubstantial parts of the database contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever (where such authorisation is granted, it only applies to such insubstantial part of the database).

However, a lawful user may not (i) perform acts which conflict with the normal exploitation of the database or unreasonably harm the legitimate interests of the maker of the database; nor (ii) cause prejudice to the holder of a copyright or related right in respect of the works or subject matter contained in the database.

The concept of "lawful users" has been implemented differently across the Member States.

[478] Case C-203/02 *British Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 78: (…) *the intrinsic value of the data affected by the act of extraction and/or re-utilisation does not constitute a relevant criterion for assessing whether the part in question is substantial, evaluated qualitatively* (…). *In other words, it is the value of the investment which must be taken into account*.

[479] Database Directive, art 7(5)

[480] Case C-203/02 *British Horseracing Board Ltd and others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 55; C-304/07 *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg* [2008] ECLI:EU:C:2008:552, para 51

 **Belgium**	 **France**	 **Germany**
The law refers to lawful users (*"utilisateurs légitimes"* – Articles XI.311 et seq. of the Belgian Code of Economic Law).	French law refers to the person who has lawful access (*"la personne qui y a licitement accès"* – Article L342-3 of the French Code of Intellectual Property).	The legislator has deliberately not used the term of lawful user like in the Directive, so that all consumer groups are recognised.[481]

Table 26: The concept of "lawful users" in Belgium, France and Germany

3.7.2.4.6 Exceptions to the Sui Generis Protection

The Database Directive proposes exceptions that Member States may transpose under their national laws. These exceptions cover:

- cases of extraction for private purposes of the contents of a non-electronic database;
- cases of extraction for the purposes of illustration for teaching or scientific research; as well as
- cases of extraction and/or re-utilisation for the purposes of public security or an administrative or judicial procedure.

The exception related to "scientific research" may prove interesting, to a certain extent only, in the context of big data projects. Under Article 9(b) of the Database Directive:

> *"Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents: (…) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved (…)."*

Member States are provided with the possibility of limiting such exception to certain categories of teaching or scientific research institutions.[482] As a result, some Member States have not implemented such exception, and those that have, have done so in diverging ways and notably by providing additional conditions.

[481] Bundestag-Drs. 13/7934, 54. Furthermore, according to Article 87e of the German Copyright Act a contractual agreement by which, inter alia, the owner of a lawful copy of the database undertakes vis-à-vis the producer of the database to refrain from reproducing, distributing or communicating to the public quantitatively or qualitatively insubstantial parts of the database shall be ineffective insofar as these acts neither run counter to any normal utilisation of the database nor unreasonably impair the legitimate interests of the producer of the database.

[482] Database Directive, Recital 51 of the Preamble

 **Belgium**	 **France**	 **Germany**
Belgian law requires that the name of the database maker and the title of the database are mentioned, and thus not only the source (Article XI.310 of the Belgian Code of Economic Law).	The legislator has adopted a rather restrictive approach: it excludes from the benefit of the exception some databases [483] and certain uses[484] and it limits the list of beneficiaries. [485] Also, the user must pay a compensation. [486] French law also does not specify that the research must be scientific. (Article L342-3 of the French Code of Intellectual Property).	Article 87(C) of the German Copyright Act refers to "personal scientific use".

Table 27: Overview of the differences of implementation of the scientific research exception

3.7.2.5 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.

[483] Databases created for educational purposes and databases created for a digital written edition.
[484] *"To the exclusion of entertainment or recreational activity".*
[485] *"So far as the public to whom the extraction and the re-utilization are intended is mainly composed of pupils, students, teachers or researchers directly involved".*
[486] *"The use of the extraction […] is compensated by a remuneration negotiated on a lump sum basis".*

In consequence, the ownership of rights to a database does not confer the rights to the individual data as such.[487] 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 (see more in sub-Section 3.7.3 below).

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. The database protections do provide however an incomplete and unsatisfactory protection of the data as such in the event a third party possesses the datasets.

3.7.2.6 Contractual Aspects

Databases protected by copyright can be transferred or licensed under the same general rules as those described for works in sub-Section 3.7.1.5 above. The rights of lawful users should nevertheless be preserved. The latter is particularly important, since according to the Database Directive[488] any provisions that would limit the rights of lawful users would be considered null and void.

As for the *sui generis* right, the Database Directive[489] indicates that such right may be transferred, assigned or granted under contractual licence. Hence, in practice, it will normally be conveyed under a transfer or licence agreement. Also, in case of the *sui generis* protection, the prerogatives given to lawful users cannot be limited, and contractual clauses contrary to the Directive's provisions on the rights of lawful users[490] will be considered null and void.

Finally, the judgment of the CJEU of 15 January 2015 in the *Ryanair v. PR Aviation*[491] case is of significant importance in the context of databases and in particular in relation to contractual terms related to databases.

The facts were in substance as follows. PR Aviation operated a website on which consumers could search through the flight data of low-cost airlines companies, compare prices and, subject to the payment of a commission, book a flight. PR Aviation obtained the necessary data to respond to an individual query by automated means, inter alia, from a dataset linked to the Ryanair website, which was also accessible to consumers.[492]

The use of the Ryanair's website was subject to a "tick-the-box" acceptance of the website terms and conditions. Under said terms and conditions, (i) Ryanair is the exclusive seller of Ryanair flights; (ii) use of Ryanair's website is permitted only for certain defined private and non-commercial purposes; and (iii) use of automated systems or software to extract data from the Ryanair website for commercial purposes is prohibited.

---

[487] 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*."
[488] Database Directive, art 15
[489] Database Directive, art 7(3)
[490] Database Directive, art 8
[491] C-30/14 *Ryanair v PR Aviation BV* [2015] ECLI:EU:C:2015:10
[492] Ibid para 15

The case ended up before the Dutch Supreme Court, which referred a question to the CJEU as to whether the limits on contractual freedom as set out in Article 15 of the Database Directive[493] apply to databases which are not protected by database copyright or the *sui generis* right.

The CJEU held that where a dataset falls within the general definition of a 'database' under Article 1(2) of the Database Directive, but when at the same time the database does not qualify for protection as database copyright and/or the *sui generis* right, the provisions governing database copyright and the *sui generis* right do not apply to that database.

As a consequence, the provisions of the Database Directive on lawful use (Articles 6(1) and 8) and on the limits to contractual freedom (Article 15) do not apply to such databases. Accordingly, the author/producer of such a database has the right to lay down contractual provisions on the use of the database of its choosing, subject to compliance with any applicable national laws.[494]

Legal scholars have concluded that "*non-protected sole-source databases benefit from the full scope of contractual protection*"[495] and that "*national laws on unfair competition and contract have the ability to defeat the objectives of the Directive and to upset the efficiency of the internal market*".[496]

In result, in such circumstance the database authors/makers receive strong protection by having the lawful option to unilaterally exclude third parties from making free use of their databases' content.[497] The practical outcome of the interpretation given by the CJEU in the *Ryanair* judgment is that contractual terms may bring significant limitations to the possibility to use databases and that in certain cases, one may be better off not to benefit from the protections laid down under the Database Directive in order to be able to set strict contractual terms. Indeed, if, and to the extent that, a particular dataset does not qualify for protection by database copyright and/or the *sui generis* right under the Database Directive, the conclusion of the *Ryanair* case may have consequences on the contractual terms related to the use of said dataset. If one enjoys no intellectual property rights on its databases, in light of the *Ryanair* judgment, one would be entitled, at least from an intellectual property rights perspective, to be stricter in its contractual terms and to prohibit any kind of extraction or re-utilisation of its databases, and even of unsubstantial parts of said databases.

3.7.2.7 Conclusion on database rights in a big data environment

The protection established by the Database Directive is dual, and supplements the possible protection granted to the data as such. In particular, the *sui generis* right provides an interesting protection of the investment made in obtaining, verifying or presenting the

[493] Article 15 of the Database Directive on the 'Binding nature of certain provisions' provides that "*Any contractual provision contrary to Articles 6(1) and 8 shall be null and void.*"
[494] In this context, it shall be noted that the CJEU did not examine the validity or the enforceability of the contractual terms concerned.
[495] Maurizio Borghi and Stravoula Karapapa, 'Contractual Restrictions on Lawful Use of Information: Sole-source Databases Protected by the Back Door?' (2015) 37(8) EIPR 505, 513
[496] Poorna Mysoor, "Protecting the Unprotected Database" (2015) 131 LQR 556, 561
[497] Maurizio Borghi and Stravoula Karapapa, 'Contractual Restrictions on Lawful Use of Information: Sole-source Databases Protected by the Back Door?' (2015) 37(8) EIPR 505

contents of the database. However, as demonstrated above, such right may have reached its limits in the current data- and technology-rich landscape.

Another feature that should be considered interesting is the possibility for the *sui generis* protection to run anew in case a substantial change to the contents of a database amounts to a substantial new investment. This rule allows securing reward in the investment made, for instance, in updating or upgrading the data collected. However, as demonstrated above, in a big data context this may lead to having an unlimited protection in case of dynamic databases.

In addition, database protection secures the interests of the right holder even in situations where the database is made available to the public, and its content disclosed.

However, database protection does not award protection to individual data.

Also, the eligibility for protection needs to be evaluated on a case-by-case basis in order to verify whether the criterion of originality (copyright protection) or of the substantial investment (*sui generis* protection) has been fulfilled in case of a specific database.

In addition, and similarly to the traditional copyright protection, the copyright protection for databases entails the necessity to account for the following features:

- the eligibility for protection and its scope can only be confirmed *a posteriori* by a court, leading to lack of legal certainty in the meantime,
- the exclusivity of copyright protection does not allow to acquire a right to the same or similar database, even if the data were collected independently or on the basis of different sources,
- in case the moral rights cannot be transferred, it is necessary to:
    - guarantee that the author will not exercise these rights against the entity who acquired the economic rights, and
    - ensure that the entity who acquired the economic rights is entitled to modify the data and/or database as required.

In case of the *sui generis* protection of databases, the Database Directive prevents only extraction or re-utilisation of "substantial" parts of data, or "insubstantial" parts of data if they are repeatedly and systematically extracted, and not any individual data. Even if the database protection is complemented by the traditional copyright protection, it would still leave the non-original individual data unprotected.

Also, the level of protection ensured across the Member States, especially concerning the copyright to database, 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.

Focusing on the enforcement, in some cases, it might also be difficult to demonstrate that data used by third parties, identical to the ones forming the database content, were actually copied, and not created or collected by this third person in parallel. Without possessing sufficient evidence that the data was actually copied, the database maker may have serious difficulties in preventing the use of its database content.

3.7.3 Trade Secrets and Confidentiality

While the mechanisms described in sub-Sections 3.7.1 and 3.7.2 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.

3.7.3.1 Legal Framework

Similarly to databases, only general rules requiring protection of trade secrets have been embedded in international law. Specific measures implementing this protection have been laid down at national level across the Member States. The importance of trade secret protection is growing thanks to recent harmonisation efforts undertaken at EU level.

3.7.3.1.1 International Legal Framework

The first step towards general recognition of trade secret protection was ensured by the TRIPS Agreement, which introduced a definition of "undisclosed information".[498] Pursuant to that provision, information qualifies for protection if:

- 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;
- it has commercial value because it is secret; and
- it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.

More specific measures complementing this general principle were introduced in the national law of the contracting parties. Traditionally, relevant provisions were introduced within employment law (such as the obligation of an employee to preserve confidentiality of information relating to the business activity of his employer) and/or legislation providing measures against unfair competition.

3.7.3.1.2 European Union Legal Framework

Trade secret protection has been established in EU legislation relatively recently. In June 2016, the European Parliament and the Council adopted 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**").[499] The Directive aims to standardise the national laws of the Member States as regards the unlawful acquisition, disclosure and use of trade secrets.

[498] TRIPS Agreement, art 39

[499] 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

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 a form of exclusive intellectual property right).

The Member States had to implement the Trade Secrets Directive by 9 June 2018.

3.7.3.2 Possibility to Protect Data as Trade Secrets

Since the Member States had to adjust their national laws to the protection standards introduced in the Trade Secrets Directive, this sub-Section will further refer to its text in order to describe the scope of trade secret protection and to evaluate the possibility to rely on this protection in case of individual data.

3.7.3.2.1 Definition and Scope of Protection

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:

- 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;
- it has commercial value because it is secret;
- it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.[500]

Such definition follows the one provided in Article 39 of the TRIPS Agreement. Since the definition is anchored in an international agreement binding all Member States, it was already (to a certain extent) reflected in the national laws.

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 IPR. Therefore, trade secrets are often at the origin of patents (new inventions), copyright (a new novel or a song), trademarks (a new branded product), and designs (a design of a new car model). 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.[501] 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).

3.7.3.2.2 Who Owns Trade Secrets?

The Trade Secrets Directive grants protection to the benefit of a trade secret holder, i.e. any natural or legal person who is lawfully controlling a trade secret.

---

[500] Trade Secrets Directive, art 2

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

3.7.3.2.3 Rights Conferred

As such, a trade secret holder has no private or exclusive rights to its use. Trade secrets are thus different from intellectual property rights that are safeguarded through an exclusive right that is legally enforceable. Instead, the Directive provides rights enabling to oppose the misappropriation of trade secrets.

Consequently, the holder of a trade secret cannot prevent competitors from copying and using the same solutions – reverse engineering (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).[502]

Thus, the differentiation between lawful and unlawful acquisition, use, and disclosure of trade secrets are at the core of protection granted by the Trade Secrets Directive.

The acquisition of a trade secret will be considered lawful when the trade secret is obtained by any of the following means:

- independent discovery or creation;
- observation, study, disassembly, or testing of a product or object that has been made available to the public or that is lawfully in the possession of the acquirer of the information who is free from any legally valid duty to limit the acquisition of the trade secret;
- exercise of the right of workers or workers' representatives to information and consultation in accordance with the EU law and national laws and practices;
- any other practice conforming with honest commercial practices.

On the other hand, the acquisition of a trade secret without the consent of the trade secret holder will be considered unlawful, whenever carried out by:

- unauthorised access to, appropriation of, or copying of any documents, objects, materials, substances or electronic files, lawfully under the control of the trade secret holder, containing the trade secret or from which the trade secret can be deduced;
- any other conduct which, under the circumstances, is considered contrary to honest commercial practices.

As to the use or disclosure of a trade secret, it will be considered unlawful whenever carried out, without the consent of the trade secret holder, by a person who:

- acquired the trade secret unlawfully;
- is in breach of a confidentiality agreement or any other duty not to disclose the trade secret;
- is in breach of a contractual or any other duty to limit the use of the trade secret.

[502] 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

The acquisition, use or disclosure of a trade secret shall also be considered unlawful whenever a person knew or ought, under the circumstances, to have known that the trade secret had been obtained directly or indirectly from another person who was using or disclosing the trade secret unlawfully.

In addition, the production, offering or placing on the market of infringing goods, or the importation, export or storage of infringing goods for those purposes, shall also be considered an unlawful use of a trade secret where the person carrying out such activities knew, or ought, under the circumstances, to have known that the trade secret was used unlawfully.

The Directive also harmonises the civil means through which victims of trade secret misappropriation can seek protection, such as:

- stopping the unlawful use and further disclosure of misappropriated trade secrets;
- the removal from the market of goods that have been manufactured on the basis of a trade secret that has been illegally acquired;
- the right to compensation for the damages caused by the unlawful use or disclosure of the misappropriated trade secret.

3.7.3.2.4 Exceptions

The Trade Secrets Directive introduces four obligatory exceptions to the rights it confers. The Member States need to ensure that an application for the measures, procedures and remedies provided for in this Directive is dismissed where the alleged acquisition, use or disclosure of the trade secret was carried out in any of the following cases:

- for exercising the right to freedom of expression and information as set out in the EU Charter of Fundamental Rights, including respect for the freedom and pluralism of the media;
- for revealing misconduct, wrongdoing or illegal activity, provided that the respondent acted for the purpose of protecting the general public interest;
- disclosure by workers to their representatives as part of the legitimate exercise by those representatives of their functions in accordance with EU or national law, provided that such disclosure was necessary for that exercise;
- for the purpose of protecting a legitimate interest recognised by EU or national law.

3.7.3.2.5 Data Protected as Trade Secrets

The protection established for trade secrets will expand to every piece of information, as long as it fulfils the protection requirements (mentioned above). Recital 16 of the Trade Secrets Directive however clarifies that "*in the interest of innovation and to 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.[503]

[503] Trade Secrets Directive, Recital 16

It is important to note that data can be protected as trade secrets as long as they remain secret. 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 it follows from the foregoing that 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.

3.7.3.3 Confidentiality: Contractual Aspects

The transfer of confidential information should be contractually regulated. Specific provisions imposing obligations on the acquiring party to ensure that the information will remain confidential, should be stipulated.

Confidentiality of data is important not only as an integral part of trade secret protection, but also as a security measure used at different stages of cooperation between entities.

Since trade secret protection is based on securing confidentiality of information, it is required for the information to be subject to reasonable steps under the circumstances.

Such reasonable steps can be based on the following technical measures backed by organisational (legal and policy) measures:

Technical measures	Legal and policy measures
• Software limiting access to data, blocking the possibility to copy, send or otherwise disseminate	• Establishing in-house rules on data access and management
• Login & password to the computer used by an employee/other contractor	• Introducing policy on data disclosure to third parties
• Requirement of using access cards to enter the building/floor/data room	• Including confidentiality clauses in the employment and cooperation contracts
• Anonymisation and/or pseudonymisation of the data (see also Section 3.4 on anonymisation and pseudonymisation)	• Concluding non-disclosure agreements ("**NDA**") for the purpose of business negotiation, at an early phase of cooperation (if the parties engage in pre-contractual discussions, confidentiality clauses play an essential role in allowing for an open dialogue and give a measure of security in the event that the parties do not reach a final agreement)

Table 28: Reasonable measures to secure confidentiality of information

In any instance where a company or an individual needs to disclose information to its/his partner (in business relations, research collaboration etc.), it needs to make sure that the information will not be further transmitted. It is therefore necessary to put in place

appropriate technical and organisational measures to ensure a sufficient level of protection. Confidentiality therefore plays an important role at different points of the contracting process and throughout the life of any big data service.

Safeguard procedures should be implemented and complied with, and the necessary filters should be applied in order to avoid that data the publication of which would potentially be harmful to the commercial interest of any party concerned, is made publicly available.

This can however pose several practical issues in the framework of big data projects, as the several actors of the data value cycle should be compelled to keep confidential the valuable data used. This would lead to a multitude of agreements with different actors and, most likely, with different terms and conditions. Such myriad of contracts would render the practical implementation of any big data project extremely burdensome and risky for the preservation of trade secrets.

A popular measure to ensure that an appropriate level of confidentiality is guaranteed is by concluding NDAs and including confidentiality clauses in the employment, cooperation and other types of similar agreements.

An NDA is a legal contract between at least two parties that outlines confidential material, knowledge, or information that the parties wish to share with one another for certain purposes, but wish to restrict access to or by third parties. It is a contract through which the parties agree not to disclose information covered by the agreement. An NDA creates a confidential relationship between the parties to protect any type of confidential and proprietary information or trade secrets. As such, an NDA protects non-public business information.

A confidentiality clause is a contractual provision imposing the obligation not to disclose certain information that a party acquires while performing its tasks under the contract. It is usually secured by means of contractual penalties.

3.7.3.4 Conclusion on trade secrets and confidentiality in a big data environment

In view of big data projects, trade secret protection 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.

Nevertheless, it requires 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.

3.7.4 Summary

As demonstrated by the above sub-Sections, all intellectual property rights examined may have, to some extent, an impact on the use of big data, including in the transport sector. Indeed, depending on the manner in which and the extent with which a right holder may exercise its exclusive rights attached to the intellectual property right concerned, IPR may pose a barrier to data access, interoperability, and exploitation.[504]

Opportunities in relation to intellectual property in the context of big data in the transport sector	Challenges in relation to intellectual property in the context of big data in the transport sector
If stakeholders in the big data analytics lifecycle are able to rely on intellectual property rights to protect their investment (be it monetary or intellectual) in (parts of) the datasets, they may be more eager to engage in data sharing in a big data analytics context, including in the transport sector.	Many different actors in the big data analytics lifecycle may try to claim intellectual property rights in (parts) of the datasets intended to be used and may therefore try to exercise the exclusive rights linked to the intellectual property right concerned. Any unreasonable exercise of rights may stifle data sharing and thus innovation through big data, including in the transport sector.
	Even in circumstances where no intellectual property rights subsist in individual data or in datasets, an assertion of (non-existent) IPR claims by economically stronger actors in the big data analytics lifecycle and/or in the field of transport against those with fewer resources may in practice have effects equivalent to the exercise of strong intellectual property rights. Such behaviour would thus constitute a barrier to the uptake of big data in the transport sector.

Table 29: Summary table of opportunities and challenges in relation to intellectual property in the context of big data in the transport sector

[504] Teresa Scassa, 'Public Transit Data Through an Intellectual Property Lens: Lessons About Open Data' (2014) 41(5) Fordham Urb. L.J. 1759

3.8 *Open data*

In many instances, big data analytics may rely on data made publicly available by public authorities. This may be the case for instance with weather data, data on sustainable energy sources, or public transport data. However, the making available, but also the use, of such datasets is subject to rules at both EU and national level. Those rules may restrict the possibilities of companies that wish to rely on such 'open data'.

3.8.1 Concept of open data and PSI

Pursuant to the 'Open Definition'[505], open data is "*data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike*".

Its most important aspects are:

Availability and access	The data must be available as a whole, at no more than a reasonable reproduction cost, preferably by downloading over the Internet. The data must also be available in a convenient and modifiable form.
Re-use and redistribution	The data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
Universal participation	Everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Table 30: Most important aspects of open data[506]

In the context of the Digital Single Market of the EU Commission, the concept of open data is defined as referring to "*the idea that certain data should be freely available for use and re-use*".[507] It should not be confused with the concept of big data. The latter refers to the size and complexity of datasets, whereas "open" indicates the accessibility and machine readability of and no (or low) cost of access and rights related to data.

Open data therefore increasingly refers to so-called open public sector information ("**PSI**"), i.e. material produced, collected, paid for and/or held by public sector bodies at national, regional and local levels, such as ministries, agencies, municipalities, but also organisations mainly funded by or under the control of a public authority.[508] Public entities generate and hold enormous amounts of data. Governments and public institutions have an inherent interest in

[505] Open Definition is a project of Open Knowledge International, available online at <http://opendefinition.org/>
[506] Open Definition 2.1, available online at <http://opendefinition.org/od/2.1/en/>
[507] European Commission, 'Open Data' (*European Commission*, 8 June 2018) <https://ec.europa.eu/digital-single-market/en/open-data> accessed 18 October 2018
[508] 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

managing this large amount of data carefully, both to improve their performance and to generate savings to their budget, but also to be able to provide open data to their citizens and business entities.

Access to and use of data is essential for any organisation developing or making use of big data applications. Not all private companies however have the resources to generate or collect the huge amounts of data required for big data analytics. As a result, they are forced to turn to other data sources and in this respect increasingly rely on open (public sector) data. The Open Data movement and governments around the world, including the EU, are committed to make data, and more particularly 'government data' or PSI publicly available and usable. The European Commission claims to support open data for various reasons, including for the re-use thereof in new products and services, to face societal challenges, to achieve efficiency gains, and to increase transparency.[509] Open data is moreover considered to be crucial for EU research and development.[510]

3.8.2 Non-legislative measures on open data

The EU institutions have taken both legislative and non-legislative measures to encourage the uptake of open data.

On the non-legislative front, the EU 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 Open Data incubator, a 6-month incubator for open data entrepreneurs across Europe;[511]
- funding the ePSI platform;
- funding the Legal Aspects of Public Sector Information (LAPSI) network;
- commissioning studies related to the re-use of public sector information;
- developing the EU Open Data Portal, which provides access to data from the EU institutions and bodies for re-use;[512] and
- contributing to the G8 Open Data Charter, which aims to open up government information.

[509] European Commission, 'Open Data' (*European Commission*, 8 June 2018) <https://ec.europa.eu/digital-single-market/en/open-data> accessed 18 October 2018
[510] Commission, 'Open Data. An Engine for Innovation, Growth and Transparent Governance' (Communication) COM (2011) 882 final, 4
[511] More information available online at <https://opendataincubator.eu/>
[512] Accessible online at <http://data.europa.eu/euodp/en/data>

3.8.3 Legislative measures on open data: Directive on the re-use of public sector information

3.8.3.1 PSI Directive of 2003

Already in 2003 the EU adopted its first set of rules on the re-use of public sector information, Directive 2003/98/EC (the "**PSI Directive**"). The aim of that Directive was not so much to make public data more accessible and to encourage the re-use of this information, as it was to ensure that when public sector bodies decided to make data available, they did so in a fair and non-discriminatory manner.[513]

To this end, various obligations were imposed on public sector bodies. These included among others that documents should be made available in pre-existing formats[514], applicable conditions and standard charges should be pre-established and published[515], and conditions for re-use could not unnecessarily restrict possibilities for re-use or be used to restrict competition.[516] On the issue of remuneration, the initial version of the PSI Directive still allowed public sector bodies to achieve a "*reasonable return on investment*".[517]

It should be noted that in its initial version, the PSI Directive was still primarily aimed at paper documents rather than electronic data, even though electronic data already fell within its scope of application. Moreover, public authorities had to comply with these requirements when they decided to make data available, but the making available of data as such had not been made mandatory.

3.8.3.2 *PSI Directive of 2013*

In 2013, the PSI Directive was given a thorough makeover. Directive 2013/37/EU ("**consolidated PSI Directive**") amended the PSI Directive in order to keep pace with technological developments, which had led to the rise of the data economy, and unlock the potential of big data held and accumulated by government authorities.

The European legislators introduced concepts such as "machine-readable format", "open format", and "formal open standard"[518] as part of an effort to facilitate the re-use of electronic data of different platforms and different formats. Public sector bodies are required to make data available, where possible and appropriate, in open and machine-readable formats together with their metadata.[519]

In a significant departure from the first PSI Directive, an obligation was introduced for public sector bodies to make public sector information available.[520] This effectively eliminated the possibility for public bodies to avoid application of the Directive by not making available information. Still, the Directive also includes a number of exceptions to the principle of

[513] PSI Directive, Recital 8
[514] PSI Directive, art 5
[515] PSI Directive, art 7
[516] PSI Directive, art 8(1)
[517] PSI Directive, art6
[518] PSI Directive, art 1(2)
[519] Consolidated PSI Directive, art 5(1)
[520] Consolidated PSI Directive, art 3 (1)

mandatory data provision. 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.[521]

The principle governing remuneration was also amended and the right of a "reasonable return on investment" was removed. In the amended PSI Directive, any fees charged must in principle be limited to "*the marginal costs incurred for their reproduction, provision and dissemination*".[522]

The Directive stipulates 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.[523] They must also make practical arrangements to facilitate the search for documents available for re-use.

More particularly regarding the licensing, 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.[524] In these Guidelines, the Commission recommends using the Creative Commons CC0 public domain dedication, as it "*allows waiving copyright and database rights on PSI, ensures full flexibility for re-users and reduces the complications associated with handling numerous licences*".[525]

The contents of the CC0 licence can be graphically summarised as follows:

	CC0
Commercial use	Yes
Non-commercial use	Yes
Modification	Yes
Distribution	Yes
Sub-licence	Yes
Responsibility for content	No
Attribution (credit)	No
Fees	No
Duration	Unlimited

Table 31: Graphic overview of Creative Commons licence CC0

In the event it is impossible to use the CC0 public domain dedication, the Guidelines encourage public sector bodies to use open standard licences appropriate to the Member

[521] Consolidated PSI Directive, art 2
[522] Consolidated PSI Directive, art 6(1)
[523] Consolidated PSI Directive, art 8
[524] Commission Notice Guidelines on recommended standard licences, datasets and charging for the reuse of documents [2014] OJ C 240/1
[525] Ibid 2

State's intellectual property and contract law, and which comply with the licensing provisions recommended in the Commission's Guidelines. The Guidelines further stimulate Member States to develop a suitable national open licence.[526]

The licensing provisions recommended by the Commission in its Guidelines cover the following topics:

- **Scope**: temporal and geographical scope of the rights covered by the licence, the types of rights granted and the range of re-use allowed;
- **Attribution** (only when simple notices are not possible): obligation on the re-user to acknowledge the source of the documents in a manner specified by the public sector body (licensor);
- **Exemptions**: indication of the datasets, if any, that are not covered by the licence;
- **Definitions**: concise definitions of the main terms of the licence;
- **Disclaimer of liability**: indication that the information is provided 'as is' and that the public sector body (licensor) assumes no responsibility for its correctness or completeness;
- **Consequences of non-compliance**: consequences of non-compliance with the licence terms;
- **Information on licence compatibility and versioning**: indication of the other licences the licence in question is compatible with.[527]

The above constitutes a framework for the minimum harmonisation of national rules and practices, through which the PSI Directive is meant to create a level playing field across the EU.[528]

In April 2018, the European Commission published a proposal for a revised PSI Directive. This proposal will be discussed in sub-Section 3.8.5 below.

3.8.4 Challenges and opportunities of sharing public sector information

The following sub-Sections offer an overview of various challenges and opportunities related to the sharing of public sector information.

3.8.4.1 Opportunities of sharing public sector information

Public sector information 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.[529] This sub-Section addresses some of the opportunities that arise from opening up PSI.

[526] Ibid
[527] Ibid 3-4
[528] Mireille Van Eechoud, 'Making Access to Government Data Work' (2015) 9(2) Masaryk University Journal of Law and Technology 61
[529] 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)

Lowering barriers to market entry

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. The non-discrimination 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.[530]

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.

Creating value for the economy

Perhaps the most important potential benefit of opening up PSI is the value that can be generated by commercialising this data. The fees charged by public sector bodies providing the data are a crucial factor in this respect. Studies have shown that reduced pricing of PSI (e.g. through zero or marginal cost charging) entails a significant increase in the use of PSI, being used directly or indirectly in different types of applications across the economy.[531] This may lead to increased revenues, job creation and increased innovation, and thus overall an increase in competition. Where, on the other hand, cost-plus fees are charged to access and re-use the PSI, the impact on the development of new products and services is likely to be significantly smaller.[532]

When it comes to innovating in the age of the data economy, start-ups and SMEs also play an important role. While many established companies have embraced open data, there is still a risk of these companies viewing open data as a threat to their existing business and revenue models. This is not the case for SMEs and start-ups, which are more prone to see data as an opportunity. As their business models are not yet set in stone, these companies are more

[530] 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

[531] Dinand Tinholt, 'The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data' (Capgemini Consulting 2013) 9 <https://www.capgemini.com/wp-content/uploads/2017/07/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf> accessed 18 October 2018

[532] Heli Koski, 'Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth?'(Keskusteluaiheita Discussion Papers N 1260, Etla, the Research Institute of the Finnish Economy, 2011) 6 <https://www.etla.fi/wp-content/uploads/2012/09/dp1260.pdf> accessed 18 October 2018

likely to experiment and display the flexibility required to maximize the economic potential of PSI.[533]

Additionally, re-use of public sector information does not take place in silos. Open data is often used in combination with other datasets (e.g. from other public bodies, private companies, or even private individuals). This combination is likely to lead to more innovation, thereby increasing the economic potential of each separate data source.[534] It shall be kept in mind however that the combination of datasets of different origin may collide with the several limitations discussed in this Deliverable (e.g. see Section 3.1 for the use of personal data or Section 3.7 for the use of IPR-protected data and databases).

Open data in the transport sector – Example 1

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. AIS transponders are mandatory on all passenger ships, ships above 300 gross tonnage, and fishing vessels of over 15 meters in length. 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.[535]

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.

An initial weekly forecast is already being presented. For week 36 of 2018, a communication mentioned a significant drop in the rivers' water level as a result of a continuing dry period. It announced that ships would be able to carry less cargo and therefore more ships would be needed to transport the same amount of freight. On the other hand, the communication also stated that due to the smaller cargo sizes, a faster turnaround time in the port could be expected.[536]

[533] 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

[534] Wendy Carrara, Wae San Chan, Sander Fischer and Eva van Steenbergen, 'Creating Value Through Open Data. Study on the Impact of Re-use of Public Data Resources' (European Commission 2015) 36-37 <https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf> accessed 18 October 2018

[535] 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

[536] Port of Rotterdam, 'Barge Performance Monitor' (*Port of Rotterdam*) <https://www.portofrotterdam.com/nl/zakendoen/logistiek/verbindingen/barge-performance-monitor> accessed 18 October 2018

Collaboration between public sector bodies

Not only private companies can benefit from open data policies with respect to public sector information, public sector bodies can equally benefit themselves. First, the use of data made available by one public body in an open data format can reduce searching and processing times for other public bodies. This results in an increased speed of decision-making and an overall improved efficiency in the provision of public services. Additionally, the administrative burden on data subjects may be reduced.[537] It can also reduce unnecessary public spending, as data will not need to be collected twice.[538]

Public-private sector collaboration and synergies

While public sector bodies play a key role with regard to data publication, even more opportunities can be realised when the public sector moves from being a mere data supplier and/or publisher to taking up a more proactive role in this respect.[539] This could for example result in collaborations with the private sector to maximise the economic potential of PSI.

In an initial stage of open data policy, these proactive efforts to engage and cooperate with the private sector could focus on identifying demand to prioritise data publication, increase user awareness, and help build capacities for data re-use. Beyond this first step, public sector bodies could however move to the implementation of public-private initiatives focusing on increasing re-use with a problem-solving approach.[540] Open data can for instance help with the optimisation of the fleet of vehicles owned and operated by the government. This has been done in a US city, where open data allowed the local government to understand the usage patterns of its vehicles and was able to cut the fleet by 30%. Also in the US, the state of California published budget data on public fleet spending. Citizen advocates spotted examples of unnecessary costs which resulted in the state reducing its fleet by 15%.[541]

[537] Law Commission, 'Data Sharing between Public Bodies' (Consultation Paper No. 214, Law Commission 2013) 5 <http://www.lawcom.gov.uk/app/uploads/2015/03/cp214_data-sharing.pdf> accessed 18 October 2018

[538] European Open Data Portal, 'Benefits of Open Data' (*European Open Data Portal*, 2018) <https://www.europeandataportal.eu/en/using-data/benefits-of-open-data> accessed 18 October 2018

[539] OECD, 'Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact' (OECD Digital Government Studies, OECD Publishing 2018) 206 <https://doi.org/10.1787/9789264305847-en> accessed 18 October 2018

[540] Ibid

[541] Michael Chui and Diana Farell, 'A Closer Look at Open Data: Opportunities for Impact' (2014) Innovation in local government: open data and information technology 24, 28-29

Open data in the transport sector – Example 2

Brussels' bike-sharing scheme, Villo!, is run through a public-private partnership with the French advertising company JCDecaux, who uses the bikes' stops as natural candidates for outdoor advertising space. When the scheme became more popular, it was increasingly plagued by frequent unavailability of either bikes or parking spaces. This made many of the stations near unusable, as the bikes were simply not reallocated often enough. Villo! Now makes real-time data available about the status of each station on their servers to help users find out which stations have bikes or parking spaces available. This is done through the use of open government data (the location of the bike stations) and adding crowd-sourcing for quality monitoring and feedback to the city council. "*By pulling data from Villo!'s servers every five minutes and writing them to a database, Where's My Villo? computes a daily list of the worst places to find a bike or park one. The website also allows visitors to see how well the station(s) they use most frequently perform, and to report their own encounters with problems (from finding a bike to technical problems at stations).*"[542]

3.8.4.2 *Challenges of sharing public sector information*

Many challenges remain with respect to optimal use of public sector information. Those challenges include a need for more real-time access to dynamic data[543], technical challenges related to standardisation and interoperability, and questions of how to ensure sufficient data quality and timely data publication. These challenges are outside the scope of this sub-Section, which will focus on some of the key legal issues that arise.

Licensing

The PSI Directive allows public sector bodies to make the re-use of data subject to conditions, notably through the use of licences. The only limitation in this respect is the fact that conditions may not "*unnecessarily restrict possibilities for re-use and shall not be used to restrict competition*".[544]

While 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.[545] Despite the guidelines on recommended standard licences adopted by the Commission in 2014[546], very little uniformity is seen in practice. Analysis shows that the EU Member States have embraced very different licensing practices. In some Member States, notably Poland, public authorities do not promote any model licence agreements.[547] In others,

[542] Jonathan Van Parijs, 'Open Data in Public Private Partnerships: How Citizens can Become True Watchdogs' (*Open Knowledge International Blog*, 29 October 2010) <http://blog.okfn.org/2010/10/29/open-data-in-public-private-partnerships-how-citizens-can-become-true-watchdogs/> accessed 18 October 2018

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

[544] Consolidated PSI Directive, art 8(1)

[545] Ibid art 8(2)

[546] See *supra* sub-Section 3.8.3.2.

[547] Wendy Carrara, Cosmina Radu and Heleen Vollers, 'Open Data Maturity in Europe 2017' (European Data Portal 2017) 26 <https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n3_2017.pdf> accessed 18 October 2018

like France[548] and the United Kingdom[549], standard licences are in force. In other Member States such as Belgium, a lack of unity even exists within the different levels of government. As a consequence, any company that wishes to reuse public sector information from different Member States to develop a product is obliged take into account as many (and perhaps even more) licences as the number of Member States in which it operates.

Privacy

In theory, the relationship between the PSI Directive and the GDPR evokes little questions. 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.[550] 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 public sector information to be taken into account when applying the GDPR.[551]

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 personal data part of PSI could lawfully be disclosed.

That assessment can be summarised as follows:

1. Determine whether the relevant public sector dataset contains personal data;
2. Determine whether:
    a. national access regimes exclude or restrict access to the public sector data on grounds of personal data protection; and/or whether
    b. the re-use of public sector data accessible through such national access regimes has been defined by law as being incompatible with personal data protection;
3. Ensure that the public sector dataset that contains personal data and which is disclosed is processed in accordance with data protection law.[552]

The first and third parts of that assessment however give rise to a number of additional challenges.

An important challenge stems from the broad interpretation of personal data, which has been addressed in Section 3.1 above but is repeated here for ease of reference: "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an*

[548] See<https://www.etalab.gouv.fr/licence-ouverte-open-licence>
[549] See <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/>
[550] Consolidated PSI Directive, arts 1(2)(cc) and 1(4)
[551] GDPR, Recital 154
[552] Consolidated PSI Directive, art 1(2)(cc)

identifier […]".[553] Additionally, and notably, Recital 26 of the GDPR states that to "*determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person*". So while in theory, the GDPR is not applicable to anonymous data[554], in practice it will be very hard for a dataset to be rendered truly anonymous. It cannot be excluded that a PSI dataset, following combination with other information from third-party sources, indirectly enables the identification of individuals. That would mean that it has always constituted personal data. For a more detailed analysis of the issues surrounding anonymisation, we refer to Section 3.4 of this Deliverable. For the purposes of this Section, it suffices to state that the first part of the assessment is not as straightforward as it might seem.

Public sector bodies will moreover encounter additional difficulties in the third part of the assessment, when attempting to ensure that a PSI dataset containing personal data is processed in accordance with the relevant data protection rules, most notably those laid down in the GDPR. For instance, the making available of 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 latter prohibits any processing of personal data that is not compatible with the specific purposes for which the data was initially collected. 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.[555]

It follows from the above 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.

Relationship with the Database Directive

Uncertainty also exists about the precise relationship between the PSI Directive and the Database Directive discussed in sub-Section 3.7.2. 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*".[556] It appears that this has been frequently relied upon by public bodies to exclude applicability of the PSI Directive to their information, with transport data being one of the categories of PSI for which this reasoning has been invoked.[557] 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. [558] This issue will be addressed further below.

[553] GDPR, art 4(1)
[554] GDPR, Recital 26
[555] Consolidated PSI Directive, art 11(1)
[556] Consolidated PSI Directive, art 1(2)(b)
[557] 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
[558] Ibid 6-7

3.8.5 Towards a new PSI Directive?

3.8.5.1 Proposal for a revision of the PSI Directive

On 25 April 2018, the EU Commission presented a proposal for a revision of the PSI Directive (the "**Recast Proposal**"). The most fundamental change relates to the Directive'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 included in 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 covering 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).[559] The Recast Proposal is thus to a large degree targeted at public undertakings in the transport sector at large.

The Recast Proposal limits the 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.[560] It will thus be important to consider whether a public undertaking has produced the requested information in the context of the provision of services of general interest.

However, even where the revised Directive would be applicable, the public undertaking in question could still decide whether or not to disclose the information or to keep it for itself 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 information available, it will have to respect the rules laid down in the Directive. Additionally, these undertakings would also not be required to comply with the (mainly procedural) requirements on the processing of requests for re-use.[561]

Taking account of the limitation in scope and the distinction between hard and soft requirements, the Recast Proposal would lead to a three-pronged regime which can be visually represented as follows:

[559] Recast Proposal, art 1(1)(b)
[560] Recast Proposal, art 2(1)(a)
[561] Recast Proposal, art 4

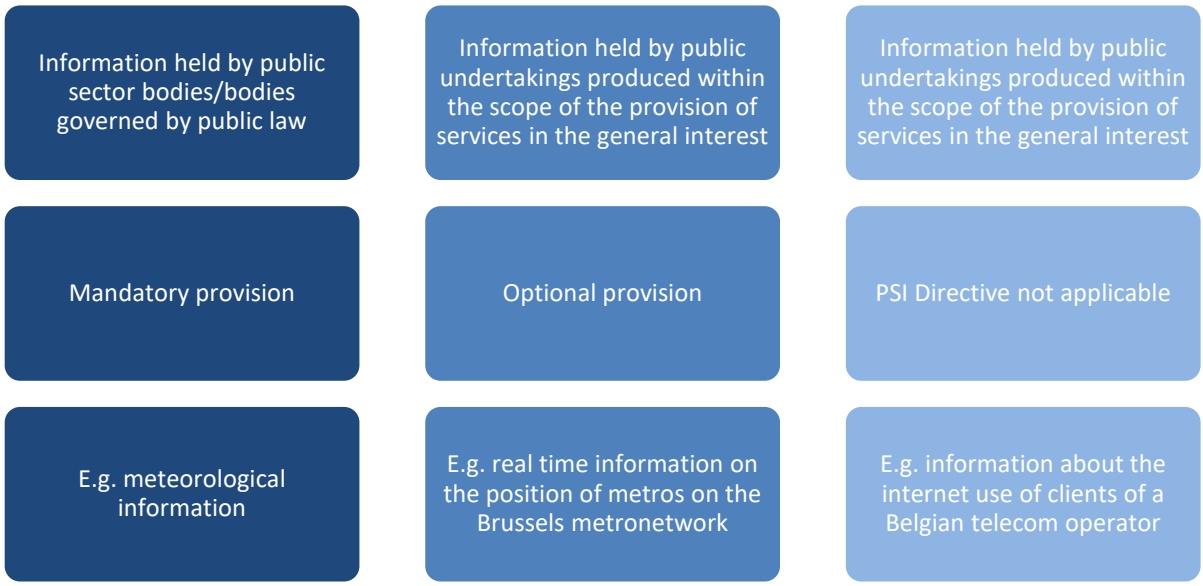


Figure 9: Three-pronged legal regime of the Recast Proposal for a revision of the PSI Directive.[562]

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 API.[563] The text also clarifies that costs related to data anonymisation[564] may be included in the fees charged to re-users.[565]

Open data in the transport sector – Example 3

In 2015, the German railway and infrastructure operator Deutsche Bahn, which is 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.[566]

[562] Thomas Verellen, 'Het voorstel tot herziening van de PSI-Richtlijn: Hoe open is open data?' Draft article submitted to the *Revue du droit des industries de réseau* (forthcoming)
[563] Recast Proposal, art 5(4)
[564] For more information on data anonymisation, please refer to Section 3.4.
[565] Recast Proposal, art 6(1)
[566] 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

3.8.5.2 Benefits and shortcomings of the Recast Proposal

Attention to the commercial interests of public undertakings

The Recast Proposal does not simply extend the compulsory regime of making data available to all information held by public undertakings but introduces an optional regime, applicable only to a particular category of information (i.e. that which falls within the scope of the public tasks of the public undertaking). It is therefore not blind to the fact that, unlike public sector bodies, public undertakings have commercial interests and are subject to competitive market forces. Data held by public undertakings may have significant commercial value or may even be considered a trade secret (for more information on trade secrets, please refer to sub-Section 3.7.3). If such an undertaking would be forced to share commercially valuable data, it would equally be forced to undermine its own competitive position on the market.[567] The PSI Directive's objective was however not to frustrate competition on the European markets but to strengthen it, notably by giving undertakings the opportunity to unlock already existing data produced by the public sector in the context of its public tasks and enable them to re-use that data, mainly for commercial purposes.

Legal uncertainty

The Recast Proposal also gives rise to a number of questions and uncertainties. First, one can wonder what the consequences will be of introducing the abovementioned 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 requirements on the processing of re-use requests were not made applicable to public undertakings.

Second, while it is commendable that the regime has been limited to data falling within the scope of the public tasks of a public undertaking, the rationale behind making such regime optional is unclear and creates legal uncertainty. This is reminiscent of the optional regime for public sector bodies introduced by the initial PSI Directive and begs the question whether this is only a first step towards a mandatory regime in the future.

List of high-value datasets

Another novelty in the Recast Proposal is the introduction of the category of so-called “high-value datasets”. These are datasets that are associated with important socio-economic benefits and the re-use of which should in principle be free of charge. The datasets are however not defined in the Recast Proposal itself, but would be adopted by the European Commission through a Delegated Act.[568]

Public undertakings fear that such future Delegated Acts could force them to make high-value datasets for free and would thereby significantly affect their competitive position on the market. Public undertakings would be put in an inferior position compared to private

[567] Thomas Verellen, 'Het voorstel tot herziening van de PSI-Richtlijn: Hoe open is open data?' *Revue du droit des industries de réseau*, 11 (forthcoming)
[568] Recast Proposal, art 13

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. [569]

Interplay between the Recast Proposal and other legal instruments

We already mentioned that certain public sector bodies assume that they have a *sui generis* right to their database(s). Consequently, they claim that these databases are excluded from the scope of the PSI Directive and impose conditions on the use of the database. In the public consultation preceding the Recast Proposal, public transport undertakings even insisted that these database rights over publicly funded databases were essential to the protection of their legitimate interests in a competitive market.[570]

While this practice seemed to be permitted under the current PSI Directive, the Recast Proposal aims to close this potential loophole. It clarifies that a public sector body which is the rightholder cannot invoke its *sui generis* right as a ground to prevent or restrict the re-use of the content of the database.[571] How exactly the terms 'prevent' and 'restrict' are to be understood is still unclear but in any case it has been made clear that the Database Directive does not offer public sector bodies the right to restrict re-use of public sector information.

Finally, the interplay between the Recast Proposal and the GDPR merits examination. As explained above, the Recast Proposal aims to include public undertakings in the scope of the PSI Directive but does not impose a mandatory regime on those undertakings. The Recast Proposal merely stipulates that *in the event* that a public undertaking decides to make PSI available, it must comply with the rules and requirements laid down by the PSI Directive.

In light hereof, it should be noted that the GDPR is likely to affect public sector bodies and public undertakings in a different manner. While public sector bodies are not subject to the regime of administrative fines of Article 83 GDPR (except when individual Member States declare the regime applicable), that is not the case for public undertakings which are in any case subject to the regime. The possibility of incurring a huge administrative fine under the GDPR may therefore have a deterrent effect on public undertakings to which only the optional PSI Directive regime applies. It can therefore not be excluded that these undertakings will err on the side of caution and decide not to disclose information rather than running the risk of infringing the GDPR.[572]

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

[570] European Commission, 'Consultation on PSI Directive Review, Synopsis Report' (European Commission 2018) 6-7 <https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-revision-directive-reuse-public-sector-information> accessed 18 October 2018

[571] Recast Proposal, art 1(5)

[572] European Data Portal, 'The PSI Directive and GDPR' (*European Data Portal*, 2018) <https://www.europeandataportal.eu/en/highlights/psi-directive-and-gdpr> accessed 18 October 2018

3.8.5.3 *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 as put forward in sub-Section 3.8.4.1, it seems that this broadened scope can only be applauded. The present sub-Section however aims to show that there are limits to the desirability of making available public sector data, through the example of essential services and critical infrastructure.

As explained in Section 3.2, the NIS Directive requires Member States to identify so-called operators of essential services. The latter are services that a Member State deems essential for the "*maintenance of critical societal and economic activities*".[573] Annex II of the NIS Directive lists the sectors and subsectors for which such operators must be identified. It covers 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.

By way of reminder, the NIS Directive essentially requires Member States to impose security and incident notification requirements on all operators of essential services in their territory. Public and private operators alike will need to implement technical and organisational measures to manage any risks to the security of the network and information systems used in their operations. With the objective of ensuring service continuity, they will moreover be obliged to adopt measures to prevent and minimise the impact of incidents affecting the security of their systems.[574]

Directive 2008/114/EC[575] (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.[576] Similarly to the NIS Directive, security requirements are introduced for such European critical infrastructures. Among others, Member States must 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, as in the other sectors identified in Annex II of the NIS Directive, many essential services operators are public undertakings. The essential services covered by the NIS Directive are moreover likely to constitute services provided in the

[573] NIS Directive, Recital 20
[574] NIS Directive, art 14
[575] 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
[576] Critical Infrastructure Directive, art 2(a)

general interest. This would mean that, under the Recast Proposal, the PSI regime would cover those services offered by essential services operators.

There is however an inherent tension between the PSI Directive'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 the sensitivities presented in this sub-Section should be taken into account in current and subsequent revisions of the PSI Directive.

While the PSI regime currently proposed would still be optional for public undertakings, uncertainty exists about whether or not this might be made mandatory in the future.[577] Additionally it should be noted that the PSI Directive does not apply to documents which are excluded from access by virtue of the access regimes in the Member States, including on the grounds of protection of national security, defence or public security, statistical confidentiality or commercial confidentiality.[578] Still, this may not cover all data generated by providers of essential services or operators of critical infrastructures, which would then still be subject to the PSI Directive.

3.8.6 Summary

The Open Data movement and governments around the world, including the EU, are committed to make 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 public sector information throughout the EU.

The benefits of opening up public sector data are many. Open PSI can help eliminate barriers to market entry that start-ups and SMEs would otherwise be faced with due to the limited nature of their resources when compared to those of larger companies. Importantly, the potential commercialisation of previously inaccessible public data entails significant opportunities for economic value creation by being applied directly and indirectly in applications across all sectors of the economy. This can lead to increased revenues, increased competition, increased innovation and job creation. Benefits can also be achieved through efficient use of PSI by other public sector bodies, and by establishing collaborations between public and private entities.

[577] See "*Legal uncertainty*" under sub-Section 3.8.5.2
[578] Recast Proposal, art 1(2)(d)

Still, open data regimes also encounter a number of challenges, on a technical, economic and legal level, that cannot be ignored. Public sector data is often made available under a licence, and the PSI Directive encourages Member States to adopt standard licences. It appears however that, despite guidelines in this respect, very little uniformity can be observed and Member States have embraced very different licensing practices. This imposes a burden on companies that wish to re-use public sector information from different Member States in the development of their product. Another significant challenge, mainly for public sector bodies, is presented by the GDPR and how it relates to the PSI Directive. While the relationship is clear in theory, in practice it is no easy task to determine whether a dataset contains personal data and whether or not it should be made public. Finally, companies have voiced concerns about public sector bodies relying on intellectual property rights protection, notably through the Database Directive, as an excuse not to have to make their datasets available.

The proposal of April 2018 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. This is a significant development for the transport sector, where services are often provided by public undertakings. The same applies to critical transport infrastructure. Positively, the Recast Proposal takes into account the commercial interests of public undertakings and the fact that they are subject to competitive market forces. The regime would moreover be optional for public undertakings, but there is uncertainty as to whether this is only regarded as a first step and will be made mandatory in the future.

There is however an inherent tension between the PSI Directive'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 essential services. 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. These considerations should be taken into account in current and subsequent revisions of the PSI Directive.

Opportunities in relation to open data in the context of big data in the transport sector	Challenges in relation to open data in the context of big data in the transport sector
Open data can help eliminate barriers to market entry for start-ups and SMEs, by giving them access to resources which they would not be able to access otherwise.	Although standard open data licences are encouraged by the PSI Directive, it is shown that in practice, licences are still widely diverging in different Member States.
Open data may lead to an increase in competition and opens up potential for private companies to innovate, which creates value for the economy.	Public bodies are faced with the difficult task of reconciling their obligations under the PSI Directive and requirements under the GDPR. Most importantly, they must deal with the question whether a dataset contains personal data and should be subject to the PSI Directive or not.
Open data can create public-private sector collaborations and synergies, which in turn create economic value and are another ground for innovation.	Where the PSI regime would be applied to public undertakings, their commercial interests should be taken into account to prevent distortion of competition in the relevant market.
	The example of essential services and critical infrastructures shows that there are limits to the desirability of open data policies, which should be taken into account by the EU legislator in current and future reviews of the PSI Directive.

Table 32: Summary table of opportunities and challenges in relation to open data in the context of big data in the transport sector

3.9 *Data sharing obligations*

3.9.1 Introduction

While the PSI Directive examined in the previous Section imposes data sharing obligations on public sector bodies, this Section will focus on legal instruments imposing data sharing obligations on private undertakings.

In 2016, 254,850 data companies were operating in the EU.[579] Many of these companies generate and collect data themselves. That data is for a large part machine-generated (generated from real time sensors in industry machinery, vehicles or other products)[580] but also includes human-generated and organisation-generated data.[581] The huge amounts of data involved pave the way for increasingly innovative big data-enabled applications. Previous Sections have already mentioned some of the benefits of such applications, including enhanced safety and security, maximizing economic benefits, fostering innovation, lowering barriers to market entry and increased transparency and efficiency. But when left to themselves, private companies do not necessarily engage in data sharing among each other. This is due to the large number of challenges associated with private sector data sharing.

The different Sections of this Deliverable 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. These notably relate to privacy and data protection when personal data are involved, (cyber-)security, intellectual property rights, trade secrets and confidential information, data ownership and access to data, barriers to the free flow of data, liability concerns and competition law concerns. We refer to the relevant Sections for a detailed analysis of those obstacles as well as their potential mitigants.

Barriers to private sector data sharing are obviously not only of a legal nature. Many commercial and technical barriers also come into play. Any business having made an investment which resulted in the generation of commercially valuable data will naturally wish to protect that investment. It will moreover seek to maximise and retain a competitive advantage over its competitors. Finally, technical obstacles such as portability, interoperability and the lack of standards also constitute considerable barriers. Lack of interoperability presents a high cost for businesses that want to access data, particularly for SMEs and start-ups.[582]

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.[583] In the context of the LeMO Project, we have limited our analysis of data sharing obligations to those that were particularly relevant to the transport sector. The below is in no way an exhaustive list, as many other EU instruments impose data

[579] Commission, 'Towards a common European data space' (Communication) COM (2018) 232 final, 2
[580] http://www.iosrjournals.org/iosr-jce/papers/conf.15013/Volume%202/1.%2001-05.pdf, 3.
[581] Ibid 3-4
[582] Mugdha Ghotkar and Priyanka Rokde 'Big Data: How it is Generated and its Importance' (2016) 2 IOSR-JCE
[583] SWD (2017) 2 final, 21

sharing obligations, but attempts to offer a succinct examination of pieces of legislation most relevant to the LeMO Project.

3.9.2 Data sharing obligations in the transport sector

It appears that the data sharing obligations 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 services. Furthermore, while creating increased consumer transparency is an objective of many of the examined data sharing obligations, 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 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. 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.

The advent of Intelligent Transport Systems 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 Intelligent Transport Systems 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, and will be briefly examined in the following sub-Sections.[584]

While the delegated regulations adopted pursuant to the ITS Directive focus on road transport, Intelligent Transport Systems 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

[584] European Commission, 'Intelligent Transport Systems: Action Plan and Directive' (*European Commission*, 2018) <https://ec.europa.eu/transport/themes/its/road/action_plan_en> accessed 18 October 2018

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.[585] This is in part due to the fact that Intelligent Transport Systems 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.

3.9.2.1 Travel Information Services

The Commission Delegated Regulation with regard to the provision of EU-wide multimodal travel information services (2017/1926/EU)[586] is aimed at establishing the specifications necessary to ensure the cross-border accuracy and availability of multimodal travel information services to ITS users. Multimodal travel information services help travellers plan their route by comparing different travel options that combine various transport modes. The Delegated Regulation therefore necessarily applies to all modes of transport, including e.g. air and rail, as well as ride sharing, bike-hire, cableways and walking. [587]

In order to achieve its goal of providing seamless Union-wide multimodal travel information services, the Delegated Regulation introduces a number of 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. [588] A distinction is made between static and dynamic data. The former is data that (almost) does not change and is deemed essential for the end-user's pre-trip phase. Static data must therefore in any case be shared. The latter relates to e.g. travel disturbances and delays and can thus help travellers make informed decisions, but Member States are free to decide whether or not such data must be shared through their national access point.

The data cannot simply be supplied *as is*, but certain conditions will have to be fulfilled. Most notable is the fact that data on several modes of transport will have to meet specific standards and that there is an obligation to update the data in a timely manner whenever changes occur. This obligation also applies to the correction of inaccuracies.[589] Additionally, the Delegated Regulation sets conditions for the reuse of the data and the linking of travel information services, including non-discrimination, transparency and timely provision of services. It furthermore foresees that data reuse may be regulated through licence agreements, but these should, by default, impose as few restrictions on reuse as possible. Finally, financial compensation may be demanded but must be reasonable and cost-based.[590]

---

[585] 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

[586] 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

[587] Travel Information Services Delegated Regulation, Recital 8

[588] Travel Information Services Delegated Regulation, arts 4 and 5

[589] Travel Information Services Delegated Regulation, art 6

[590] Travel Information Services Delegated Regulation, art 8

Sharing obligations in the transport sector – Example 1

We are already seeing the deployment of multimodal travel information services today. One of these is the start-up Qixxit, founded by the German railway operator Deutsche Bahn. Qixxit is a digital travel planner that connects various modes of transport and providers to find optimal routes. It combines train, flight and even long-distance bus to offer users a coherent route and allow them to arrive at their destination in a fast and affordable way. The start-up went live in May 2018 and is available as an end-application for smartphones.[591]

3.9.2.2 eCall

In the context of eCall, two instruments were adopted with the aim of improving road safety by reducing fatalities as well as the severity of injuries caused by road accidents, and by improving incident management and reducing congestion and secondary accidents.

Pursuant to the mandate received through the ITS Directive, the European Commission adopted the e-Call Delegated Regulation (305/2013/EU), which laid 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.

That Delegated Regulation has been complemented by the e-Call Regulation (2015/758)[592], requiring 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.[593] It explicitly foresees the possibility for car manufacturers and independent operators to develop and deploy additional emergency services or added value services in parallel with or building on the mandatory e-Call system, thereby creating opportunities for private companies to offer value-added technology-based services. The Regulation moreover requires manufacturers to make the e-Call system accessible to all independent operators for a reasonable fee and without discrimination for repair and maintenance purposes in accordance with Regulation EC/715/2007 (see below).[594]

3.9.2.3 Minimum universal traffic information

Another instrument that was adopted to improve road safety is Delegated Regulation 886/2013/EU on road safety-related minimum universal traffic information. This imposes on both public and private road operators and/or service providers an obligation to detect and

[591] https://www.deutschebahn.com/en/Digitalization/startups/db_startups/qixxit_en-1214910, accessed 24 September 2018.
[592] 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
[593] Claudiu-Dan Bârcă, Rareş Ropot and Sorin Dumitrescu, 'eCall – Minimum Set Of Data (MSD)' (2009) Journal of Information Systems & Operations Management 428, 429
[594] e-Call Regulation, art 5(7)

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. Users to which the data are accessible include private road operators, traffic managers, service providers, and broadcasters dedicated to traffic information. Accessibility must moreover be in accordance with the access rights and procedures for public bodies laid down in the PSI Directive, be non-discriminatory, and ensure timely provision of the service. These must then in turn provide the relevant information to end-users, where possible free of charge.

3.9.2.4 Information services for parking for trucks and commercial vehicles

The objective of Commission Delegated Regulation 885/2013/EU with regard to the provision of information services for safe and secure parking places for trucks and commercial vehicles 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 standardized machine-readable formats to a national access point. Access to this information must be accessible for exchange and reuse by any public or private information service provider or parking operator on a non-discriminatory basis and in accordance with access rights and procedures defined in the PSI Directive. If costs are charged, these must be 'reasonable' as set out in the PSI Directive. Service providers are in turn required to disseminate this information to drivers.

3.9.2.5 EU-wide real-time traffic information services

The Delegated Regulation with regard to the provision of EU-wide real-time traffic information services (962/2015/EU) 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. This is expected to generate higher quality information services for both passengers and freight operations and to enhance the EU industry's competitive position.[595]

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. Similar conditions apply as already mentioned for some of the other delegated regulations under the ITS Directive: access and reuse of the data must be allowed on a non-discriminatory basis and must allow timely provision of the information services they are supposed to enable. Additionally, metadata must also be provided, as well as information on the quality thereof. Similarly to the Travel Information Services Delegated Regulation, an obligation is introduced to update the data in a timely manner whenever changes occur, and

[595] Commission, 'The provision of EU-wide real-time traffic information services Accompanying the document Commission Delegated Regulation 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' (Staff Working Document) SWD (2014) 356 final, 2

to correct any inaccuracies detected. Unlike the abovementioned ITS-related instruments, this Delegated Regulation remains silent on the issue of costs for the mandatory data sharing.

3.9.2.6 Infrastructure for Spatial Information in the European Union

Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (the "**INSPIRE Directive**") lays down rules to set up an infrastructure for spatial information for the purpose of EU environmental policies. It targets spatial information, which is information directly or indirectly referencing a specific location or geographical area and includes information related to transport networks. 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. [596]

The INSPIRE Directive allows Member States to limit public access to spatial data and services for various reasons, such as the confidentiality of commercial or industrial information or intellectual property rights. It follows that there is an explicit acknowledgement of the issues of confidentiality of commercial or industrial information, as well as of a possible protection of such information by intellectual property rights. Hence, while there is a general transparency obligation applicable to documents held by authorities[597], there are limits to such principles in favour of companies.

Sharing obligations in the transport sector – Example 2

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.[598]

3.9.2.7 Advance Passenger Information

Certain legislations have a narrow scope, requiring the disclosure of limited amounts of data for well-defined purposes. This is for instance the case in the aviation sector with the Advance

[596] INSPIRE Directive, Recital 18

[597] See for instance Regulation (EC) 1049/2001 of the European Parliament and of the Council regarding public access to European Parliament, Council and Commission documents [2001] OJ L 145/43 ("Regulation 1049/2001"), implementing Article 15 of the Treaty of the Functioning of the European Union on transparency. Such Regulation relates to the transparency obligation applicable to the EU institutions and agencies.

[598] 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

Passenger Information Directive (2004/82/EC). Air carriers must communicate information concerning passengers, and thus "personal data" to certain authorities for the purpose of combating illegal immigration.[599] 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.

3.9.2.8 *Rail Passengers' Rights*

Regulation 1371/2007/EC[600] on rail passengers' rights and obligations is primarily an instrument of consumer protection, aiming to guarantee the passengers' receipt of information both before and during the journey.[601] As a corollary, its more indirect goal is to improve the quality and effectiveness of rail passenger services and thereby help increase rail transport's market share compared to other modes of transport.[602]

Pursuant to this Regulation, railway undertakings must also 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.[603] Railway undertakings must additionally provide a limited amount of information during the journey.

The Rail Passengers' Rights Regulation furthermore requires all information to be provided through a "Computerised Information and Reservation System for Rail Transport" or CIRSRT for which technical specifications for interoperability would be established.[604]

3.9.2.9 Vehicle Emissions

The Vehicle Emissions Regulation (715/2007/EC) 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**"). 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.[605]

The Vehicle Emissions Regulation 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.[606] Therefore, when a consumer buys a

[599] Advance Passengers' Information Directive, art 1
[600] Regulation (EC) No 1371/2007 of the European Parliament and of the Council on rail passengers' rights and obligations [2007] OJ L 315/14
[601] Rail Passengers' Rights Regulation, Recital 4
[602] Rail Passengers' Rights Regulation, Recital 1
[603] Rail Passengers' Rights and Obligations Regulation, art 8
[604] Commission Regulation (EU) No 454/2011, Recital 5
[605] Vehicle Emissions Regulation, art 3(14)
[606] 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

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.[607]

These information-sharing obligations were primarily introduced to eliminate competition-restricting barriers in the market. It had appeared that the market for repair and maintenance services and for vehicle spare parts was significantly less competitive than the new car sales market, as spare parts and technical knowledge are often specific to a brand or model. The information-sharing obligations should however ensure competition in the vehicle aftermarket sector and broaden consumer choice.[608]

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[609], which will apply as of 1 September 2020.

3.9.2.10 Car Labelling

The Car Labelling Directive (1999/94/EC) 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.

3.9.2.11 *Vessel Traffic Monitoring*

The Vessel Traffic Monitoring Directive (2002/59/EC) was adopted in 2002 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 use of AIS promotes maritime safety, enables a coastal state to obtain information about ships and can moreover serve as a tool for vessel traffic services, which are services designed to improve the safety and efficiency of vessel traffic and to protect the environment.[610] Mandatory fitting of AIS may also offer commercial opportunities as AIS is, by its nature,

---

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

[607] Vehicle Emissions Regulation, art 7

[608] COM (2016) 782 final, 3

[609] Regulation (EU) 2018/858 of the European Parliament and of the Council on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC [2018] OJ L 151/1

[610] Vessel Traffic Monitoring Directive, art 3(o)

transmitted in an un-encrypted way over open frequencies, anyone with appropriate equipment can receive it and make use of it.[611]

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. Data to be reported typically include ship identification and type, navigation information (course speed), as well as details of cargo type and the total number of persons on board.[612] Additionally, the Vessel Traffic Monitoring Directive introduces reporting requirements on dangerous and/or polluting goods aboard ships and on hazardous ships and incidents at sea.[613]

Sharing obligations in the transport sector – Example 3

In the context of the ESSnet Big Data project, the quality of AIS data for the use of big data analytics was investigated. Several proofs of concepts were developed and examined, which generated promising results. One proof of concept evolved around the development of an algorithm to calculate the intra-port journey of a vessel by using AIS. The proof of concept proved successful. These intra-port travel distances could be paired with other information to create value-added services for port authorities, shippers and other stakeholders. In another proof of concept, it was shown that AIS data is useful to investigate fluvio-maritime transport, i.e. transport by ships that travel across both seas and inland waterways. One could imagine this data being used for the development of end-user services offering insights into traffic intensity and vessel emissions.[614]

3.9.3 Other data sharing obligations

3.9.3.1 Unfair Contract Terms and Unfair Commercial Practices

To a limited extent, data sharing obligations may arise under the legislation relating to unfair contract terms[615] and unfair commercial practices[616] when a data-holding company is preventing access to data in a particularly unfair manner.

[611] Directorate-General for Maritime Affairs and Fisheries, 'Legal Aspects of Maritime Monitoring & Surveillance Data: Summary Report' (European Commission 2009) 3 <https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/docs/body/legal_aspects_maritime_monitoring_summary_en.pdf> accessed 18 October 2018

[612] Ibid 4

[613] Vessel Traffic Monitoring Directive , Titles II "Notification of dangerous or polluting goods on board ships (Hazmat)" and III "Monitoring of hazardous ships and intervention in the event of incidents and accidents at sea", Vessel Traffic Monitoring Directive

[614] Anke Consten and others, 'Deliverable 4.3 Report about Sea Traffic Analyses using AIS-data. Version 2017-07-21' (ESSnet Big Data 2017) 20-29 <https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/5/5c/WP4_Deliverable_4.3_2017_07_21_v1.0.pdf> accessed 18 October 2018

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

[616] 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")

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 respective rights and obligations of the contracting parties. While this legislation is in principle applicable only to contracts in a business to consumer relationship, some Member States apply it (or its principles) to B2B relations as well.[617] 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.[618]

3.9.3.2 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**").[619] Its aim is to create a fair, transparent and predictable business environment for smaller businesses and traders when using online platforms.

The Regulation would 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.[620]

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.[621]

[617] SWD (2017) 2 final 21
[618] 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
[619] 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
[620] Proposed Platform-to-Business Regulation, art 1
[621] Ibid art 2(5)

The Platform-to-Business Regulation as proposed 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;[622] 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.[623]

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.[624] The draft Regulation however provides that such obligation should not require online intermediation services or online search engines to disclose any of their trade secrets (see also sub-Section 3.7.3).

3.9.3.3 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. Otherwise the business models of Google and Facebook, for a large part built on the control of user data, might be sent to the garbage can.[625]

Striking the right balance between access to and legitimate control of data is thus a delicate task. The CJEU 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 would prevent the second from being developed at all; and (iv) the refusal cannot be justified by objective reasons.[626]

---

[622] Ibid art 3
[623] Ibid art 7
[624] Ibid art 5
[625] 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
[626] 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

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.[627] 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, please refer to Section 3.13.

3.9.3.4 Data sharing obligations imposed through 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.[628]

3.9.4 Summary

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. This Section focused on such legally imposed data sharing obligations on private undertakings.

We succinctly examined the body of legislation specific to the transport sector that could impact a company's control of, the access to, or the rights in data. Our analysis shows that data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems. In the framework of the ITS Directive, numerous data sharing obligations were established, mostly in the domain of road transportation. It should be noted however that ITS is not limited to transport by road, but extends to all modes of transport including rail transport, maritime transport and transport using inland waterways, and air transport. It is therefore not unimaginable that regulation containing data sharing obligations will be adopted for those modes of transport as well.

Another interesting finding is the fact that increasingly, technical specifications are adopted for different modes of transport. This is largely due to the interoperability issues that would otherwise arise for Intelligent Transport Systems and which could render those systems incompatible and potentially inoperable. In light of the above, it may be expected that more

[627] 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

[628] 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

technical specifications will be adopted in the future, which might in turn entail additional data sharing obligations.

In general, the 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.

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. The former should usually be provided free of charge, while for the latter some kind of remuneration may be demanded. A similar distinction applies based on the nature of the purpose pursued. If an instrument is mainly concerned with data sharing for public safety purposes, the obligation should be fulfilled free of charge. Where data sharing obligations are however enforced to foster development of innovative and competitive services, 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. 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.

Overall, a clear increase can be observed in legislation imposing data sharing obligations, which can be linked to the development of Intelligent Transport Systems. 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.

Opportunities in relation to data sharing obligations in the context of big data in the transport sector	Challenges in relation to data sharing obligations in the context of big data in the transport sector
Data sharing obligations imposed through legislation may offer opportunities for increased competition and innovation by opening up data to private actors which would otherwise not have access.	Data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems. The European legislator should however make sure always to consider the necessity of these obligations.
Typically, when data must be shared with other private actors, we see that some kind of remuneration may be demanded, allowing the businesses involved to recover the related costs.	We recommend carefully considering whether or not to allow private actors to recover the costs of mandatory data sharing.
The potential benefits (and challenges) of alternative ways of imposing data sharing obligations, such as through public tenders, should be further investigated.	The rise of technical specifications and standardisation requirements, which are often necessary to ensure interoperability in the context of Intelligent Transport Systems, could also lead to an increased adoption of data sharing requirements.

Table 33: Summary table of opportunities and challenges in relation to data sharing obligations in the context of big data in the transport sector

3.10 *Data ownership*

The EU Commission has voiced on multiple occasions the most important legal issues in a data environment. In its data-driven economy Communication of July 2014, 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*".[629]

Indeed, if they cannot rely on any of the other exclusive rights discussed in this Deliverable (e.g. intellectual property rights), stakeholders in the (big) data analytics lifecycle increasingly try to claim "ownership" in (parts of) the datasets used in the analytics. This Section examines the legal implications of such evolution.

3.10.1 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*."[630] 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.[631] 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.

Facing such different meanings, an author suggested in 1998 already to use the term "data stewardship" as it would be more appropriate[632], capturing the "*responsibility that organisations are actually looking to promote with the ownership concept*."[633]

In this report, 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

---

[629] 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 18 October 2018

[630] Jonathan Law and Elizabeth A. Martin, *A Dictionary of Law* (7th edition, Oxford University Press 2014) <http://www.oxfordreference.com/view/10.1093/acref/9780199551248.001.0001/acref-9780199551248-e-2745?rskey=2MFh2r&result=2900> accessed 18 October 2018

[631] OECD, *Data-driven Innovation: Big Data for Growth and Well-being* (OECD Publishing 2015) 195

[632] Michael Scofield, 'Issues of Data Ownership', (*Information Management*, 1 November 1998) <http://www.information-management.com/issues/19981101/296-1.html> accessed 18 October 2018

[633] OECD, *Data-driven Innovation: Big Data for Growth and Well-being* (OECD Publishing 2015) 195

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".

3.10.2 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 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.[634]

The following Figure aims to depict the data value cycle.[635]

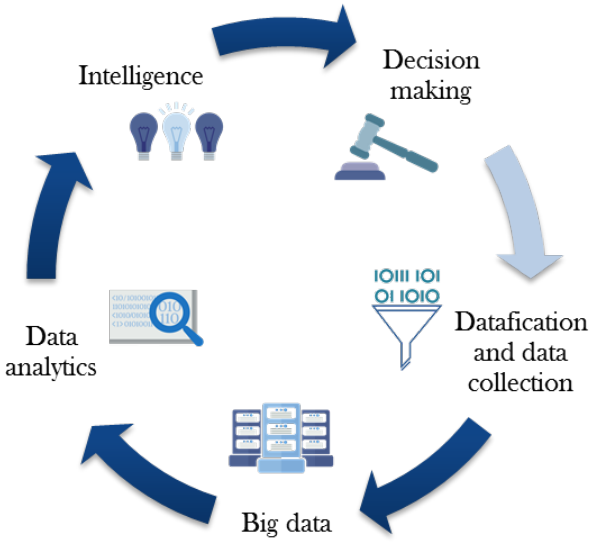


Figure 10: Data value cycle

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.[636]

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 and their central role are summarised below.

[634] Ibid
[635] Ibid 33
[636] See in this respect the recent Free Flow of Data Initiative of the EU Commission as part of the Digital Single Market

3.10.2.1 Internet Service Providers

Internet Service Providers ("**ISPs**") are at the heart of the data ecosystem through which data is exchanged.[637] They play an important role at the beginning of the process, as they provide the necessary technical foundations to end-users (organisations or individuals), or to other ISPs. Certain ISPs also provide supplementary IT services, such as cloud computing and data analytics services. Consequently, ISP's play a fundamental part in big data analytics, including – for some of them – by offering specific data-related services and/or availing of such services for their own needs.

3.10.2.2 IT Infrastructure Providers

IT infrastructure providers make available to other companies the toolkit, including both software and hardware, to handle and analyse big data. They offer tools for data analytics, data management, critical computing, data storage and transport, cloud computing, software allowing database management and analytics, etc.[638] A typical example is Hadoop, which has almost become a 'standard' technology allowing to deal with complex unstructured large volumes of data.

3.10.2.3 Data Providers

Various kinds of data (service) providers are active in the data environment.

- *Data brokers and marketplaces*

Data brokers and marketplaces compile and aggregate information (including personal data) obtained from a broad range of sources with the ultimate objective to sell, license or otherwise distribute such data to companies, consumers or other data brokers. Possible data sources include[639]:

  - Data disclosed or provided by organisations or individuals;
  - Data from sensors;
  - Data mined or crawled on the Internet;
  - Data obtained from not-for-profit organisations;
  - Open data (see Section 3.8);
  - etc.

- *Individuals (such as data subjects, consumers, patients, etc.)*

Certain individuals play an active role in the data economy either by providing their data (be it personal or not) to organisations (including data brokers), or by assembling, storing and managing their own (personal) data; including in the cloud.

[637] OECD, *Data-driven Innovation: Big Data for Growth and Well-being* (OECD Publishing 2015) 72
[638] Ibid
[639] Ibid 82

- *Public sector*

Public authorities have been active for several years in making certain sets of data 'freely' available – a concept which is also known as "open data" (see also Section 3.8). In the EU, for example, the EU institutions adopted a Directive on the re-use of public sector information (government-held data), which aims at unlocking the potential of big data held and accumulated by government authorities.[640]

3.10.2.4 Data Analytics Service Providers

The analysis of data is oftentimes performed by ISPs, IT infrastructure providers or data providers. Nevertheless, the data ecosystem still includes specific providers of data analytics services, including for the development of dedicated software and visualisation tools based on data analytics.[641] The role of data analytics service providers tends to be assumed by start-ups or SMEs specifically active in the development of new techniques, such as predictive analytics, simulations, scenario development, and advanced data visualisations.[642]

Peculiar, however, is the fact that data analytics service providers, contrary to data brokers, generally obtain their data directly from their customers, rather than from third party sources. This naturally has consequences for the identification of actors as data controllers or processors in a data protection context (see sub-Section 3.1.2). Taking into account that particularity of their service, data analytics service providers usually qualify as "data processors", rather than data controllers. Data brokers, on the contrary, are generally considered to be independent data controllers.

3.10.2.5 Data-driven Entrepreneurs

The last category of actors we will discuss covers those organisations developing cutting-edge products, services and technologies based on the use of data and data analytics for different purposes; the so-called data-driven entrepreneurs. These include start-ups and incumbents, but also innovative (ICT and non-ICT) companies and governments. Not only do they use data as the core enabler for their business operations; for a majority of them it can even be said to be the fundamental economic value behind the service they provide. Against such background, data becomes a valuable asset due to the transformation of data into know-how and intelligence, and thus it can be used for decision-making purposes.

3.10.2.6 A Layered Approach of the Key Roles of Actors

The actors as well as their roles briefly explained above can be depicted in layers, whereby the underlying layers supply the upper layers with goods and services[643]:

[640] Directive 2003/98/EC of the European Parliament and of the Council on the re-use of public sector information [2003] OJ L 345/ 90
[641] OECD, *Data-driven Innovation: Big Data for Growth and Well-being* (OECD Publishing 2015) 86
[642] Ibid
[643] Ibid 72

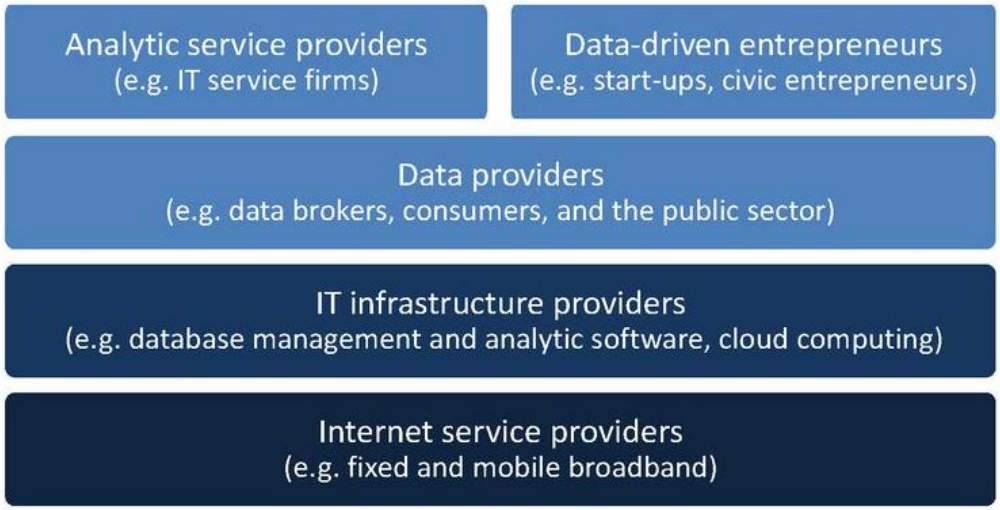


Figure 11: The data ecosystem as layers of (key roles of) actors

3.10.3 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 – see Section 3.7).

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

3.10.4 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.

European Union

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 (case C-128/11). 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.[644] 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.

Despite such ruling and the possible interpretation deriving from it, a high legal uncertainty remains.

Certain issues that may be of particular interest when considering the "ownership of data" have been addressed by the **German** Courts.

The first landmark decision comes from the Higher Regional Court of Karlsruhe[645] and concerns destruction of data. The Court 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. On the other hand, later decisions of German courts opposed the possibility to hold ownership over data as such, since data lacks the necessary material character[646] and since it is not considered a 'thing' under the German Civil Law Code.[647]

Subsequently, the Court of Appeal of Nuremberg[648] 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, the employees were allowed to delete the data.[649] The Nuremberg Court has however indicated that whilst the rule derived from the "Skripturakt"-theory also applies in the employment context, the situation may be different if the data have been already passed over to the employer; in such case the employer would become the owner of data. In

[644] Thomas Hoeren, 'Big Data and the Ownership in Data: Recent Developments in Europe' (2014) 36(12) EIPR 751
[645] OLG Karlsruhe, Urt. v. 07.11.1995 – 3 U 15/95 – *Haftung für Zerstörung von Computerdaten*
[646] LG Konstanz, Urt. v. 10.05.1996 – 1 S 292/95 = NJW 1996,2662
[647] OLG Dresden, Beschl. v. 05.09.2012 – 4 W 961/12 = ZD 2013,232
[648] OLG Nürnberg 1. Strafsenat decision of 23.01.2013, 1 Ws 445/12
[649] 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

addition, in the Court's opinion, the data will originally belong to the employer if they were created completely according to his demands.[650]

The Labour Court of Appeal of Saxony (Landesarbeitsgericht) had to decide a similar case in 2007, from a civil law perspective.[651] This decision is however somewhat contradictory to the one issued by the Nuremberg Court later on. The Saxon Court claimed that because the employee installed software (Microsoft Outlook) on a company-owned laptop, the employer has obtained the property in the software. In consequence, when the employee deleted the software from this laptop he destroyed the data of the employer, and could therefore be dismissed.

While the question of "ownership" of data was also indirectly addressed by the courts in the **United Kingdom**, they did not set out clear rules on that matter. So far, the UK courts held that data is not property and therefore cannot be stolen[652], that data are not eligible to be the subject of a common law lien[653], and that there is no proprietary right in the content of an email.[654]

Finally, the **French** Supreme Court ("*Cour de cassation*") rendered a ruling[655] in 2015 that could open a way to recognising the ownership of "data". The Court found that downloading (remotely) computer data without taking away their support may amount to the offence of theft, acknowledging therefore indirectly that such independent data may be owned.

3.10.5 Commission Communications having an impact on the Data Ownership Debate

3.10.5.1 "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*".[656]

---

[650] OLG Nürnberg 1. Strafsenat decision of 23.01.2013, 1 Ws 445/12, par. 16-17
[651] LAG Sachsen, decision of 17.01.2007, 2 Sa 808/05, MMR 2008, 416
[652] *Oxford v Moss* [1979] 68 Cr App Rep 183
[653] *Your Response v Datateam Business Media* [2014] EWCA Civ 281
[654] *Fairstar Heavy Industries v Adkin*, [2013] EWCA Civ 886
[655] May 20, 2015 (No14-81336)
[656] COM (2014) 442 final

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.

3.10.5.2 "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.*"[657]

Finally, in the context of the DSM, the EU Commission voiced its intentions to propose in 2016 a European 'Free flow of data' initiative that would notably address the restrictions on the free movement of data and the emerging issues of ownership, interoperability, usability and access to data in situations such as business-to-business, business-to-consumer, machine generated and machine-to-machine data (see also Section 3.6 on the free flow of data).[658]

3.10.5.3 "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. In such context, it reiterates its objective voiced in the Digital Single Market strategy "*to create a clear and adapted policy and legal framework for the data economy, by removing remaining barriers to the movement of data and addressing legal uncertainties created by new data technologies*."[659]

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*."[660] 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*."[661]

---

[657] 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

[658] Commission, 'A Digital Single Market Strategy for Europe' (Communication) COM (2015) 192 final

[659] COM (2017) 9 final, 4

[660] 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.

[661] Ibid

The Commission further stated that "*the use of existing general contract law and competition law instruments available in the Union might be a sufficient response*" and that "*voluntary or umbrella agreements covering certain sectors might be envisaged*."

Finally, the Communication suggests several possibilities[662], to be discussed with stakeholders, to move forward on the issue of access to machine-generated data, with the aim of achieving several objectives:

Objectives to achieve		Possible ways forward (non-exhaustive and not mutually exclusive)
Improve access to anonymous machine-generated data	Non-legislative measures	Guidance on incentivising businesses to share data
Facilitate and incentivise the sharing of such data		Fostering the development of technical solutions for reliable identification and exchange of data
Protect investments and assets		Model contract terms
Avoid disclosure of confidential data	Legislative measures	Default contract rules
Minimise lock-in effects		Access for public interest and scientific purposes
		Data producer's right
		Access against remuneration

Table 34: Moving forward on access to machine-generated data

3.10.5.4 "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.[663]

Such initiative was supported and driven by a stakeholder dialogue and replies to the Public Consultation on "Building the European Data Economy".[664] As regards business-to-business

[662] Such possibilities are detailed in the Commission Staff Working Document SWD (2017) 2 final, 30 ff
[663] COM (2018) 232 final

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*".[665]

For further information on the contents of this Communication, please refer to Section 3.6 on data sharing obligations.

3.10.6 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 right[666], 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.[667] Other academics do not necessarily dismiss the idea of an exclusive right in data, but claim its advent to be premature.[668] The authors of this Deliverable 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.[669]

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 **English** law was addressed for instance by Christopher Rees[670], 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 such[671], 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

---

[664] 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

[665] COM (2018) 232 final 9

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

[667] Josef Drexl 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 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; 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/fb02/makro/forschung/magkspapers/paper_2016/37-2016_kerber.pdf> accessed 18 October 2018

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

[669] 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 18 October 2018; 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 18 October 2018.

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

[671] See e.g.: Michael Dorner, 'Big Data und "Dateneigentum"' (2014) 9 CR 617, Malte Grützmacher; 'Dateneigentum – ein Flickenteppich' (2016) 8 CR 485

Courts. In particular, Prof. Dr. T. Hoeren examined the issues of data ownership under the current German legal framework and jurisprudence[672], 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, 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 such[673], others believe that in view of the abovementioned ruling of the French Supreme Court the ownership over data cannot be called into question.[674] Having said that, most discussions on the recognition of ownership seem to focus on individuals' ownership over their personal data.[675]

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

[673] 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

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

[675] The particular issue of personal data ownership has been discussed extensively in the context of D2.3 'Report on Ethical and Social Issues'; 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.

Data ownership in the transport sector – Example 1

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.[676] 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. The different measures listed under the second options are as follows: (i) targeted elimination of existing vulnerabilities; (ii) promotion of a single market for data through standardisation; (iii) removal of barriers for data mining and big data applications; (iv) promotion of open data (see also Section 3.8); (v) promotion of the awareness that data is a marketable good; and (vi) consolidation and merging of data-related regulations into one data law.

Data ownership in the transport sector – Example 2

The developments in relation to connected and autonomous vehicles have also raised questions with respect to data ownership.[677] 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.[678]

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

[677] 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

[678] Please also refer to Section 3.1 on 'Privacy and Data Protection' and Deliverable D2.3 'Report on Ethical and Social Issues'.

3.10.7 Summary

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 Section (and this Deliverable at large) 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 Section 3.11 below.

Opportunities in relation to data ownership in the context of big data in the transport sector	Challenges in relation to data ownership in the context of big data in the transport sector
Given the nature of the issue, no opportunity in relation to data ownership has been identified.	It may prove difficult to establish ownership of different data components within a set of data of various types and coming from various sources.
	Multiple actors involved in big data analytics may try to claim ownership of the data concerned, which may lead to a gridlock.

Table 35: Summary table of opportunities and challenges in relation to data ownership in the context of big data in the transport sector

3.11 Data Sharing Agreements

It follows from the previous Sections, including Section 3.10, that there is a multitude of actors on the market actively reaping the benefits of the data economy. The relationship between such actors is at the heart of the data value cycle.

It is however apparent from the previous Sections of this Deliverable 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 (if any).[679] 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.

This sub-Section offers a brief overview of what can be defined as a Data Sharing Agreement ("**DSA**") and of the rules that may apply to these agreements arising both from the law and from the contractual obligations established by the parties.

Although there are no specific regulations on DSAs *per se*, it shall nonetheless be noted that some national authorities have published guidance on some aspects of DSAs – mainly in relation to personal data (e.g. some Data Protection Authorities have provided guidance on it).[680]

3.11.1 Data sharing agreement definition

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 Section does not intend to cover any contractual relationships with natural persons in their capacity as consumers or data subjects.

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

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

In case of violation of these rules, such as when one of the parties discloses the data received from the other party to another party not authorised to receive the data, the disclosing party

[679] 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)

[680] See Irish Data Protection Commissioner <https://www.dataprotection.ie/docs/Commissioner-launches-new-guidance-on-data-sharing-in-the-private-sector/530.htm> accessed on 23 October 2018.

may be held liable for its contractual breach, but also other possible sanctions may arise from the law, including liability against third parties.

Under some circumstances it is also possible that policies internal to an organisation address the sharing of data and information among two or more departments of the same organisations. In such circumstances, in several jurisdictions it is not possible, from a strict legal perspective, to consider such policies as "agreements" since there are no separate legal entities.

Depending on the specific needs of the parties, the sharing of data may therefore take different forms, such as for instance:

- reciprocal exchange of data;
- one or more organisations providing data to a third party or parties;
- several organisations pooling information and making it available to each other or to a third party / parties;
- exceptional, 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, and thus do not necessarily concern personal data but other protected categories of information, such as for instance[681]:

- data about identified or identifiable natural persons ("personal data" – see also Section 3.1 on Analysis of key legal issues
- Privacy and 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.

3.11.2 General rules applicable to data sharing agreements

The DSA shall first of all be in line and comply with the applicable (national) laws and regulations concerning the formation and execution of a contract, notably relating to the activity of data sharing.

In addition to a general compliance with the possible restrictions laid down by the applicable legislations and/or regulations on the sharing of data, 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 kind of data or information.

[681] Non-exhaustive and non-exclusive list

Finally, data sharing will be construed in line with the will of the parties who, within the limits imposed by laws and regulations, will decide on terms and conditions for the sharing of data.

3.11.3 Completion and execution of a data sharing agreement

A DSA must comply with all possible rules, terms and conditions concerning the execution or completion of an agreement relating to the sharing of data. Most of such rules, terms and conditions 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 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[682]
- 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
- 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)

3.11.4 Terms and conditions set forth by the parties

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;

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

- 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.

3.11.5 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.[683]

The European Commission then issued guidance on 'Sharing private sector data in the European data economy'.[684] 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. business-to-business or "**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 data sharing agreements and six principles to govern B2G data sharing agreements. These will be briefly addressed in the next sub-Sections.

[683] 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

[684] 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

Data sharing agreements in the transport sector – Example 1
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 by the end of 2018. One of the issues to be addressed by the Recommendation, and which is therefore included in the Roadmap, 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 is open for feedback on the Better Regulation platform until 16 November 2018.[685] Any feedback will be taken into account for further development of the initiative.

3.11.5.1 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*".[686] These principles are displayed in Table 36 above:

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.

Table 36: Principles for B2B data sharing[687]

[685] 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 23 October 2018
[686] SWD (2018) 125 final, 3
[687] Ibid 3

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 2019.[688]

The section on DSAs also contains a list of considerations to help companies in the preparation and/or negotiation of data sharing agreements. 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.[689] 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.[690]

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.[691]

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.[692]

[688] Ibid 6
[689] Ibid 6-8
[690] Ibid 6
[691] Ibid 7
[692] 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.[693]

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.[694]

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.[695]

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, presented in Figure 12 above.

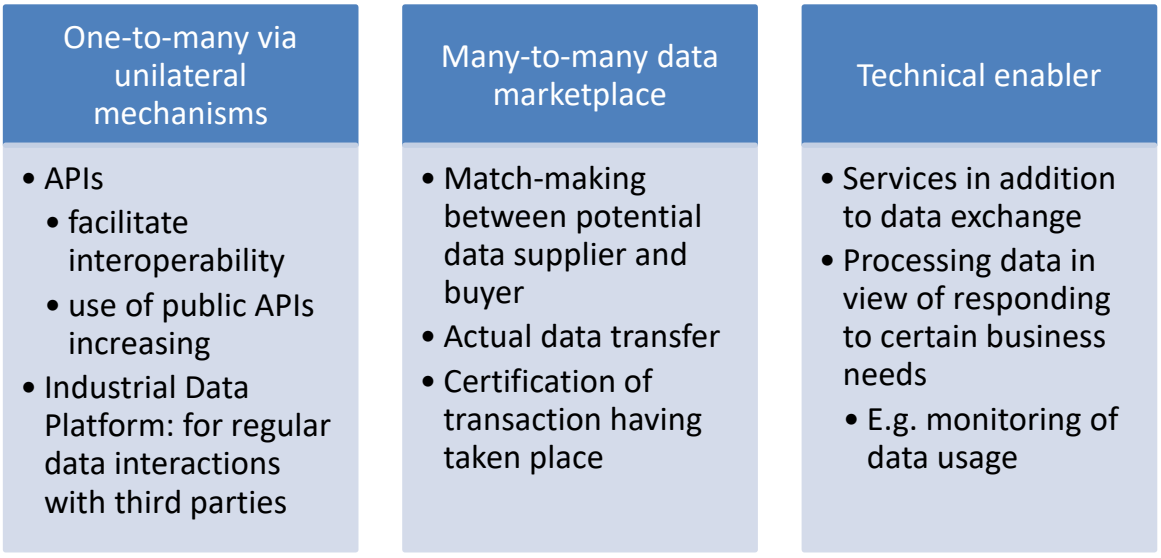


Figure 12: Technical mechanisms for data sharing according to the Commission[696]

[693] Ibid
[694] Ibid
[695] Ibid7-8
[696] Ibid 8

The two following examples on data sharing in the transport sector illustrate the importance of the convergence between the legal (contractual) aspects and the technical aspects of data sharing:

Data sharing agreements in the transport sector – Example 2

Traffic, navigation and mapping products provider TomTom shares data with other companies (B2B) on the basis of a unilateral business model.[697] From a legal perspective, comprehensive licensing agreements are put in place in order to define usage rights and remuneration, but also restrictions to re-use data. The data value is determined based on the features of the data to be shared, but also on the use that will be made of the data. The main business users of TomTom's data are original equipment manufacturers ("**OEMs**"), large vendors, technology companies, and geographical information system ("**GIS**") providers. From a technical perspective, TomTom uses APIs to enable the data sharing.[698] APIs also allow for metering and monitoring how data are used and for swift intervention in cases of misappropriation or misuse of data.

Data sharing agreements in the transport sector – Example 3

Nallian[699] has created a cloud-based customisable platform that facilitates real-time data sharing in a controlled, flexible and agile environment and supports process synchronisation.[700] The users of Nallian's platform are currently logistics hubs and companies, vertical supply chains and multimodal transport networks.[701] 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).[702] Examples of communities relying on the Nallian platform relevant to the transport sector are BRUcloud[703], CargoStream[704], NxtPort[705], and Heathrow CargoCloud[706].

[697] 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

[698] TomTom for Developers, 'TomTom Maps APIs for Developers' (*TomTom*, 2018) <https://developer.tomtom.com/tomtom-maps-apis-developers> 23 October 2018

[699] See <https://www.nallian.com>

[700] 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

[701] Ibid

[702] SWD (2018) 125 final, 11

[703] See <https://brucloud.com>

[704] See <https://www.cargostream.net>

[705] See <https://www.nxtport.eu>

[706] See <https://www.heathrow.com/company/cargo/cargocloud>

3.11.5.2 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 Table 37 above.

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.

Table 37: Principles for B2G data sharing[707]

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 Deliverable 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.[708]

The Commission also outlines different technical means to achieve B2G data sharing.

3.11.6 Data Sharing Agreements: a Critical Analysis

As already mentioned in this Deliverable, 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 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 data sharing agreements. The same is true for big data analytics in the transport sector. In view of the aforementioned complexity and multitude of actors, data sources, data flows, algorithms, etc., an intricate chain of data sharing agreements should be put in place in order for the big data analytics to (legally) function in practice.

However, the authors of this Deliverable are reticent to settle for data sharing agreements 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

[707] SWD (2018) 125 final, 3-4
[708] Ibid 14-16

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 data sharing agreement 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 an IPR holder) 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 data sharing agreements 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.

3.11.7 Summary

This Section examined the common practice to use contracts, i.e. 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.

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.[709]

[709] 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-

Opportunities in relation to data sharing agreements in the context of big data in the transport sector	Challenges in relation to data sharing agreements in the context of big data in the transport sector
At the time of writing, data sharing agreements provide the only solution to govern access to and/or exchange of data between the numerous stakeholders active in the big data value cycle, including in the transport sector.	The complexity of data flows as well as the multitude of actors, data sources, algorithms, analyses, and activities that can be performed on data in a big data context requires the conclusion of a myriad of intricate data sharing agreements.
	Contractual agreements cannot be enforced vis-à-vis third parties. This entails that no recourse is available against third parties that obtain unjustified access to or misuse the data.
	It proves extremely difficult to clearly define the concepts of "data" and "data ownership" in data sharing agreements as no legal definitions of these concepts exist.
	The actual terms of the DSA may be so restrictive that in essence the recipient cannot do anything with the data.
	Governing the big data analytics cycle through multiple data sharing agreements requires integrating complex back-to-back warranty clauses in respect of the upstream data sources as well as the downstream uses of data. In absence of such clauses, the use of data may be prohibited or restricted and the whole big data analytics chain may be affected.

Table 38: Summary table of opportunities and challenges in relation to data sharing agreements in the context of big data in the transport sector

the-european-data-economy> accessed 18 October 2018; 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 18 October 2018

3.12 Liability

3.12.1 Setting the scene

Liability issues are undoubtedly a major concern to take into account in the context of new technologies applied in the transport sector, including with respect to big data.

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.[710]

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.[711] 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 in relation to new / 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 Artificial Intelligence, 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.

---

[710] 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

[711] SWD (2018) 137 final, 9

This Section 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.

Before digging into the legal intricacies of the applicable liability regimes, an overview of the complexity of the technologies is needed in order to demonstrate the different layers of components/players that compose the new technologies, and that ultimately increase the complexity of liability issues.

In a nutshell, disruptive technologies are complex "*due to their interdependency between the different components and layers*"[712]:

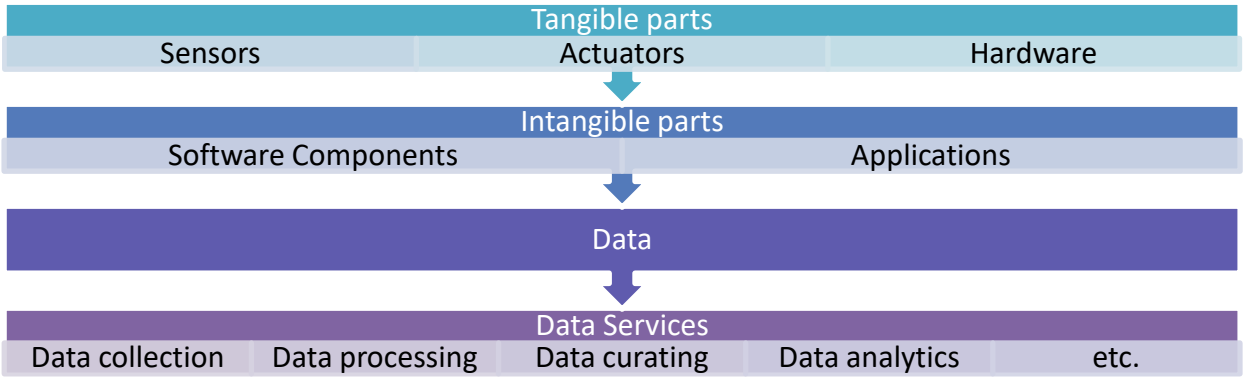


Figure 13: Overview of the components and layers of disruptive technologies[713]

It derives from the above diagram that the lower, intangible, layers create an unforeseen complexity and necessarily involve a plethora of new actors that may have some kind of responsibility / liability in case damage is caused. Necessarily, big data service providers find their place in the (lower level) complex ecosystem and are thus concerned by the debates surrounding the future liability regime.

Numerous questions related to liability arise, such as in particular:

- Who should be held liable in case the technology causes damage?[714]
- How to identify the root cause of the problem?[715]
- How to deal with liability in case damage is caused by a machine operating with a certain degree of autonomy but which cannot be linked to a defect or a human wrongdoing (e.g. by the driver or the car manufacturer)?[716]
- How to attribute liability where the expected outcome of the technology was not identified before the market launch or after that launch?[717]

[712] SWD (2018) 137 final, 9
[713] A fifth lower layer includes the connectivity features.
[714] SWD (2018) 137 final, 9
[715] Ibid
[716] Ibid 10
[717] Ibid

3.12.2 Extra-contractual and statutory lability 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 follows[718]:

[718] 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-FB14-08A4-31F51A564C60F2A7_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

Liability		Safety
Member States-driven	EU-driven	
Extra-contractual liability (including tort)	Statutory liability (including product liability)	Safety requirements
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). Extra-contractual liability can be imposed by general civil law rules, but also by specific legislation. Two main categories exist: • 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. • 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).	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.	The EU safety legislation aims at ensuring that only safe products can be placed on the internal market of the EU. This includes various instruments such as for instance: • Directive 2001/95 on general product safety • Directive 2006/42 on machinery • Directive 2014/53 on radio equipment Such system is further reinforced by harmonised standards, where such standards provide a presumption of conformity with the EU safety legislation.

Table 39: Schematic overview of the EU legal framework on liability

As clearly affirmed by the 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*."[719] It goes without saying that it should be clearly assessed whether changes to the above legal systems are needed in order to

[719] SWD (2018) 137 final, 9

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.

Liability in the transport sector – Example 1

It is already possible to rely on fully autonomous unmanned aircrafts, or "drones", for instance for the delivery of packages.[720] 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 imagined[721]:

- 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.

[720] SWD (2018) 137 final, 11-13
[721] Ibid

Liability in the transport sector – Example 2

Only few EU Member States have thus far adopted specific rules covering highly or fully automated vehicles.[722] 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.

It follows from the foregoing that many questions still need to be looked into when examining the current legal framework, such as for instance:

- To what extent are services (and embedded / non-embedded software) included in the product liability legislation?
- How can one identify whether the damage has been caused by the product itself or by other elements interconnected to it in a digital ecosystem?[723]
- Are concepts such as the "liability of a guardian" appropriate to new technologies?
- Should liability of autonomous systems be fault-based or strict?
- Does it matter, when determining liability, whether the damage could have been avoided or not?[724]
- What specific burden of proof should be adopted in relation to new technologies, including autonomous devices?
- What type of damage should be compensated when caused by new technologies?
- Should the liability be capped, and to what extent/amount?
- How are the liability issues addressed in situations where a complex ecosystem of market operators enables the roll-out and functioning of the emerging digital technology?
- Should the actors in the data value chain be obliged to take out insurance coverage (as it is currently the case for cars)?

All questions listed above have been rightfully identified by the EU institutions who are currently delving into their complexities. On the basis of their ongoing work, it will be possible to determine whether regulatory intervention is required.

In all likelihood, the following interventions will be required[725]:

[722] SWD (2018) 137 final, 13
[723] Ibid 18
[724] Ibid

Short- and mid-term	Long-term
Non-regulatory options	**Regulatory options**
• Recommended liability provisions, including model contract clauses and best insurance practices in a specific sector or in general • Identifying appropriate standards for safety assessments and certification in a specific sector or in general • Establishing Member State specific coordination and cooperation mechanisms to address cross border data economy challenges • Funding research and innovation, including in particular in relation to industrial / big data platforms	• Sector-specific legislation on minimum liabilities to be borne by certain service providers in certain sectors • General revision of liability law (such as expanding the scope of product liability legislation, revising the rules for products with an elevated risk profile, and harmonising certain extra-contractual liability aspects) • Legislating the insurance-related obligations[726]

Table 40: Expected interventions regarding the EU liability legal framework

3.12.3 Contractual liability

While the previous sub-Section only looked into the extra-contractual liability aspects, one should not ignore the contractual liability issues, which are particularly relevant in relation 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 B2C and 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.

---

[725] 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-FB14-08A4-31F51A564C60F2A7_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

[726] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), para 57 ff

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 business-to-consumer 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.[727]

In a B2B context, contrary to a B2C environment, 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.

3.12.4 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.

The safe harbour regime takes due account of fundamental rights, requiring striking a balance between conflicting rights: on the one hand the freedom or neutrality of intermediaries which should be guaranteed, and on the other hand, the rights of third parties (e.g. owners of intellectual property rights, personality rights, etc.), which deserve to be protected.

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

More specifically, Directive 2000/31/EC on Electronic Commerce[728] ("**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*"[729], "*caching*"[730] and "*hosting*".[731]

In short, the core idea is to protect intermediaries who are not the authors of the infringing / 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.

More specifically, Article 14 of the e-Commerce Directive concerns hosting,[732] which is of particular interest in the context of big data. The applicability of the exemption of liability will however depend on whether the big data analytics host fulfils the legal conditions in light of the characteristics of the service provided.

Article 14 provides that in order to benefit from the exemption, the host cannot have actual knowledge of the illegal activity or information and, as regards claims for damages, cannot be aware of facts or circumstances from which the illegal activity or information is apparent. As a result, according to case law of the CJEU, the host must have a passive, technical, and automatic role in the storage of data.[733]

It derives from the conclusions of the CJEU that in all likelihood, when a big data analytics host is playing a mere passive role, it should benefit from the safe harbour regime. To the contrary, if the big data analytics host is more active and thus somehow controls, selects or determines the content, it will most probably not benefit from the liability exemption.

As a result, it is of utmost importance to examine in depth the precise nature of the service provided by the big data analytics host as this will impact whether or not it may benefit from the favourable liability regime.

In any event, Article 14 of the Directive provides that the host must, upon obtaining knowledge or awareness of any illegal activity or information, act expeditiously to remove or disable access to the information. Also, the safe harbour regime does not affect the possibility for a court or administrative authority to require, in accordance with Member States' legal systems, the service provider to terminate or prevent an infringement, nor does it affect the possibility for Member States to establish procedures governing the removal or disabling of access to information.

---

[728] 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

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

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

[731] Ibid art14

[732] It consists of the storage of information provided by a recipient of the service.

[733] Case C-237/08 *Google v Louis Vuitton* [2010] ECLI:EU:C:2010:159; Case C-324/09 *L'Oréal v eBay* [2011] ECLI:EU:C:2011:474

Effectively, this requires the big data analytics host to implement adequate notice and take down procedures.

In this context, it shall further be noted that Article 15 of the e-Commerce Directive provides that there is (i) no general obligation on providers, when providing hosting services, to monitor the information that they transmit or store; and (ii) no general obligation to actively seek facts or circumstances indicating illegal activity. This has been confirmed by the CJEU, which held that Article 15 does not allow imposing a filtering obligation on intermediaries.[734]

In light of the above, it is recommended for a big data analytics host that wishes to preserve the benefit of the safe harbour regime to:

- Reserve the right to delete or disable content alleged to be infringing and to terminate accounts of repeat infringers;
- Put in place a response mechanism should the service provided by the big data analytics host infringe a third party's rights (e.g. provide a non-infringing equivalent service, modify the service, terminate the provision of a particular service and provide a refund, etc.).

Finally, it shall be noted that the EU Commission is currently examining the rules related to intermediaries, as part of its Digital Single Market strategy. To this end, it launched a public consultation in which the Commission sought the views of market actors and the wider public to better understand the social and economic role of platforms, market trends, the dynamics of platform-development and the various business models underpinning platforms. The EU Commission is currently assessing the regulatory framework applicable to platforms in order to determine whether changes are necessary.

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.

[734] Case C-70/10 *Scarlet v Sabam* [2011] ECLI:EU:C:2011:771; Case C-360/10 *Sabam v Netlog* [2012] (ECLI:EU:C:2012:85

Liability in the transport sector – Example 3

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 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. Further regulation and compliance obligations might apply in such case.

3.12.5 Liability aspects of the draft Directive on the supply of digital content

The Draft Directive on the Supply of Digital Content (hereinafter the "Draft Directive") aims to deal with the liability of suppliers of digital content towards the consumer.[735] This sub-Section merely aims to demonstrate the necessary evolution of liability regimes in the EU in order to tackle new technologies. The below analysis is based on the Commission's proposed text, which has not yet been adopted.

3.12.5.1 Overview of the liability aspects of the Draft Directive on the supply of digital content

According to the Draft 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 and/or service is provided on a continuous basis, the liability of the supplier is extended over the time of said supply. In

[735] Commission, 'Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content' ("Draft Directive"), COM (2015) 634 final, art 10

other words, the digital content and/or service supplier remains liable for defects existing at the time of supply without any time limit.

Because of the complexity characterising digital content and digital services, 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 Draft Directive provides for a reversal of the burden of proof, i.e. the burden of proof will lie with the supplier.[736]

Pursuant to Recital 44 of the Draft Directive the supplier's liability is an essential element "*to increase consumers' trust in digital content (…), consumers should be entitled to a compensation for damages caused to the consumer's digital environment by a lack of conformity with the contract or a failure to supply the digital content.*" The Draft Directive however leaves it up to the EU Member States to define the complete conditions for the exercise of this right to damages.

3.12.5.2 Analysis of the potential applicability and consequences on big data

The scarce stakeholders who have examined the issues of liability in the context of the supply of digital content have raised the following concerns[737]:

- The delivery of digital content depends on a number of factors located outside of the supplier's control, e.g. internet connection and available storage on a user's hard drive space.

  In that regard, suppliers may experience difficulties to prove that their digital content is not faulty and is not responsible for the user's poor experience. It has further been mentioned that providing the evidence would require the replication of the consumer's digital environment which is potentially time and cost consuming.[738]

- Another issue concerns the several players who may fall under the general notion of "suppliers".[739] Although only one final supplier is liable towards the consumer, there are different crucial suppliers along the value chain, e.g. online shops, platforms, and distributors.

  Digital businesses have expressed their concern that the Draft Directive will make those actors responsible for faulty content, while they do not necessarily have the adequate means to repair the digital content.[740] Some commentators therefore

[736] Draft Directive, art 9
[737] Very few commentators have examined the liability aspects of the Draft Directive.
[738] Deloitte, 'Impact of the European Commission's Draft Directive on Contract Rules for the Supply of Digital Content. Final Report' (Deloitte 2016) 28 <http://edima-eu.org/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf> accessed 26 October 2018
[739] Draft Directive, art 2.3
[740] Deloitte, 'Impact of the European Commission's Draft Directive on Contract Rules for the Supply of Digital Content. Final Report' (Deloitte 2016) 27 <http://edima-eu.org/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf> accessed 26 October 2018

estimate that a "secondary" liability should be introduced in the digital age for those third parties in order to avoid undermining consumers' rights.[741]

- Concerns were also expressed regarding user-generated content since a large part of the digital content supplied by digital business has been generated by their users. The liability provision of the Draft Directive could therefore make those businesses responsible for such user-generated content.[742]

It goes without saying that the liability aspects of the Draft Directive should necessarily be (re-)examined in light of the broader assessment of the adequacy of the current liability regime in the EU. Indeed, one may wonder about the precise necessity and opportunity to regulate liability in the specific context of the supply of digital content, rather than more broadly. This should be carefully evaluated in order to avoid additional complexity of the liability legal framework, and the associated legal certainty.

3.12.6 Summary

The authors of this Deliverable 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 Section 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.

The e-Commerce Directive provides for an exemption of liability for intermediaries, the so-called safe harbour regime. One of the exemptions concerns hosting,[743] which is of particular interest in the context of big data. The applicability of the exemption of liability depends however on whether the big data analytics host fulfils the legal conditions in light of the characteristics of the service provided. In its Uber judgment, the CJEU clarified that information society services that form an 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 and will therefore not be able to benefit from the safe harbour regime.

The Draft Directive on the Supply of Digital Content aims to deal with the liability of suppliers of digital content towards the consumer. One may however wonder about the precise

[741] Directorate-General for Internal Policies – Policy Department C: Citizens' rights and constitutional affairs, 'Sale of Goods and Supply of Digital Content - Two Worlds Apart?' (European Commission 2016) 18-19 <http://www.europarl.europa.eu/cmsdata/98774/pe%20556%20928%20EN_final.pdf> accessed 26 October 2018

[742] Deloitte, 'Impact of the European Commission's Draft Directive on Contract Rules for the Supply of Digital Content. Final Report' (Deloitte 2016) 27 <http://edima-eu.org/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf> accessed 26 October 2018

[743] It consists of the storage of information provided by a recipient of the service.

necessity and opportunity to regulate liability in the specific context of the supply of digital content, rather than more broadly. The liability aspects of the Draft Directive should therefore be (re-)examined in light of the broader assessment of the adequacy of the current liability regime in the EU. This should be carefully evaluated in order to avoid additional complexity of the liability legal framework, and the associated legal certainty.

Opportunities in relation to liability in the context of big data in the transport sector	Challenges in relation to liability in the context of big data in the transport sector
The ongoing work of the EU institutions regarding extra-contractual and statutory liability specifically, and the envisaged non-regulatory and regulatory interventions, may be beneficial for the uptake of big data in the transport sector.	The status of contractual liability rules across the EU is likely to limit the uptake of new technologies, including big data in the transport sector.
	In general, an unclear, non-harmonised EU legal framework on liability entails legal uncertainty and may accordingly stifle the uptake of big data in the transport sector.

Table 41: Summary table of opportunities and challenges in relation to liability in the context of big data in the transport sector

3.13 Competition

Competition law aims at ensuring a level-playing field between undertakings and providing unrestricted choice of products and services for consumers. To this end, competition authorities are tasked with the enforcement of competition laws at supra-national (for example, the European Commission's Directorate-General for Competition ("**DG COMP**")) and at national level (national competition authorities in each country).

Most competition law dates from an era when companies had tangible assets, and when the economy was based primarily on industry and products. If they were not making nuts and bolts, they were providing a service with a more or less well-defined purpose, value, and customer.

But 'data' has now emerged as a new commodity where we have moved from the previous era dominated by tangible products, into an era based more on data and services in large quantities, data can define company value. To demonstrate this point, Uber's value is estimated at USD 68 billion not because of its business model, but because it owns the biggest pool of data about supply and demand for personal transportation.[744] Data can be bought, sold and used to leverage new product or service offerings. In the words of the EU Commissioner of Competition, Margrethe Vestager, "I*f data can help you compete, by improving your services and cutting costs, then having the right set of data could make it almost impossible for anyone else to keep up. So we need to be sure that companies which control that sort of data don't use it to stop others from competing*".[745]

Companies active in what has become known as big data – blocks of mass information that can be processed – are trading in much more than one's online shopping details. Weather, pollution levels, traffic flow, energy use, prices, values, water levels, crop yields, exchange rates, trade flows and countless more provide data points that can be collated, compounded, correlated (or not), and marketed on.

That has put competition regulators in a difficult position, as they apply the legal principles of the nuts-and-bolts world to a reality in which anything from a person's date of birth to their holiday photos and Internet browsing habits are the new corporate assets. Data can in effect be regarded as a new type of currency, thus bringing into question the applicability of financial yardsticks in measuring a multitude of things ranging from determining market power, to assessing mergers.

In view of the above, big data has become a key aspect of competition law, as the Organization for Economic Co-operation and Development ("**OECD**") noted as early as 2013, by stating that "*big data now represents a core economic asset that can create significant competitive advantage for firms*".[746] A similar report written by the OECD, on data-driven

[744] The Economist, 'The World's Most Valuable Resource. Data and the new rules of competition' [2017] The Economist 14

[745] Margrethe Vestager, 'Making Data Work for us' (Data Ethics Event on Data as Power, Copenhagen, 9 September 2016) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/making-data-work-us_en> accessed on 19 September 2018

[746] OECD, 'Exploring Data-driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"' in OECD (ed), *Supporting Investment in Knowledge Capital, Growth and Innovation* (OECD Publishing 2013)

innovation, noted that data-driven mergers have been steadily increasing over the years.[747] An example of this is the Facebook/WhatsApp merger which was given clearance in October 2014. This merger was purely data-driven in Facebook's interests which was more clearly revealed later in 2017, when the Commission fined Facebook €100 million for having given it misleading information regarding Facebook's ability to combine user data from WhatsApp with that of Facebook's.

The interplay between big data and competition law has also been expressly addressed by the European Data Protection Supervisor ("EDPS") in 2014, who stressed that if regulators fail to acknowledge the increasing importance of personal data as an intangible asset, more and more services that rely on mass personal data processing could in effect be 'ring-fenced' outside the scope of enforcement of competition rules.[748]

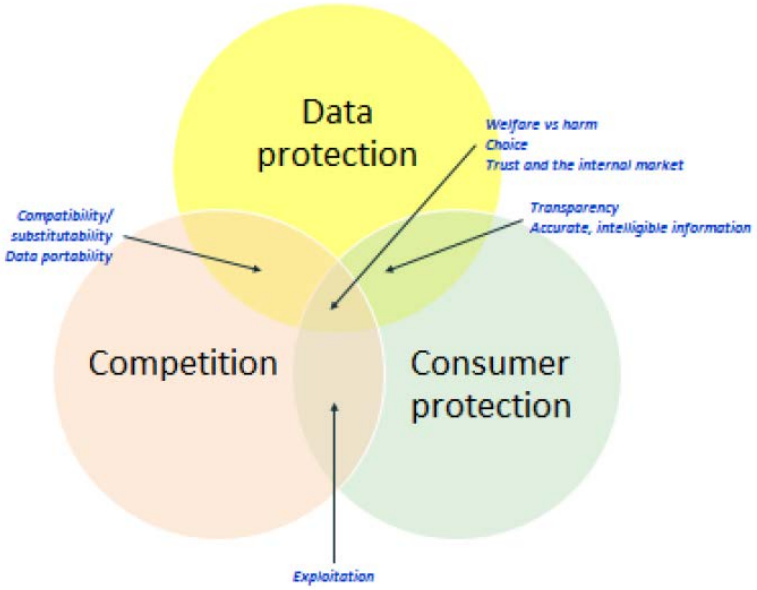


Figure 14: The interplay between data protection, competition and consumer protection (as well as [unfair] trade practice)[749]

This Section will analyse the impact of big data on different aspects of EU competition law and will seek to create more clarity on when and how the ownership or (mis)use of (big) data can give rise to competition law concerns.

Figure 15 above provides a summary and overview of the key areas of competition law that have been or potentially will be relevant to big data, which will be analysed in turn in sub-Sections 3.13.1 to 3.13.5 below.

[747] OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being* (OECD Publishing 2015) 94

[748] European Data Protection Supervisor, 'Preliminary Opinion of the European Data Protection Supervisor: Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy' (EDPS 2014) <https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf> accessed 18 October 2018

[749] Ibid 2

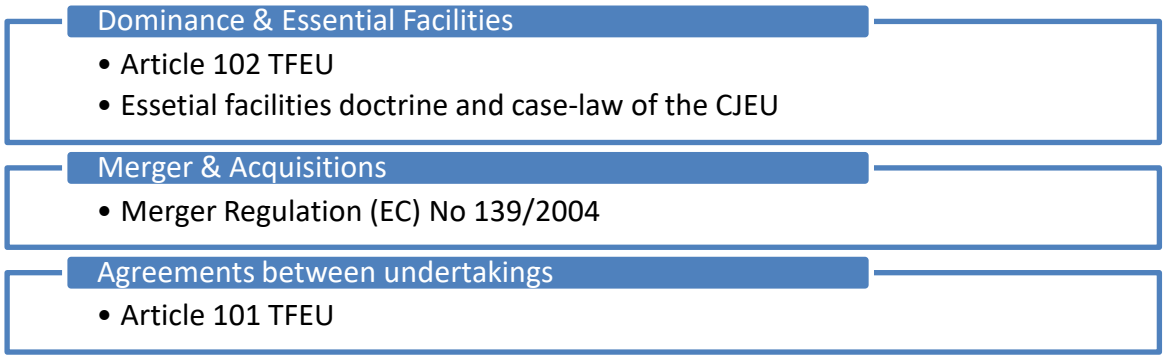


Figure 15: Overview of key areas of competition law relevant to big data

3.13.1 The role of big data in competition law analysis – the approach of different competition authorities in recent years

The first stage in the legal analysis of any anti-competitive behaviour (such as cases of anti-competitive agreements, mergers and abuse of dominant market position) is the definition of the so-called "relevant market". This allows competition regulators to identify the market operators, that is, suppliers, customers and consumers, and to calculate the total market size and the market share of each supplier with reference to the relevant product or service in the relevant area. The relevant market is defined by the following parameters[750]:

- the so-called "product market", which includes products and services which are considered by consumers to be interchangeable or substitutable;
- the so-called "geographic market", which covers the area where generally similar competition conditions prevail that are distinct from neighbouring areas.

The same approach, i.e. starting off the competition law analysis from the market definition, applies to cases where competition law and big data are interwoven. It has even been suggested by certain authors that competition cases concerning big data may not actually require a special treatment under competition law, as the main competition issues identified in relation to data-intensive industries are, in fact, not novel.[751]

In recent years, the European Commission's DG COMP, in addition to several national competition authorities have demonstrated a growing interest in big data. Although there are many ways in which competition law can affect the way in which big data is collected, used and shared, the primary areas of interest are currently the role of big data in merger control as well as the question whether the ownership of (or privileged access to) big data gives a company a dominant position. Furthermore, questions have been raised as to the efficiency of ex post measures as data can easily shift and be used for different purposes. It has been discussed whether ex ante measures could feasibly be implemented in order to address this issue, particular reference being made to the avoidance of exploitative behaviour and lack of

[750] See Commission Notice on the definition of relevant market for the purposes of Community competition law (97/C 372/03) [1997] OJ C 372/5

[751] Marixenia Davilla, 'Is Big Data a Different Kind of Animal? The Treatment of Big Data under the EU Competition Rules' (2017) 8(6) Journal of European Competition Law & Practice 370

innovation in the market. However, this would require a measurement of data as contributing to market power in a given market.[752]

Germany

Following their 2016 joint report on "Competition Law and Data", where the German competition authority and the French competition authority analysed the implications and challenges for competition authorities resulting from data collection in the digital economy and other industries,[753] the German Ministry for Economic Affairs has set up a new body, Commission Competition Law 4.0, tasked with proposing reforms to competition law to better support digital companies based in Europe. It has been asked to prepare "concrete recommendations for action on European competition law" by Autumn 2019.[754]

- assessing whether competition law should be adapted to support cooperation and standardisation efforts, including in the context of 'industry 4.0' – the term coined in Germany to describe the fourth industrial revolution that is seeing the latest technologies modernise manufacturing.
- examining whether new competition laws are needed to regulate access to data and how the development of a competitive data economy can be aligned with the requirements of data protection law.
- examining whether changes in contract law are needed to account for the increasing use of algorithms and AI for example for 'matching' and 'ranking' purposes as well as for dynamic pricing to ensure markets remain fair and competitive.

The new body has also been asked to examine how competition rules for "powerful platform companies" could be developed and whether procedural changes are needed to enable regulators to better respond to "dynamically changing digital market platforms and companies".

[752] Harry van Til, Nicolai van Gorp and Katelyn Price, 'Big Data and Competition' (Ecorys 2017) <https://www.rijksoverheid.nl/documenten/rapporten/2017/06/13/big-data-and-competition> accessed 23 October 2018

[753] Bundeskartellamt and Autorité de la concurrence, 'Competition Law and Data' (2016) <https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2> accessed 18 October 2018

[754] Bundesministerium für Verkehr und digitale Infrastruktur, 'Einsetzung der Kommission Wettbewerbsrecht 4.0' (BMVI) <https://www.bmwi.de/Redaktion/DE/Downloads/E/einsetzung-der-kommission-wettbewerbsrecht-4-0.pdf?__blob=publicationFile&v=6Its> accessed 18 October 2018

France

Following its 2016 sector inquiry[755] into the relevance of data in the online advertising sector, the French competition authority ("**FCA**") identified potential competition concerns that may trigger individual antitrust investigations in the online advertising sector[756]. The FCA's aim was to further its understanding of the digital economy and the challenges of platform regulation, notably with respect to personal data collection and use.

The Netherlands

The Dutch competition authority ("**ACM**") has recently published a study[757] into the market for online video streaming platforms following a market study[758] into online platforms.

Although the study did not find any indications of anticompetitive conduct or dominant market power, the ACM warned that it intended to keep a close eye on developments in this sector.

The study notes that online video platforms such as YouTube and Netflix have become increasingly popular over the last decade. Some of these platforms (e.g. YouTube) generate revenue by selling online advertising space to advertisers, while others use a subscription based model (e.g. Netflix). Although the ACM did not undertake a fully-fledged market definition analysis, its findings indicated that consumer data plays a key role in the offering, buying and reselling of online advertising space.

[755] Autorité de la concurrence, 'L'Autorité de la concurrence se saisit pour avis afin d'analyser les conditions d'exploitation des données dans le secteur de la publicité en ligne' (*Autorité de la concurrence*, 23 May 2016) <http://www.autoritedelaconcurrence.fr/user/standard.php?id_rub=629&id_article=2777&lang=fr> accessed 18 October 2018

[756] Autorité de la concurrence, 'Avis n° 18-A-03 du 6 mars 2018 portant sur l'exploitation des données dans le secteur de la publicité sur internet. The FCA examined various practices including bundling/tying, "low prices" and exclusivity, leveraging effects, discriminatory treatment, restrictions on interoperability, and restrictions on the ability to collect and access data' (2018) <http://www.autoritedelaconcurrence.fr/pdf/avis/18a03.pdf> accessed 18 October 2018

[757] Authority for Consumers & Markets, 'Report: Taking a Closer Look at Online Video Platforms' (Authority for Consumers & Markets 2017) <https://www.acm.nl/en/publications/publication/17575/Report-Taking-a-closer-look-at-online-video-platforms> accessed 18 October 2018

[758] Authority for Consumers & Markets, 'Online platforms onder de loep!' (*Authority for Consumers & Markets*, 21 September 2016) <https://www.acm.nl/nl/publicaties/publicatie/16332/Online-video-platforms-onder-de-loep/> accessed 18 October 2018

Italy

The Italian Department of Economics and Statistics ("**Agcom**"), Competition and Market Authority and Italian Data Protection Supervisor have begun to consider, in a cross sectoral fashion, the main problems and opportunities arising from the use of big data with particular reference to the markets (communication) and to subjects (media, political pluralism and consumer protection) within their institutional competence, with Agcom producing an interim report in June 2018.

3.13.2 Big Data & Abuse of Dominance

3.13.2.1 Overview

Abusive behaviour of companies which have a dominant position on a given market is prohibited under competition law. Such behaviour can harm competitors and customers alike, and can take the form of predatory or excessive pricing, loyalty rebates and restrictions on access to necessary infrastructure.

In practice, when a company holds a dominant position on a particular market, and uses this position to distort competition by, for example, trying to eliminate competitors, or by creating a barrier to entry for competitors, this will be considered an abuse of dominance, which is prohibited by competition law. To be dominant is not illegal; but the company has a duty to act in a manner that does not abuse that dominance which could constitute a breach of competition law.

As explained above, to determine whether a company is dominant, one must first define the relevant product market and the relevant geographic market in which the company is active.

The simple fact that a company has access to large amount 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 analysis 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 and monetisation;
- 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 authorities suggests that future cases could be based on the logic that abuse of dominance can arise from a firm's ability to 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.[759]

All of the above is also relevant for the question whether certain data can be categorised as being an "essential facility" and thus must be made available to competitors on non-discriminatory, fair and reasonable conditions. A facility or infrastructure is deemed essential when its use is indispensable for an activity in a market upstream or downstream of the market where the company holding the data is active. It is clear from the above that these requirements are likely to be satisfied only in very limited circumstances. For instance, an energy market investigation conducted by the CMA in 2016 devised a remedy requiring energy suppliers to disclose certain details of their domestic customers for a period of three years, to their competitors.[760] However, another aspect to consider could be the interplay between data as an essential facility and GDPR. Can companies share their data (if deemed as an essential facility) without infringing GDPR provisions? This is a valid question that could be explored further.

3.13.2.2 Examples

In December 2017, the German competition authority informed the company Facebook of its preliminary legal assessment in the abuse of dominance proceedings which the authority was conducting against the company.[761] Based on the current stage of the proceedings

[759] 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

[760] Competition & Markets Authority, 'Energy Market Investigation' (CMA 2016) para 233 <https://assets.publishing.service.gov.uk/media/5773de34e5274a0da3000113/final-report-energy-market-investigation.pdf> accessed 23 October 2018

[761] See Bundeskartellamt, 'Preliminary Assessment in Facebook Proceeding: Facebook's Collection and Use of Data from Third-party Sources is Abusive' (*Bundeskartellamt*, 19 December 2017)

(September 2018), the competition authority appears to assume that Facebook is dominant on the German market for social networks. The authority has publicly stated its view that Facebook is abusing this dominant position by making the use of its social network conditional on its being allowed to limitlessly amass every kind of data generated by using third-party websites and merge it with the user's Facebook account. These third-party sites include firstly services owned by Facebook such as WhatsApp or Instagram, and secondly websites and apps of other operators with embedded Facebook APIs.

According to the German competition authority's preliminary assessment, when operating this business model, Facebook, as a dominant company, must consider that its users cannot switch to other social networks. Participation in Facebook's network is conditional on registration and unrestricted approval of its terms of service. Users are given the choice of either accepting the "whole package" or doing without the service.

Apart from the above recent probe which is currently pending and its results are expected to provide much clarity on the importance of big data in competition law cases concerning abuse of a dominant position, the EU Courts have dealt in the past with the issue of whether the refusal of a dominant company to supply to other undertakings certain data which could be considered as an "essential input" for those undertakings could be considered as an abuse of a dominant position.[762]

A further recent example at EU level is the European Commission's fine (on 27 June 2017) on Google of €2.42 billion for abuse of its market dominance as a search engine by giving an illegal advantage to another Google product, its comparison shopping service. Comparison shopping services rely to a large extent on traffic to be competitive. More traffic leads to more clicks and generates revenue. By giving prominent placement only to its own comparison shopping service and by demoting competitors, Google gave, according to the Commission, its own comparison shopping service a significant advantage compared to rivals. This is because consumers click significantly more often on search results at the top of a ranking. The effects on mobile devices were even more pronounced given the much smaller screen size. By purposely placing Google's own products on the first page when returning results, Google's products were favoured over competitors' products.

Moreover, on 18 July 2018, the Commission fined Google €4.34 billion for illegal practices used to cement the dominance of its search engine. The Commission decision concerned three specific types of contractual restrictions that Google had imposed on device manufacturers and mobile network operators.[763] According to the Commission, these practices enabled Google to use Android as a vehicle to cement the dominance of its search engine.

<https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2017/19_12_2017_Facebook.html> accessed 18 October 2018

[762] See Case C-418/01 *IMS Health GmbH & Co. OHG v NDC Health GmbH & Co. KG.* [2004] ECLI:EU:C:2004:257; and T-201/04, *Microsoft v Commission [2007]* ECLI:EU:T:2007:289

[763] In particular, the Commission decision found that Google: (1) required manufacturers to pre-install the Google Search app and browser app (Chrome), as a condition for licensing Google's app store (the Play Store); (2) made payments to certain large manufacturers and mobile network operators on condition that they exclusively pre-installed the Google Search app on their devices; and (3) prevented manufacturers wishing to pre-install Google apps from selling even a single smart mobile device running on alternative versions of Android that were not approved by Google (so-called "Android forks").

3.13.3 Big Data & Mergers

3.13.3.1 Overview

In the EU, transactions which qualify as concentrations on the basis that they involve a change of control over a business and the parties involved meet the turnover thresholds laid down in the EU Merger Regulation need to be notified to the European Commission for clearance before they are implemented. Where a transaction does not qualify as a concentration under the EU Merger Regulation, it may be subject to the national merger rules of one or more EU Member States.

A report by the OECD on data-driven innovation noted that there is an increasing number of mergers between big data companies over the past years.[764] This is not surprising given that a growing number of established market players are seeking to strengthen their position by acquiring either data-rich companies or investing in start-ups collecting large volumes of data.

Whether or not the acquisition of a data-rich company raises competition law concerns will vary from case to case. The fact that one of the parties owns a large amount of raw data does not automatically give rise to concerns.

The relevance of big data in the context of merger control has been discussed only in a handful of cases. This may, however, change in the very near future, following the conclusion of the EU Commission's public consultation on the functioning of the EU Merger Regulation.[765] The consultation explores inter alia whether the purely turnover-based thresholds in the EU Merger Regulation result in an enforcement gap regarding acquisitions of data-rich companies that do not yet generate significant turnover but have a high market potential.[766] One of the solutions considered by the Commission is the introduction of a complementary “deal size threshold” which would be based on the value of the transaction (at the national level, Austria and Germany[767] have already decided to embrace this approach).

Financial yardsticks may not be sufficient in assessing mergers and bringing them under the scope of the EUMR. Transactions involving products that are offered to consumers for free (like Facebook or WhatsApp) may fall outside the Commission’s jurisdiction. For instance, Germany has included a provision in their Act against Restraints of Competition (GWB) under article 18 that states " The assumption of a market shall not be invalidated by the fact that a good or service is provided free of charge."[768] Other financial yardsticks may no longer be applicable to data-driven mergers, such as the traditional tool for quantifying market share, the SSNIP (small but significant and non-transitory increase in price) which helps define relevant markets (used both in mergers and abuse of dominance cases) and is based on price

[764] OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being*, (OECD Publishing 2015) 94

[765] European Commission, 'Consultation on Evaluation of Procedural and Jurisdictional Aspects of EU Merger Control' (*European Commission*, 2017) <http://ec.europa.eu/competition/consultations/2016_merger_control/index_en.html> accessed 19 October 2018

[766] By way of example, the proposed Apple/Shazam deal was initially notified to Austria for regulatory clearance, as the transaction did not meet the turnover thresholds of the EU Merger Regulation. Austria submitted a referral request to the Commission pursuant to Article 22(1) of the EU Merger Regulation and subsequently the Commission received requests from France, Iceland, Italy, Norway, Spain and Sweden to assess the acquisition under the EU Merger Regulation.

[767] See the amendment of section 35 para. 1 of the German Competition Act.

[768] Bundeskartellamt, Act against Restraints of Competition (Competition Act – GWB), article 18 para 2(a)

increases to determine substitutability for a product, which assesses whether customers would switch to other readily available substitute products or to suppliers located elsewhere in response to a hypothetical, small (5-10%) but permanent increase in price of the product in question.[769] This tool cannot readily be applied to cases in which one company does not demand monetary payments for its goods or services but rather uses personal data as payment for a service by users, such as certain social media companies. Competition can be affected even when a service is offered free of charge. The things that matter can be assets such as a company's' customer base, its' data or ability to innovate.[770]

Mergers between a large undertaking (such as Google or Facebook) 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, we can assert that competition law falls short of addressing the issue as thresholds and market share are not addressed by the EUMR with data in mind, but rather with financial turnover which a data newcomer may not have a lot of, thereby failing to meet the thresholds and falling outside of the scope of competition legislation.

The combination of data resulting from a merger between an established undertaking and a newcomer could lead to new data troves whose information would not be replicable by competitors, thereby making a substantial possession of that data a barrier to entry for newcomers and competitors, unless it is deemed an essential facility; but this is only assessed on a case-by-case basis. 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.

Another specific issue facing the EUMR and mergers between big data companies is the classification of data itself. The concept is very young and therefore requires further development, but as of now it can be asserted that data is in fact an asset, following the Commission's analysis of the Google/DoubleClick merger[771] which is reasonable since sometimes data is the only aspect driving a company's business, not generating financial profits. For instance, Uber's value is estimated at USD 68 billion because it controls the biggest pool of data about supply and demand for personal transportation.[772] This view is supported by Commissioner Vestager who recognized during her speech in September 2016 that data has an inherent value and that the aggregation of data can create competition concerns.[773] The largest web-based service providers such as Google, Facebook, Twitter etc. owe their

[769] Notice on Market Definition, para 17.
[770] Philip Lee, 'Competition Law Focus on "Big Data"' *Lexology* (2016) <https://www.lexology.com/library/detail.aspx?g=48c764ce-33f2-493c-9760-056008b20082> accessed 19 October 2018
[771] *Google/Double* Click (Case No COMP/M.4731) Commission Decision of 11/03/2008 [2008] OJ C (2008) 927 final, para 359
[772] The Economist, 'The World's Most Valuable Resource. Data and the new rules of competition' [2017] The Economist 14
[773] Margrethe Vestager, 'Big Data and Competition' (EDPS-BEUC Conference on Big Data, Brussels, 29 September 2016) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/big-data-and-competition_en> accessed 19 October 2018

success to the plethora of data that they possess and are able to synthesize (and they also have a frightening amount of first mover advantage in terms of insights that they are generating on top of the data). Therefore it logically follows suit that data is an asset, and more specifically an essential asset for the functioning of certain undertakings. The quality and quantity of the data they control as well as the intellectual property rights that they possess in order to make use of data in certain ways constitute a new parameter for business operations in the digital world. It is precisely these assets as opposed to financial turnover that entices larger undertakings to acquire smaller companies with a large data-asset collection. For instance, Facebook acquired WhatsApp for $21 billion, whilst WhatsApp's turnover was very moderate. Therefore, important mergers can fall under the radar.

To address these concerns, new thresholds such as consumer thresholds could be implemented concomitantly to the size of the transaction and financial turnover thresholds and would also address the issues of assessment of network effects. A solution could be adding a new requirement of a number of consumers affected in the transaction, meaning that if companies merge and it involves consumer data, then the actual number of consumers should be taken into account as the threshold for merger control. A consideration in this could also be growth in the consumers in a given online platform as a measure for merger control, subject to a percentage of the addressable market. For instance, If Germany has 15 million users and Facebook + WhatsApp are only 1% of the 15 million then it is not significant. But if they are growing at a 100% compound annual growth rate then it is significant and valid to take into consideration.

3.13.3.2 Examples

Combinations of new (or newly-discovered) and different data troves could prove detrimental to competition and raise concerns if the combination of the data makes it impossible for other competitors in the market to also reproduce the same type of information extracted from it.[774] This could result in competition concerns with barriers to entry or potential abuse of dominant position, something that was brought to light in the Facebook/WhatsApp merger which was cleared by the Commission but was reinvestigated due to new information which pointed to the fact that Facebook could in fact combine the user data of both its own and WhatsApp's database. The Commission's concern and examination was essentially whether the merger would be likely to lead to any merger-specific substantial strengthening of network effects, which could result from the combination of the separate user networks between both Facebook and WhatsApp into one substantially larger network[775] therefore making the product more valuable as more individuals use it which would entrench and increase the company's market share. Facebook had responded to this by claiming during the proceedings that 'integration between Facebook and WhatsApp would pose significant technical difficulties'[776] which therefore pointed to the impossibility of the Commission's

[774] "Bundeskartellamt, 'The French Autorité de la Concurrence and the German Bundeskartellamt Publish Joint Paper on Data and its Implications for Competition Law' (*Bundeskartellamt*, 10 May 2016) <https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2016/10_05_2016_Big%20Data%20Papier.html> accessed 19 October 2018

[775] *Facebook/Whatsapp* (Case No COMP/M.7217) Commission Decision of 3/10/2014 [2014] OJ C (2014) 7239 final, para 136

[776] Ibid para 138

concerns materializing. However later developments contradicted Facebook's claim during the proceedings as Facebook began sharing data from WhatsApp users with the rest of its business, sparking complaints from both users and rivals.[777] Essentially WhatsApp was linking its users' numbers to Facebook in order to create more targeted ads and provide better services. This can even be observed by the average individual by the fact that their WhatsApp contacts (whom WhatsApp only has through the other user's phone number) appear as suggestions in their Facebook profiles. While this issue falls significantly under the scope of data-privacy issues, it also fits quite concerningly within the scope of competition law. This still leaves the possibility of the materialization of the EC's initial concern, being the substantial strengthening of network effects (meaning that a platform becomes more attractive for consumers if the total number of consumers grows). In markets exhibiting strengthening of network effects, concentration tends to be high as a result of the market's nature due to the fact that the undertaking or platform with the highest amount of users essentially 'takes it all'. The reason is that while a particular platform grows, the network effects make it increasingly difficult for competitors to challenge the position of that platform since users find it more and more difficult to multi-home and switch to other platforms which may not have enough users to make it desirable. As such, first mover advantages can make huge differences and the competitive game may result in a winner-takes-all outcome.[778]

Data can shift very easily and current methods of analysis of market power may not be adequate in the data context. This quite clearly demonstrates the potentially competition-detrimental properties of big data in the competition parameter and the issues that can be caused by it. The concern that the Facebook/WhatsApp case demonstrates is the clearing of certain undertakings under the scope of the merger regulation, only to later have concerns arise as a result of the merger which was either not a possibility at the time that the merger took place due to network effects, or could not have been foreseen as a result of specific technological availability at the time.

In the Google/DoubleClick[779] transaction, neither the European Commission's DG COMP nor the US Federal Trade Commission concluded that the transaction would give rise to competition law concerns. Although many believed that this would be different in the Microsoft/LinkedIn[780] case, Microsoft was ultimately able to avoid an in-depth investigation by the European Commission by offering commitments. The Microsoft/LinkedIn transaction was also cleared by the US competition authority as well as multiple national competition authorities.[781] The European Commission did not approve the transaction as quickly as the other authorities and extended the review period to discuss the proposed commitments.

---

[777] Duncan Robinson, 'Facebook Faces EU Complaint over WhatsApp Deal' *Financial Times* (20 December 2016) <https://www.ft.com/content/7ed82560-f534-3d78-bc84-8043301b6c85> accessed 19 October 2019

[778] "Nicolai Van Gorp and Olga Batura, 'Challenges for Competition Policy in a Digitalised Economy' (Directorate-General for Internal Policies Policy Department A Economic and Scientific Policy 2015) 10 <http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU%282015%29542235> accessed 19 October 2018

[779] *Google/Double* Click (Case No COMP/M.4731) Commission Decision of 11/03/2008 [2008] OJ C (2008) 927 final (summary available on OJ C 184/10)

[780] *Microsoft/LinkedIn* (Case M.8124) Commission Decision of 6/12/2016 [2016] OJ C (2016) 8404 final (OJEU summary not yet available)

[781] Flavia Fortes, 'Microsoft Given Go-ahead by US Antitrust Enforcers to Buy LinkedIn' *MLex* (2016) <http://www.mlex.com/GlobalAntitrust/DetailView.aspx?cid=830831&siteid=191&rdir=1> accessed on 18 October 2018

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)[782]. 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 databases.[783] 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.

3.13.4 Big Data & Agreements between undertakings

Article 101 of the Treaty on the Functioning of the European Union ("**TFEU**") "*prohibits all agreements between undertakings, decisions by associations of undertakings and concerted practice…which have as their object or effect the prevention, restriction or distortion of competition in the common market*." It is clear that this cartel prohibition applies to both direct and indirect price fixing.

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 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[784]. 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 artificial intelligence 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 where there exists price-fixing -thereby having detrimental effect on the market- but which cannot be proven following the traditional definitions of collusion despite the definition of 'agreements' itself being rather broadly construed under the relevant legislation. Article 101(1) stipulates 'All agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention,

[782] *Apple/ Shazam* (Case M.8788) Commission Decision of 06/09/2018 [2018] (OJEU summary not yet available)

[783] See Commission press lease IP/18/5662 'Mergers: Commission Clears Apple's Acquisition of Shazam' (6 September 2018) <http://europa.eu/rapid/press-release_IP-18-5662_en.htm> accessed on 19 October 2018

[784] There are few precedents concerning competitors using algorithms to fix pricing. An exception is the 2016 CMA investigation into price fixing between two competing online sellers Trod Ltd (Trod) and GB eye Ltd (GB eye) who each sold posters and frames online, including on Amazon Marketplace. Rather than compete, however, Trod and GB eye agreed not to undercut each other's prices in certain circumstances for particular products sold through Amazon's UK website. To give effect to the agreement, both sellers used automated re-pricing software to monitor and adjust their prices, and ensure that neither was undercutting the other. The CMA found both companies liable for breaching competition law. Online sales of posters and frames, Case 50223, Decision of the CMA, dated 12 August 2016

restriction or distortion of competition within the internal market.'[785] The issue, however, is that in certain circumstances it may be difficult or near impossible to prove human input in the collusion due to the fact that it is automated by an algorithm and acts independently of human agreements. This, in turn, is not covered by Article 101(1) stemming from the fact that the requirements for an agreement to constitute anticompetitive behaviour are not satisfied between two parties; rather, it is automated and not captured by the traditional meaning of agreements as stipulated in Article 101(1).

The UK's Competition & Market Authority ("**CMA**") in a report to the OECD[786] 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.

3.13.5 Big Data and Transport: competition law issues

The use of big data in transport has significant potential. Big data is (or could potentially be) involved in all types and modes of transport, be it traditional modes of transport (such as motorcars, rail, urban transportation, aviation) or in novel concepts in the transport sector (such as transportation apps, electric vehicles, autonomous vehicles etc.). Examples of the effects of big data in the transport sector have been discussed in detail in the previous Sections of this Deliverable D2.2 and in the previous Deliverables of the LeMO Project, and include effects on road, air, water and rail transport.[787]

Further, the role of data and analytics in transportation is not limited to urban centres. The logistics industry will benefit from applied mapping technology and algorithms. Location-based services use GPS and other data to pinpoint where a person (vehicle or device) is situated in real time. The widespread adoption of GPS enabled smartphones, and the role of digital giants such as Google and Apple in driving location-based services forward for billions of smartphone users, cannot be underestimated.

[785] TFEU, art 101, para 1

[786] OECD, 'Algorithms and Collusion – Note from the United Kingdom' (DAF/COMP/WD(2017)19, OECD 2017) <https://one.oecd.org/document/DAF/COMP/WD(2017)19/en/pdf> accessed 19 October 2018. See 4.2 written contribution from the United Kingdom, 127th OECD Competition Committee on 21-23 June 2017

[787] Carl-Stefan Neumann, 'Big Data versus Big Congestion: Using Information to Improve Transport' (*McKinsey & Company*, July 2015) <https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/big-data-versus-big-congestion-using-information-to-improve-transport> accessed 19 October 2018

McKinsey Global Institute ("**McKinsey**") estimates[788] that by 2030 mobility services such as ride sharing and car sharing could account for more than 15-20% of total passenger vehicle miles globally. For personal transportation, ride sharing services use geospatial mapping technology to collect crucial data about the precise location of passengers and available drivers in real time.

Platforms such as Uber and Lyft have been able to expand rapidly without acquiring huge fleets themselves, making it easy for drivers to put their own under-utilised assets to work. Platforms matching supply and demand have already set off ripple effects in urban transportation. With the advent of autonomous vehicles, the wave of change could accelerate as these vehicles have higher utilisation rates. The cost of urban transportation could plummet. They could also transform energy markets by enabling smart grids to deliver distributed energy from many small producers.

The important, and constantly evolving role of big data in the digitalized 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 (smaller existing players or potential new entrants in the market). An example of exploitative behaviour as a result of data possession could be perceived in the possession of training data by certain companies, whereby they use this training data to mature and refine their algorithms before placing them on the market. The amount of data they possess for training algorithms means that other companies have to then buy this in order to properly train their own algorithms. As we will explore later in this sub-Section, algorithms have become an essential feature of companies' business operations. Thus, databases containing large amounts of data can translate into a market advantage for training algorithms. It is easy to imagine a scenario where companies such as IBM have to buy data sets for training their algorithms, from companies like Facebook, which controls a plethora of data.

A scenario in which large amounts of data confer a market advantage could be envisaged both in "traditional" markets (such as in the field of civil aviation, where airlines hold vast amounts of data on their customers), and digitalized markets (such as in the markets of mobile apps).

We explore such a scenario in a digitalized market below.

---

[788] Nicolaus Henke and others, 'The Age of Analytics: Competing in a Data-driven World', (McKinsey & Company 2016) <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx> accessed 19 October 2018

Competition in the transport sector – Example 1

The District Court for the Northern District of California was recently called to rule upon a dispute between urban transportation apps Uber and Lyft.

It appears from publicly available information[789] that 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. Upon the basis of the information collected through "Hell", Uber in turn targeted the so-called "double-appears" (i.e. drivers who were identified as using both apps) and offered to them, without their knowledge, certain benefits (such as sending more riders through the app and awarding special bonuses for meeting a certain number of rides), so as to ensure that they would prefer Uber over Lyft.

The possible exclusionary effect of this behaviour is clear: by offering so-called "reverse rebates" to targeted costumers and by engaging to overbuying, Uber's main competitor, Lyft, was unable to compete on the market (as more and more drivers who were considering using Lyft were abandoning that app in view of the benefits enjoyed when preferring Uber).

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.

However, competition law issues potentially involving big data in the transport sector are not limited to potential abuses of dominant position by large companies. Mergers between data-rich companies belonging to the broadly-defined "transport sector" (such as, for example, companies providing airline or rail ticketing or booking services) could raise similar competition law issues as the ones discussed in sub-Section 3.13.4 below. Another area of contention is price-fixing algorithms as has been mentioned at the beginning of this sub-Section.

Price-setting algorithms have become a widespread phenomenon amongst companies. Increased market transparency has led to pricing algorithms becoming an essential feature of a company's business operations. This scenario of pricing algorithms proves very challenging for current competition legislation and authorities to capture within the current legislative parameters. In such a scenario, the algorithm essentially does all the work as a result of its collection of market data and there is virtually no human input despite the existence of horizontal collusion on the market. To date there has been no case law on the matter of predictive algorithms. Rather, the predictive algorithm scenario is more a direction in which technology and the aggregation of big data are heading in the not-so-distant future. Namely,

[789] Ignacio Herrera Anchustegui and Julian Nowag, 'Buyer Power in the Big Data and Algorithm Driven World: the Uber and Lyft Example', (Competition policy international 2017) <https://www.competitionpolicyinternational.com/wp-content/uploads/2017/09/CPI-Anchustegui-Nowag.pdf> accessed 19 October 2018

in this scenario there would be no agreement amongst competitors, and price setting and market monitoring would instead be left entirely up to the algorithms which would act autonomously in matching competitors' prices on the market. This was also an issue and concern that was brought up during the German Competition Authority's 18th Conference on Competition in March 2017.[790]

Tacit collusion or parallel behaviour can be observed in oligopolistic markets, where there is a fine line between proper competition and the existence of a monopoly.[791] However the guidelines on the application of Article 101 do not expressly state that tacit collusion is illegal. It seems from EC case law that tacit collusion is illegal when it leads to concerted practice, which is different from agreements to collude.[792] However, the unilateral use of algorithms in order to set specific prices and monitor prices on the market, as well as companies' intelligent and rational responses to their competitors (in conduct or pricing) in oligopolistic markets as tacit collusion do not bring the case under the scope of Article 101, which was asserted by the ECJ in Gerhard Züchner v Bayerische Vereinsbank AG.[793] Therefore we can see that whilst explicit collusion, stemming from agreements between competitors in the meaning of Article 101 is quite clearly illegal, tacit collusion or parallel behaviour does not constitute an infringement of Article 101, unless it leads to concerted practice.

Algorithms could actually widen the possibilities and conditions for tacit collusion and parallel behaviour to occur in the market. Undertakings can use computer algorithms, programmed to maximize turnover, which monitor competitors' prices not as a result of exchange of information (which could be considered anticompetitive behaviour) but rather as a result of real-time prices being readily available due to increased market transparency, which may result in the anticompetitive effect of a cartel without having the necessary features to prove it's illegality under the current legislative framework. The algorithms conduct the market and pricing calculations for humans, making the strategy that much more precise and giving the certainty to competitors using similar algorithms that the other players on the market entrust their pricing to their own algorithms, thereby eliminating the possibility of secretive dealings that could undermine the 'common policy'. Furthermore, as more undertakings begin using these types of algorithms, then by definition market transparency increases exponentially since the users must post their prices online and market data becomes more plentiful and easily accessible. Companies entrust the same actions described in the previous paragraph to computers and algorithms due to their precision and speed. This is quite logical if one imagines a scenario in which he is gambling against a computer who is able to quickly and precisely calculate all the possible odds of a given gambling game. It is unquestionable that the computer would be much more successful than the human. Pricing algorithms in oligopolistic markets can match competitors' prices instantaneously as soon as they are published and dissuade discounting prices whilst at the same time matching a price increase

[790] Margrethe Vestager, 'Algorithms and Competition' (Bundeskartellamt 18th Conference on Competition, Berlin, 16 March 2017) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/bundeskartellamt-18th-conference-competition-berlin-16-march-2017_en> accessed 19 October 2018

[791] Richard Whish and David Bailey, *Competition Law* (Oxford UP 2015) 559

[792] Case 48/69 *Imperial Chemical Industries Ltd. v Commission of the European Communities* [1972] ECLI:EU:C:1972:70, para 64

[793] Case 172/80 *Züchner v Bayerische Vereinsbank AG* [1981] ECLI:EU:C:1981:178, Para 22 (Reference for a preliminary ruling)

that can prove sustainable, thus leading to an overall and steady rise in market price above the competitive equilibrium which can be deemed harmful to consumers. Therefore, as more companies in a given market switch to the use of algorithms to monitor and set their prices, it becomes the norm and any company not using such an algorithm will find itself at a severe competitive disadvantage due to the effectiveness of computers vis-à-vis humans as demonstrated in our gambling scenario. In this way pricing algorithms substantially increase market transparency, facilitate parallel behaviour and because what was previously a concern in oligopolistic markets, but sanctioned under the exchange of information, which was a steady increase in market price above competitive levels, ergo harm to consumers as a result of tacit collusion. Whilst tacit collusion is not sanctioned under Article 101, it still has anticompetitive effects, and the EUMR seeks to block mergers which could result in tacit collusions or firms which enjoy 'collective dominance'. [794]

Algorithms can now essentially eliminate the existence of information exchange and any types of agreements -which are sanctioned under Article 101- to increase prices and maximize company profit, and instead do it themselves thereby only amounting to tacit collusion but having the same detrimental effect on the market as cartels due to the increase of prices above competitive levels. The most that could be brought to light would be the knowledge by companies adopting the algorithm that tacit collusion would be a likely outcome if other players on the market adopted the same algorithm. But legal certainty is very murky in this area and therefore we can see how current legislation falls short in being able to sanction these types of behaviour, achieve the goals and uphold the fundamental principles of competition law.

As we will examine in the example below, pricing algorithms may very well facilitate and/or encourage virtual cartels and cause detrimental effect to the market, but all the while being difficult to bring under current competition legislation. Thus, it cannot be excluded that potential anti-competitive agreements between companies active in the transport sector could be affected by the existence of big data and machine-learning mechanisms and algorithms.

[794] See cases: *Karstadtquelle/Mytravel* (Case No COMP/M.4601) Commission Decision 04/05/2007 [2007] SG-Greffe(2007) D/202716; *JCI/Fiamm* (Case No COMP/M.4381) Commission Decision 10/05/2007 [2007] OJ C(2007)1863 final; *TUI/First Choice* (Case No COMP/M.4600) Commission Decision 04/06/2007 [2007] SG-Greffe (2007)D/203384, *TomTom/Tele Atlas* (Case No COMP/M.4854) Commission Decision 14/05/2008 [2008] OJ C(2008) 1859.

Competition in the transport sector – Example 2

Another more theoretical scenario using Uber as an example again is with the use of pricing algorithms.

This refers specifically to Uber's price setting algorithm which works by setting the price for the individual drivers operating under Uber in each specific city that it is present in. Uber's algorithm functions in its entire network, operating in 300 cities worldwide, and acts as a price fixing hub for its drivers, the spokes. Uber's algorithm can be considered an algorithmic monopoly due to the fact that the prices exhibited by the automated price fixing algorithm show a perceived competitive price as opposed to an actual fair market price. This follows the logic that the main premise and idea of a market is that the fair market price is determined by the willingness of both the buyer to buy the product, and the seller to sell the product. However, in Uber's scenario, Uber is neither the buyer nor the seller. Rather, it is the broker in this situation and acts as a liaison between the buyer and seller who do not nor cannot negotiate the price. But Uber is not transparent, and its algorithm is not regulated; it shows the buyer only what it decides to display and determines within its market the areas to implement a particular surged price as well as when to implement it and for how long.

The algorithm in question has not thus far caused any competition concerns since customers can easily switch to another service such as Lyft or simply taxis if they do not agree with the price. However, the rate at which Uber and its services are growing as well as the number of users that are joining adds up to a steady increase of 1,100-1,500 new riders per month since 2014.[795] This may lead to unfavourable influences on the market price in the near future, leading to potential competition concerns as a result of Uber's algorithm. This could occur in a scenario where Uber is the main market player in a particular city, such as New York where taxi companies are losing customers due to poor services and comparatively high prices, which could very realistically soon not impose any competitive restraint on Uber, thereby eliminating competition on the market. Considering the fact that Uber drivers are self-employed following the company's premise and their landmark settlements in California and Massachusetts,[796] then drivers do not actually compete with each other and it is common practice for drivers to steer clear from areas where they see that other drivers are operating so as not to compete with them. In this scenario, Uber's price setting algorithm is set to attain substantial market power in setting prices. Whilst this type of vertical agreement is not illegal under competition legislation and does not attempt price fixing, the concern arises when several similar vertical agreements arise in the market where rivals rely on one single algorithm (the hub) to set the price for the spokes which effectively limits competition and increases overall prices.

[795] Artyom Dogtiev, 'Uber Revenue and Usage Statistics' (*Business of Apps*, 23 July 2018) <http://www.businessofapps.com/uber-usage-statistics-and-revenue/> accessed 19 October 2018

[796] Mike Isaac and Noam Scheiber, 'Uber Settles Cases with Concessions, but Drivers Stay Freelancers' *The New York Times* (21 April 2016) <https://www.nytimes.com/2016/04/22/technology/uber-settles-cases-with-concessions-but-drivers-stay-freelancers.html?_r=1> accessed 19 October 2018

There is an even more challenging scenario for competition authorities in Uber's case. In the case that the algorithm's design was not meant to fix prices and facilitate collusion, it may still have an appreciable and detrimental effect on market price. It is therefore questionable whether these cases would be brought under the scope of Article 101. For instance, insofar as the drivers would be agreeing prices such as surge pricing or base prices amongst themselves, then indeed this would easily fall under horizontal collusion and price fixing as illegal by object. But there are instances in which any sort of agreement or collusion may be difficult or near impossible to spot despite adverse effects on market price. Let us return to the scenario of Uber in New York enjoying a vertical relationship between the hub (the algorithm) and its spokes (the drivers). The initial drivers that decided to drive for Uber are not in fact agreeing to fix prices or collude as a result of using Uber's algorithm, while there is still competition on the market. However, after it becomes the main market player in New York, where it uses its price setting algorithm and does not experience any significant competitive constraints by other taxi companies, the question arises whether the subsequent drivers that decide to join Uber's platform after seeing the effect on the market and utilize the algorithm are part of a traditional hub-and-spoke conspiracy since they were aware of the monopolistic effect that Uber exhibited, and agreed to join Uber and reap the benefits nonetheless, under the premise of knowledge of eventual anticompetitive effects resulting from joining. This type of case could prove difficult for the current stipulations of Article 101 to capture since the intent of the algorithm is not to facilitate collusion, but it may still have the effect of raising market price, thereby making it questionable of proving the algorithm developer's or even the driver's liability as a result of a lack of intent. Therefore, there exists a fine line between situations in which an algorithm is anticompetitive, and when it is not. If there is intent to collude, then there is knowledge and it is illegal. Conversely, if there is no intent or intent cannot be proven, then it is legal, regardless of the effects (appreciable or not) that it may have on the market since in order to establish that a practice was anticompetitive by effect, there must first be an agreement. The growth of these platforms and their use of algorithms may have appreciable effect on market prices which may look like horizontal collusion but which fail to meet the conditions and features of a hub-and-spoke conspiracy since the drivers act independently when entering into any agreement with Uber and to drive for Uber, thereby eliminating any notion of horizontal agreements amongst them.[797]

[797] Ariel Ezrachi and Maurice E. Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-driven Economy* (Harvard UP 2016) 53

Competition in the transport sector – Example 3

Airlines have in their possession large amounts of data on their customers, ranging from their preferred itineraries, place of residence, dietary requirements and seat preferences to 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 suggested[798] 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, it is accepted that 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 (i.e. the expected maximum amount that a specific customer is willing to pay); the price competition (i.e. the price that the competitors are willing to offer to the same customer, and which is normally freely available on the internet); 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 quickly 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.[799]

Competition authorities could be faced with 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.

Delving deeper in to the concept of algorithms for setting companies' business strategy, we will examine another more hypothetical scenario that has been on the OECD's and several NCAs radars, which may very well materialize with the emergence of new algorithms and products, such as Amazon's Internet of Things which can track its products in various markets.

In this scenario, the algorithm not only considers and monitors competitors' pricing, but also factors in a multitude of other market features that could be synthesized to establish the best possible strategy for the firm and maximize profits. In this scenario, the algorithms undergo two key transformations, being that: 1) they are able to monitor the entire market and possess a much larger processing power and ability, so as to have an omniscient view of the entire marketplace; 2) they are able to surpass their original programming and adapt to

[798] 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

[799] Ariel Ezrachi and Maurice E. Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-driven Economy* (Harvard UP 2016) 72

different situations, as well as learn on their own and later apply learned concepts to company strategy and the market situation. This is effectively the use of artificial intelligence and is a concept that has been subject to increasing development and use in recent years. Artificial intelligence and algorithms can predict when a customer is ready to buy; a jet engine needs servicing; or a person is at risk of disease.[800] For instance, in digital marketing, Artificial Intelligence and machine learning can understand human behaviour to the extent where not only are big data sets analysed, segmented and filtered, but the meaning is also derived from them.[801] A good example is Google's recent AI software program that is able to learn and actually develop AI software itself, without human input or additional programing.[802] These two new features together can spell potential competitive disaster on the market. Being able to have an omniscient view of the market and its data as well as having the capability to learn and adapt may even bring about undetectable forms of collusion. Algorithms would be able to track products along the entire supply and distribution chain which is a feature that we can already see with Amazon's 'Internet of Things' initiative, whereby products are connected to the internet and are trackable.[803] In this scenario, again, the algorithm seeks to maximize profit without undertaking any illegal activity in the form of agreements with other companies and their own algorithms, or market allocation. However, the difference between this scenario and the previous one is that the algorithm here does not attempt to reach tacit collusion by fluctuating prices, but rather aims to gather and process information about the market in order to conduct the optimal strategy for the company, all the while doing so independently.[804]

In the case of omniscient algorithms, they evaluate every single competitive manoeuvre by other players on the market and determine the best and most efficient course of action. It can monitor customers, products, market shares and many more aspects that it can combine to give its company a competitive advantage. This algorithm will detect and counter not only any price discount by a competitor but also other competitive initiative such as selling in the market of the company utilizing the algorithm. This in turn could result in the complete disincentivization of any competitive initiatives to begin with by limiting the expected gains from such a deviation, thereby slimming down the competition on said market. Even if a few numbers of market players possess this type of algorithm in the first place, a 'survival of the fittest' principle will kick in and other players will soon be able to observe the overarching benefits of possessing such an algorithm which will likely lead them to adopt it on their own, as was examined in the previous scenario with the gambling metaphor and the obvious superiority of computers over humans in processing power. Eventually, competitors without this algorithm will likely face a stark inability to keep up with and compete with players in possession of this omniscient algorithm and will likely retreat from the market altogether

[800] The Economist, 'The World's Most Valuable Resource. Data and the new rules of competition' [2017] The Economist 7

[801] Tara Thomas, 'Artificial Intelligence in Digital Marketing: How Can It Make Your Life Easier?' (Zeta, 25 April 2017) <https://zetaglobal.com/blog-posts/artificial-intelligence-in-digital-marketing/> accessed 19 October 2018

[802] Tom Simonite, 'AI Software Learns to Make AI Software' (*MIT Technology Review*, 18 January 2017) <https://www.technologyreview.com/s/603381/ai-software-learns-to-make-ai-software/> accessed 19 October 2018

[803] Alex Konrad, 'Amazon Jumps Into Internet Of Things Frenzy With New Cloud Platform For Devices Like Car'' *Forbes* (08 October 2015) <https://www.forbes.com/sites/alexkonrad/2015/10/08/amazon-jumps-into-internet-of-things-frenzy-with-new-cloud-platform/#4198e2eede4d> accessed 19 October 2018

[804] Ariel Ezrachi and Maurice E. Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-driven Economy* (Harvard UP 2016) 72

which will result in very high entry barriers for any potential newcomers. What ultimately results is a system of predictability whereby algorithms will be able to analyse and assess what course of action other companies' algorithms will take and without even the intention of creating tacit collusion, parallel behaviour may materialize as a result of the assessment by the algorithm that such behaviour would be the most profitable.

Competition in the transport sector – Example 4

When a competitor tries to break into another competitor's market by selling to their customers, the other retaliates by doing the same and this takes the form of a quid pro quo or 'eye for an eye' scenario which is what healthy competition looks like where each competitor attempts to win over consumers in the rival's market, usually by improving their product or by making it more appealing to consumers. Consumers therefore benefit from this quid pro quo scenario.

However, with the omniscient algorithm, the algorithm may know that this is not an ideal strategy for profit maximization. This coupled with its ability to learn and perceive/process vast amounts of data in the market as well as track products and/or users means that it can opt for the optimal long run strategy in any given situation. With this type of transparency and the ability to track, algorithms may decide not to retaliate and instead leave the status quo and not engage in the competition. The scenario of product tracking can already be seen with undertakings selling physical products which they are able to track such as Amazon's Internet of Things.

What the omniscient algorithm scenario gives rise to is virtually full market transparency resulting in the ability for competitors to know and see exactly what their competitors are doing and how they are operating. This knowledge gives them many more possibilities to coordinate on several levels. The omniscient algorithm may know for instance that a given rival has certain key consumers/customers that it sells to and may refrain from targeting those customers with its own products/services so as not to engage in a price war, due to the fact that it would know that the rival would retaliate in a similar manner and thus neither would benefit. The company is therefore restricting its own competitiveness in the market. With such an omniscient algorithm, tacit collusion and parallel behaviour are made all the easier to sustain.

If we consider ordinary collusion of any sort amongst competitors, there exists a great degree of distrust amongst company managers or whoever it was that entered into the agreement in the first place. Humans can sometimes be unpredictable and efforts to undermine the agreement for one's own gains are likely to take place and if undetected could ruin the situation for all the undertakings engaged in the agreement. This also means that it is much easier for rivals to observe and keep an eye on a small number of competitors than a large number (it must also be noted that such colliusions work better on a highly concentrated market since players with a larger percentage of market share have more incentive to sustain the parallel behaviour because any loss suffered by them would be more significant if they own 20% than if they own 5%). However, with the omniscient algorithm that problem is effectively removed since the algorithms can process and monitor a much

larger range of competitors and data on the market. Therefore, parallel behaviour and tacit collusion are sustainable on an immense scale that was otherwise unprecedented, thanks to artificial intelligence.

The algorithm would engage in behaviour that is entirely natural when seeking to maximize company profits, but which reduces competitiveness in the market. However, in this scenario there is absolutely no evidence of intent since the humans in the company are not even aware of any type of collusion present on the market. The initial programming of the algorithm did not intend for it to collude. Rather it learned this process on its own through the data collected on market behaviour and the plethora of other factors that it may take into account in choosing optimal strategies for profit maximization. Here, there is no human knowledge of eventual tacit collusion as a result of the use of the algorithm. Furthermore, under the definition of agreements by the Commission in case law, this type of behaviour would not fall under Article 101 since we cannot observe any type of agreement between competitors, nor any intent to collude. Case law in Case T-141/89 Tréfileurope Sales SARL v Commission demonstrates this:

'For there to be an agreement within the meaning of [Article 101(1)], it is sufficient for the undertakings in question to have expressed their joint intention to conduct themselves in the market in a particular way.'[805]

3.13.6 Summary

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 legal literature on this topic. 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.

We have considered here three main areas in which competition law may have an impact on the use of big data. As regards the potential application of the abuse of dominance or essential facilities, it is clear that mere access to large amounts of data does not equate to having a dominant market position. This is due to the fact that competitors often have access to the same data or can obtain data from other resources (such as data brokers) and, in addition, can analyse and monetise data using their own applications and resources.

[805] Case T-141/89 *Tréfileurope Sales SARL v Commission* [1995] ECLI:EU:T:1995:62, para 7

In the context of algorithms and artificial intelligence, this is a novel topic in which issues could very perceivably arise in the near future. Algorithms can lead to the emergence and sustainability of virtual cartels and market-sharing. Whilst in the past tacit collusion could exist in smaller concentrated markets with relatively few players, such as the maxim follows: tacit collusion is "frequently observed with two sellers, rarely in markets with three sellers, and almost never in markets with four or more sellers."[806], it now can exist on a much larger scale with a greater amount of players on the market thanks to algorithms. As more companies are forced to use algorithms or face being driven out of the market, the widespread use of algorithms given the increased transparency in the market and enhanced speed of processing could very well lead to a large number of markets being susceptible to tacit collusion leading to adverse effects on competition. We have seen that the outcomes of using algorithms can cause increased market transparency, increased certainty (for competitors), greater concentration and disincentives to players or maverick in discounting prices because it would be unprofitable for them to do so. These factors can lead to the sustainability of parallel behaviour in markets in which such behaviour was previously unsustainable. We will be left with tacit collusion leading to competitive harm, but which the current stipulations under Article 101 will not be able to tackle. Thus, in this area, improved legislation may be warranted to capture such instances. This could take the form of bringing tacit collusion under the scope of Article 101 in the presence of all the features of a cartel.

In dealing with algorithm cases, the Commission would do well to consider additional factors such as whether the companies involved 1) tacitly agreed to use algorithms with the aim of lessening competition on the market; 2) whether the business decisions they conducted were sound and whether there was a reasonable/legitimate reason behind them; 3) had knowledge that their actions could likely cause competitive harm to the market and amount to anticompetitive behaviour. These considerations can be seen in practice in another case and area of the market, namely investment banking, where the US Securities and Exchange Commission (SEC) investigated Athena Capital Research in 2014.[807] The SEC found that Athena's computer algorithms were found to have manipulated stock prices and caused detrimental harm to prices, of which the company's employees were well aware of. The company was therefore fined $1 million.

The aforementioned provision drafted after the US model and considerations would be a viable and effective addition to the Commission's antitrust toolbox in that the 'unfair practices' would not have to be specifically defined, and would therefore fit in line with the EU's competition law policy of maintaining broad stipulations so as to allow the Commission to apply reasonable means of attaining competition's law overall objective. Essentially the interpretation of this provision could evolve with the changing digital market and allow precedents to tailor the meaning of unfair practice to big data company and algorithm cases.

In the context of mergers and acquisitions, where merger control regimes have traditionally been based on market share and sales figures, consideration is being given to whether there is

[806] Jan Potters and Sigrid Suetens, 'Oligopoly Experiments in the Current Millennium' (2013) 27(3) Journal of Economic Surveys 439

[807] Decision of 16 October 2014 in case 'United States of America Before The Securities and Exchange Commission In the Matter of Athena Capital Research, LLC' <https://www.sec.gov/litigation/admin/2014/34-73369.pdf> accessed 19 October 2018

a need to add "deal-size thresholds" with a view to catching high value acquisitions currently falling outside merger control due to small or insignificant turnover figures of the target. While Germany and Austria have recently altered their merger control regimes to include "deal size" transactions above certain values, the European Commission is currently contemplating whether a similar approach is warranted at EU level. Furthermore, purely financial yardsticks may not reflect the reality of the market in assessing mergers. Concomitantly, it could be useful for the Commission to incorporate other thresholds alongside financial ones, such as consumer thresholds, which would give it a better view of the circumstances under which the merger is taking place.

What makes big data useful is the potential to analyse data sets in real time on a huge scale using powerful processors and algorithms. This has a particular relevance in transparent markets where, for example, companies are required to publish detailed pricing information to customers (such as transport). In such sectors, computer algorithms can quickly detect a price reduction by a rival and effectively deprive that rival of any significant increase in sales. Rationally, this would lead to fewer firms discounting. Thus, machines using algorithms with self-learning capabilities may render obsolete illegal cartel activities such as price fixing in smoke-filled rooms. The difficult legal question here is where to draw the line between actions that can be attributed to humans and those that may arise through machines using algorithms employing artificial intelligence technology such as deep learning.

With regard to the transport sector, in view of the important role of big data in that sector, we have discussed the competition law issues that could arise with respect to companies belonging in the broadly-defined "transport sector", providing specific examples, including both real-life examples and hypothetical situations that could potentially be envisaged in the near future.

The table below summarises the challenges and opportunities identified throughout this Section:

Opportunities in relation to competition in the context of big data in the transport sector	Challenges in relation to competition in the context of big data in the transport sector
The opportunity to acquire data-rich companies should not cause competition issues under merger control regimes. 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.	Recent changes to national merger control rules (Germany, Austria) to take account of "deal size thresholds" rather than turnover figures could result in more data-rich mergers requiring prior merger clearance. A similar approach at EU level is under consideration. Additionally, the Commission could incorporate new data-type thresholds under the EUMR to better assess the merger from the perspective of digital markets.
The transport sector has always collected and analysed large quantities of data, such as data from timetables, traffic news and air schedules. Big data allows this to be used to create more efficient and smarter transport systems for people and freight and increases the scope to monetise and sell data for new and innovative services.	Competition authorities are investing considerable resources in studies and market investigations with a view to understanding whether the existing competition rules are fit for purpose or need to be changed.
New app developer opportunities such as short journey planning for multiple modes of transport in major cities by collating open data feeds, real-time traffic information based on crowd-sourced data from smartphones and vehicle GPS data, real-time public transport journey planning by combining public transport data with information crowd-sourced from users through smartphones, suggested driving routes based on traffic information crowd-sourced from users through smartphones.	Competition compliance programmes may need to be examine whether price-fixing could result from the coordinating effect of algorithms and associated risks.

Table 42: Summary table of opportunities and challenges in relation to competition in the context of big data in the transport sector

3.14 Conclusion

This Deliverable identified and examined various legal issues that are relevant to the production of, access to, linking of and re-use of big data in the transport sector and discussed the challenges and opportunities that may arise in this respect.

The analyses performed in this Deliverable demonstrate that the legal framework as it stands does not encourage the uptake of big data, including in the transport sector, and that for all legal issues examined there is – to a certain extent – room for improvement. The improvements suggested by this Deliverable are variable between the different legal issues and range from avoiding restrictive interpretations by the relevant authorities or courts, over soft law measures (such as guidelines and codes of conduct), to regulatory intervention at EU level.

We briefly summarise below the suggested improvements for each legal issue examined in this Deliverable:

- **Privacy and data protection**: Any guidance or administrative/judicial decision regarding personal data protection must carefully take into account all interests at stake and avoid over-simplistic reasoning and illustrations. In this respect, it is essential to keep in mind Recital 4 of the GDPR which stipulates that (i) the right to the protection of personal data is not an absolute right; (ii) it must be considered in relation to its function in society and be balanced against other fundamental rights; and (iii) the principle of proportionality must be taken into account. Failing to do so would necessarily impede the development of disruptive technologies, including big data, and prohibit the emergence of a true data economy.
- **(Cyber-)Security and Breach-related obligations**: In light of the need for clarification of how to comply with security requirements in practice, particularly taking into account the specificities of big data applications, the authors of this Deliverable encourage the creation of guidance by authorities such as ENISA and the drawing up of certification mechanisms, seals, marks, and codes of conduct regarding (cyber-)security, allowing companies to comply with their legal obligations and demonstrate compliance.
- **Anonymisation and pseudonymisation**: 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 such concepts, so that legal and technical professionals may share a common understanding on the consequences of the use of such techniques. The creation of codes of conduct and similar initiatives is indispensable to support organisations 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. 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.

- **Supply of digital content and services (personal data as counter-performance)**: Personal data as counter-performance for the supply of digital content is *per se* not an undesirable concept in the context of big data, including in the transport sector. However, the legalisation of the practice and its inclusion in a legal instrument at EU level generate various practical concerns around the obligations concerning data and require further clarification. The subject calls for the establishment of guidelines, or similar initiatives, to assist the suppliers of digital content and services, and provide greater legal certainty.
- **Free flow of data**: Whereas the elimination of localisation requirements under the Free Flow Regulation is likely to lead to significant cost reductions for cloud storage and processing, necessary for big data analytics services in the transport sector, uncertainty remains regarding the exact scope of application of the Regulation. For instance, there is currently a lack of clarity about which legal instruments apply to mixed datasets composed of both personal and non-personal data, which will very often be the case for big datasets in the transport sector. Further guidance from the European Commission on those issues is encouraged.
- **Intellectual property in big data environment**: Many different actors in the big data analytics lifecycle may try to claim intellectual property rights in (parts) of the datasets intended to be used and may therefore try to exercise the exclusive rights linked to the intellectual property right concerned. Any unreasonable exercise of rights may stifle data sharing and thus innovation through big data, including in the transport sector. Given that this is mainly due to the inherent nature and purpose of intellectual property rights, and taking into account that intellectual property rights protection may at the same time provide an incentive for stakeholders to engage in data sharing for big data purposes, no specific suggestions for improvement have been identified at this stage.
- **Open data**: The proposed expansion of the PSI Directive's scope to include public undertakings is a significant development for the transport sector, where services are often provided by public undertakings. There is however an inherent tension between the PSI Directive'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 essential services. 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. This should be taken into account in current and subsequent revisions of the PSI Directive.
- **Sharing obligations**: A clear increase can be observed in legislation imposing data sharing obligations, which can be linked to the development of Intelligent Transport Systems. 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.

- **Data ownership and Data sharing agreements**: 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. 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 in the form of legislative intervention would be desirable.
- **Liability**: In terms of extra-contractual and statutory liability, 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. In terms of contractual liability, the current rules, which may differ across the EU, are likely to limit the uptake of new technologies, including big data in the transport sector. The Draft Directive on the Supply of Digital Content aims to deal with liability of suppliers of digital content towards consumers specifically. One may however wonder about the precise necessity and opportunity to regulate liability in the specific context of the supply of digital content, rather than more broadly. A broader assessment of the adequacy of the current liability regime in the EU should therefore take place.
- **Competition**: 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 legal literature on this topic. 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.

The foregoing considerations and suggestions to move forward will be borne in mind during the remainder of the LeMO Project, notably while conducting the case studies, and may notably be re-assessed in the context of Tasks 3.3, 4.3 and 4.4.

References

Books:

Berenboom A, *Le nouveau droit d'auteur et les droits voisins* (4th edition, Larcier, Brussels 2008)

Bygrave L A, *Data Protection Law: Approaching Its Rationale, Logic and Limits* (Information Law Series 10, Kluwer Law International 2002)

Lefranc D, *Droit des applications connectées* (Larcier 2017)

OECD, *Data-driven Innovation: Big Data for Growth and Well-being* (OECD Publishing 2015)

Ohlhorst F J, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012)

Ustaran E, *European Privacy: Law and Practice for Data Protection Professionals* (IAPP, 2011)

Journal articles:

Bârcă C-D, Ropot R and Dumitrescu S, 'eCall – Minimum Set Of Data (MSD)' (2009) Journal of Information Systems & Operations Management 428

Bensoussan A, 'Propriété des données et protection des fichiers' (2010) 296 Gazette du Palais 2

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

Beyneix I, 'Le traitement des données personnelles par les entreprises: big data et vie privée, état des lieux' (2015) 46-47 Semaine juridique 2113

Borghi M and Karapapa S, 'Contractual Restrictions on Lawful Use of Information: Sole-source Databases Protected by the Back Door?' (2015) 37(8) EIPR 505

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

Casassa-Mont M and Pearson S, 'Sticky Policies: An Approach for Managing Privacy across Multiple Parties' (2011) 44(9) Computer 60

Castets-Renard C, '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

César J and Debussche J, 'Novel EU Legal Requirements in Big Data Security: Big Data – Big Security Headaches?' (2017) 8(1) JIPITEC 79 <https://www.jipitec.eu/issues/jipitec-8-1-2017/4534> accessed 17 October 2018

Chui M and Farell D, 'A Closer Look at Open Data: Opportunities for Impact' (2014) Innovation in local government: open data and information technology 24

Davilla M, 'Is Big Data a Different Kind of Animal? The Treatment of Big Data under the EU Competition Rules' (2017) 8(6) Journal of European Competition Law & Practice 370

Debussche J and Van Asbroeck B, 'Cloud Computing and Privacy Series: a Legal Perspective on Data Anonymisation (part 4 of 6)' (2015) 20(2) Cyberspace Lawyer 7

Dorner M, 'Big Data und "Dateneigentum"' (2014) 9 CR 617

Drexl J, '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

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

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

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

Ghotkar M and Rokde P, 'Big Data: How it is Generated and its Importance' (2016) 2 IOSR-JCE

Grützmacher M, 'Dateneigentum – ein Flickenteppich' (2016) 8 CR 485

Hoeren T, 'Big Data and the Ownership in Data: Recent Developments in Europe' (2014) 36(12) EIPR 751

Hon K and others, 'The Problem of 'Personal Data' in Cloud Computing – What Information is Regulated? The Cloud of Unknowing, Part 1' (2011) 1(4) IDPL 211

Hutchinson T and Duncan N, 'Defining and Describing What We Do: Doctrinal Legal Research' (2012) 17/1 Deakin Law Review 83

Janecek V, 'Ownership of Personal Data in the Internet of Things' (2018) 34(5) Computer Law & Security Review 1039

Kelly R and Swaby G, 'Consumer Protection Rights and "Free" Digital Content' (2017) 23(7) Computer and Telecommunications Law Review 165

Lederman J, Taylor B D and Garett M, 'A Private Matter: The Implications of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

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

Matchi Aïvodji U, Gambs S, Huguet M-J and Killijian M-O, 'Meeting Points in Ridesharing: A Privacy-preserving Approach' (2016) 72 Transportation Research Part C: Emerging Technologies 239

Mendoza-Caminade A, 'La protection pénale des biens incorporels de l'entreprise: vers l'achèvement de la dématérialisation du délit' (2015) 7 Recueil Dalloz 415

Metzger A, 'Data as Counter-Performance: What Rights and Duties do Parties Have?' (2017) 8(1) JIPITEC 2 <https://www.jipitec.eu/issues/jipitec-8-1-2017/4528> accessed 23 October 2018

Millwood S, '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

Mysoor P, "Protecting the Unprotected Database" (2015) 131 LQR 556

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

Potters J and Suetens S, 'Oligopoly Experiments in the Current Millennium' (2013) 27(3) Journal of Economic Surveys 439

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

Scassa T, 'Public Transit Data Through an Intellectual Property Lens: Lessons About Open Data' (2014) 41(5) Fordham Urb. L.J. 1759

Spindler G and Schmechel P, '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 17 October 2018

Strowel A, 'Reconstructing the Economic Rights: Taking Copyright Seriously' (2016)

Svantesson D, 'A Legal Method for Solving Issues of Internet Regulation' [2011] 19(3) International Journal of Law and Information Technology 243

The Economist, 'The World's Most Valuable Resource. Data and the new rules of competition' (2017) The Economist 14

Van Eechoud M, 'Making Access to Government Data Work' (2015) 9(2) Masaryk University Journal of Law and Technology 61

Verellen T, 'Het voorstel tot herziening van de PSI-Richtlijn: Hoe open is open data?' Draft article submitted to the Revue du droit des industries de réseau (forthcoming)

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

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

Papers:

Bundeskartellamt, 'The French Autorité de la Concurrence and the German Bundeskartellamt Publish Joint Paper on Data and its Implications for Competition Law' (*Bundeskartellamt*, 10 May 2016) <https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2016/10_05_2016_Big%20Data%20Papier.html> accessed 19 October 2018

Drexl J 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

Ferracane M F, 'Restrictions on Cross-Border Data Flows: A Taxonomy' (ECIPE Working Paper, No. 1/2017) <http://ecipe.org//app/uploads/2017/11/Restrictions-on-cross-border-data-flows-a-taxonomy-final1.pdf> accessed 17 October 2018

Kerber W, '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/fb02/makro/forschung/magkspapers/paper_2016/37-2016_kerber.pdf> accessed 18 October 2018

Koski H, 'Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth?'(Keskusteluaiheita Discussion Papers N 1260, Etla, the Research Institute of the Finnish Economy, 2011) <https://www.etla.fi/wp-content/uploads/2012/09/dp1260.pdf> accessed 18 October 2018

Law Commission, 'Data Sharing between Public Bodies' (Consultation Paper No. 214, Law Commission 2013) <http://www.lawcom.gov.uk/app/uploads/2015/03/cp214_data-sharing.pdf> accessed 18 October 2018

Martens B, 'JRC Technical Reports: An Economic Policy Perspective on Online Platforms' (Institute for Prospective Technological Studies Digital Economy Working Paper 2016/05, European Commission 2016) <https://ec.europa.eu/jrc/sites/jrcsh/files/JRC101501.pdf> accessed 18 October 2018

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=9725A618211DF41C00207963B84C43F0> accessed 17 October 2018

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

Van Asbroeck B, Debussche J and César J, '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 18 October 2018

Van Asbroeck B, Debussche J, César J, '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 18 October 2018

Publications in press:

Anderson C, 'Swedish Government Scrambles to Contain Damage From Data Breach' *The New York Times* (25 July 2017) <https://www.nytimes.com/2017/07/25/world/europe/ibm-sweden-data-outsourcing.html> accessed 17 October 2018

Apostle J, 'The Uber Data Breach Has Implications for us all' *Financial Times* (27 November 2017) <https://www.ft.com/content/e2bf6caa-d2cb-11e7-a303-9060cb1e5f44> accessed 19 October 2018

Batey E and Baume M, 'Does a Private Company Own Your Muni Arrival Times?' *The San Francisco APPEAL* (25 June 2009) <http://sfappeal.com/2009/06/who-owns-sfmta-arrival-data/> accessed 17 October 2018

BBC, 'Heathrow Probe After 'Security Files Found on USB Stick'' *BBC News* (29 October 2017) <https://www.bbc.com/news/uk-41792995> accessed 17 October 2018

Fortes F, 'Microsoft Given Go-ahead by US Antitrust Enforcers to Buy LinkedIn' *MLex* (2016) <http://www.mlex.com/GlobalAntitrust/DetailView.aspx?cid=830831&siteid=191&rdir=1> accessed on 18 October 2018

Isaac M and Scheiber N, 'Uber Settles Cases with Concessions, but Drivers Stay Freelancers' *The New York Times* (21 April 2016) <https://www.nytimes.com/2016/04/22/technology/uber-settles-cases-with-concessions-but-drivers-stay-freelancers.html?_r=1> accessed 19 October 2018

Konrad A, 'Amazon Jumps Into Internet Of Things Frenzy With New Cloud Platform For Devices Like Car'' *Forbes* (08 October 2015) <https://www.forbes.com/sites/alexkonrad/2015/10/08/amazon-jumps-into-internet-of-things-frenzy-with-new-cloud-platform/#4198e2eede4d> accessed 19 October 2018

Lee P, 'Competition Law Focus on "Big Data"' *Lexology* (2016) <https://www.lexology.com/library/detail.aspx?g=48c764ce-33f2-493c-9760-056008b20082> accessed 19 October 2018

Robinson D, 'Facebook Faces EU Complaint over WhatsApp Deal' *Financial Times* (20 December 2016) <https://www.ft.com/content/7ed82560-f534-3d78-bc84-8043301b6c85> accessed 19 October 2019

Wartburton D, 'Terror Threat as Heathrow Airport Security Files Found Dumped in the Street' *Mirror news* (28 October 2017) <https://www.mirror.co.uk/news/uk-news/terror-threat-heathrow-airport-security-11428132> accessed 17 October 2018

Edited books:

Ezrachi A and Stucke M E, *Virtual Competition: The Promise and Perils of the Algorithm-driven Economy* (Harvard UP 2016)

Goldstein P and Hugenholtz B, *International Copyright. Principles, Law, and Practice* (3rd edition, Oxford University Press 2013)

Kalyvas J R and Overly M R, *Big Data: A Business and Legal Guide* (Auerbach Publications 2014)

Whish R and Bailey D, *Competition Law* (Oxford UP 2015)

Chapters in edited books:

Cavanillas Múgica S, 'Liability for Defective Information Provided on the Internet' in Degrave E, Terwangne C, Dusollier S and Queck R (eds), *Law, Norms and Freedoms in Cyberspace / Droit, Normes et Libertés dans le Cybermonde* (Larcier 2018)

De Hert P and Malgieri G, 'Making the Most of New Laws: Reconciling Big Data Innovation and Personal Data Protection within and beyond the GDPR' in de Degrave E, Terwangne C, Dusollier S and Queck R (eds), *Law, Norms and Freedoms in Cyberspace / Droit, Normes et Libertés dans le Cybermonde* (Larcier 2018)

Forgó N, Hänold S and Schütze B, 'The Principle of Purpose Limitation and Big Data' in Corrales M, Fenwick M and Forgó N (eds), *New Technology, Big Data and the Law* (Perspectives in Law, Business and Innovation, Springer 2017)

Gutwirth S and Gonzales Fusters G, 'L'éternel retour de la propriété des données : de l'insistance d'un mot d'ordre' in de Degrave E, Terwangne C, Dusollier S and Queck R (eds), *Law, Norms and Freedoms in Cyberspace / Droit, normes et libertés dans le cybermonde* (Larcier 2018)

Hugenholtz B, 'Against Data Property' in Ullrich H, Drahos P and Ghidini G (eds), *Kritika: Essays on Intellectual Property* (Volume 3, Edward Elgar Publishing Limited 2018)

Jacquemin H and Hubin J-B, 'Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle' in Jacquemin H and De Streel A (eds), *Intelligence artificielle et droit* (Larcier 2017)

OECD, 'Exploring Data-driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"' in OECD (ed), *Supporting Investment in Knowledge Capital, Growth and Innovation* (OECD Publishing 2013)

Overly M R, 'Information Security in Vendor and Business Partner Relationships' in Kalyvas J R and Overly M R (eds), *Big Data: A Business and Legal Guide* (Auerbach Publications 2015)

Vivant M, 'Responsabilité des intermédiaires techniques de l'internet : l'obscure clarté d'un droit sans boussole apparente' in de Degrave E, Terwangne C, Dusollier S and Queck R (eds), *Law, Norms and Freedoms in Cyberspace / Droit, normes et libertés dans le cybermonde* (Larcier 2018)

Conference papers:

Damiani E, 'Toward Big Data Risk Analysis' in IEEE (ed), *Proceedings of the 2015 IEEE International Conference on Big Data* (IEEE, 2015) DOI: 10.1109/BigData.2015.7363966

Goodfellow I, Shlens J and Szegedy C, 'Explaining and Harnessing Adversarial Examples' (2015) arXiv:1412.6572

Hernandez-Serrano J and others, 'On the Road to Secure and Privacy-Preserving IoT Ecosystems' in Podnar Žarko I, Broering A, Soursos S and Serrano M (eds), *Interoperability and Open-Source Solutions for the Internet of Things* (InterOSS-IoT 2016, Lecture Notes in Computer Science, volume 10218, Springer 2017)

Zurkinden N, 'AI and Driverless Cars: From International Law to Test Runs in Switzerland to Criminal Liability Risks' in Jacquemin H and de Streel A (eds), *L'intelligence artificielle et le droit* (Larcier 2017)

Websites:

Ackerman E, '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 17 October 201

Authority for Consumers & Markets, 'Online platforms onder de loep!' (*Authority for Consumers & Markets*, 21 September 2016) <https://www.acm.nl/nl/publicaties/publicatie/16332/Online-video-platforms-onder-de-loep/> accessed 18 October 2018

Autorité de la concurrence, 'L'Autorité de la concurrence se saisit pour avis afin d'analyser les conditions d'exploitation des données dans le secteur de la publicité en ligne' (*Autorité de la concurrence*, 23 May 2016) <http://www.autoritedelaconcurrence.fr/user/standard.php?id_rub=629&id_article=2777&lang=fr> accessed 18 October 2018

Bundeskartellamt, 'Preliminary Assessment in Facebook Proceeding: Facebook's Collection and Use of Data from Third-party Sources is Abusive' (*Bundeskartellamt*, 19 December 2017) <https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2017/19_12_2017_Facebook.html> accessed 18 October 2018

Buttarelli G, 'A Smart Approach: Counteract the Bias in Artificial Intelligence' (*EDPS*, 8 November 2016) <https://edps.europa.eu/press-publications/press-news/blog/smart-approach-counteract-bias-artificial-intelligence_de> accessed 16 October 2018

Carr A, 'Uber to Pay $148 Million in Settlement Over 2016 Data Breach' (*Bloomberg*, 26 September 2018) <https://www.bloomberg.com/news/articles/2018-09-26/uber-to-pay-148-million-in-settlement-over-2016-data-breach> accessed 17 October 2018

Clarke M, 'Big Data in Transport' (*The Institution of Engineering and Technology*, 2016) <https://www.theiet.org/sectors/transport/topics/intelligent-mobility/articles/big-data.cfm?origin=carousel> accessed 16 October 2018

Data Protection Network, 'DPN Legitimate Interests Guidance – GDPR (version 2.0)' (*DPN*, 2018) <https://www.dpnetwork.org.uk/dpn-legitimate-interests-guidance/> accessed 16 October 2018

Dogtiev A, 'Uber Revenue and Usage Statistics' (*Business of Apps*, 23 July 2018) <http://www.businessofapps.com/uber-usage-statistics-and-revenue/> accessed 19 October 2018

Drieger P, '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

EfficienSea2, 'Maritime Connectivity Platform' (*EfficienSea2*) <https://efficiensea2.org/solution/maritime-connectivity-platform/> accessed 17 October 2018

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

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

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

European Commission, 'Consultation on Evaluation of Procedural and Jurisdictional Aspects of EU Merger Control' (*European Commission*, 2017) <http://ec.europa.eu/competition/consultations/2016_merger_control/index_en.html> accessed 19 October 2018

European Commission, 'WP29 Has Established a Taskforce on the UBER Data Breach Case' (*European Commission*, 29 November 2017) <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=609786> accessed 17 October 2018

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

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

European Commission, 'Digital Single Market: EU Negotiators Reach a Political Agreement on Free Flow of Non-personal Data' (*European Commission*, 19 June 2018) <http://europa.eu/rapid/press-release_IP-18-4227_en.htm> accessed 20 September 2018

European Commission, 'Joint statement by Vice-President Ansip and Commissioner Gabriel on the European Parliament's Vote on the New EU Rules Facilitating the Free Flow of Non-personal Data' (*European Commission*, 4 October 2018) <http://europa.eu/rapid/press-release_STATEMENT-18-6001_en.htm> accessed 25 October 2018

European Commission, 'Intelligent Transport Systems: Action Plan and Directive' (*European Commission*, 2018) <https://ec.europa.eu/transport/themes/its/road/action_plan_en> accessed 18 October 2018

European Data Portal, 'The PSI Directive and GDPR' (*European Data Portal*, 2018) <https://www.europeandataportal.eu/en/highlights/psi-directive-and-gdpr> accessed 18 October 2018

European Open Data Portal, 'Benefits of Open Data' (*European Open Data Portal*, 2018) <https://www.europeandataportal.eu/en/using-data/benefits-of-open-data> accessed 18 October 2018

Goodfellow I and others, 'Attacking Machine Learning with Adversarial Examples' (*OpenAI*, 24 February 2017) <https://blog.openai.com/adversarial-example-research/> accessed 17 October 2018

Hary E, '[Cabanon] Can Anonymised Data still Be Useful?' (*LINC*, 14 November 2017) <https://linc.cnil.fr/fr/cabanon-can-anonymised-data-still-be-useful> accessed 17 October 2018

Hue Williams M and Monck-Mason J, '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 17 October 2018

Information Commissioner's Office, 'Automated Decision Taking' (*ICO*, 2016) <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-gdpr-say-about-automated-decision-making-and-profiling/> accessed 22 October 2018

Khosrowshahi D, '2016 Data Security Incident' (*Uber Newsroom*, 21 November 2017) <https://www.uber.com/newsroom/2016-data-incident/> accessed 17 October 2018

Laboratoire d'Innovation Numérique de la CNIL, 'CabAnon: Exploring and Visualizing Anonymized Datasets' (*LINC*, 22 February 2017) <https://linc.cnil.fr/fr/cabanon-exploring-and-visualizing-anonymized-datasets> accessed 17 October 2018

Law J and Martin E A, *A Dictionary of Law* (7th edition, Oxford University Press 2014) <http://www.oxfordreference.com/view/10.1093/acref/9780199551248.001.0001/acref-9780199551248-e-2745?rskey=2MFh2r&result=2900> accessed 18 October 2018

Lee P, 'Privacy, Security and Information Law. To Keep or not to Keep: Data Retention Challenges and Solutions' (*Fieldfisher*, 30 July 2018) <https://privacylawblog.fieldfisher.com/2018/to-keep-or-not-to-keep-data-retention-challenges-and-solutions> accessed 16 October 2018

Marr B, '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 16 October 2018

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

Moerel L and Prins C, 'On the Death of Purpose Limitation' (*IAPP*, 2 June 2015) <https://iapp.org/news/a/on-the-death-of-purpose-limitation/> accessed 16 October 2018

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

Neumann C-S, 'Big Data versus Big Congestion: Using Information to Improve Transport' (*McKinsey & Company*, July 2015) <https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/big-data-versus-big-congestion-using-information-to-improve-transport> accessed 19 October 2018

Pfeifle S, 'Is the GDPR a Data Localization Law?' (*IAPP*, 29 September 2017) <https://iapp.org/news/a/is-the-gdpr-a-data-localization-law/> accessed 17 October 2018

Port of Rotterdam, 'Barge Performance Monitor' (*Port of Rotterdam*) <https://www.portofrotterdam.com/nl/zakendoen/logistiek/verbindingen/barge-performance-monitor> accessed 18 October 2018

Ronzitti V, '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-risks-hindering-innovation-and-investments-in-public-services/> accessed 18 October 2018

Scofield M, 'Issues of Data Ownership', (*Information Management*, 1 November 1998) <http://www.information-management.com/issues/19981101/296-1.html> accessed 18 October 2018

Simonite T, 'AI Software Learns to Make AI Software' (*MIT Technology Review*, 18 January 2017) <https://www.technologyreview.com/s/603381/ai-software-learns-to-make-ai-software/> accessed 19 October 2018

Thomas T, 'Artificial Intelligence in Digital Marketing: How Can It Make Your Life Easier?' (*Zeta*, 25 April 2017) <https://zetaglobal.com/blog-posts/artificial-intelligence-in-digital-marketing/> accessed 19 October 2018

TomTom for Developers, 'TomTom Maps APIs for Developers' (*TomTom*, 2018) <https://developer.tomtom.com/tomtom-maps-apis-developers> 23 October 2018

Transforming Transport, 'Integrated Urban Mobility: Tampere Pilot' (*TT*, 2018) <https://transformingtransport.eu/domains/integrated-urban-mobility-tampere-pilot> accessed 17 October 2018

Vallet F, '[CabAnon] Anonymity vs Usability, Another Shot at Anonymizing the NYC Taxi Dataset' (*LINC*, 24 September 2018) <https://linc.cnil.fr/fr/cabanon-anonymity-vs-usability-another-shot-anonymizing-nyc-taxi-dataset> accessed 17 October 2018

Van Parijs J, 'Open Data in Public Private Partnerships: How Citizens can Become True Watchdogs' (*Open Knowledge International Blog*, 29 October 2010) <http://blog.okfn.org/2010/10/29/open-data-in-public-private-partnerships-how-citizens-can-become-true-watchdogs/> accessed 18 October 2018

Reports and studies:

Article 29 Data Protection Working Party, 'Opinion 10/2004 on More Harmonised Information Provisions' (2004) WP 100

Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data' (2007) WP 136

Article 29 Data Protection Working Party, 'Opinion 2/2009 on the Protection of Children's Personal Data (General Guidelines and the special case of schools)' (2009) WP 160

Article 29 Data Protection Working Party, 'Opinion 1/2010 on the concepts of "controller" and "processor"' (2010) WP 169

Article 29 Data Protection Working Party, 'Opinion 02/2013 on apps on smart devices' (2013) WP 202

Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation' (2013) WP 203

Article 29 Data Protection Working Party, 'Opinion 03/2014 on Personal data breach notification' (2014) WP 213

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

Article 29 Data Protection Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (2014) WP 217

Article 29 Data Protection Working Party, 'Statement on the role of a risk-based approach in data protection legal frameworks' (2014) WP 218

Article 29 Data Protection Working Party, 'Guidelines on the Recent Developments on the Internet of Things' (2014) WP 223

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

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

Article 29 Data Protection Working Party, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining whether Processing is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679' (2017) WP 248

Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and profiling for the purposes of regulation 2016/679' (2017) WP 251

Article 29 Data Protection Working Party, 'Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS)' (2017) WP 252

Article 29 Data Protection Working Party, 'Opinion 17/2017 on Consent under Regulation 2016/679' (2017) WP 259

Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under Regulation 2016/679' (2018) WP 250 rev.01

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)

Authority for Consumers & Markets, 'Report: Taking a Closer Look at Online Video Platforms' (Authority for Consumers & Markets 2017) <https://www.acm.nl/en/publications/publication/17575/Report-Taking-a-closer-look-at-online-video-platforms> accessed 18 October 2018

Autorité de la concurrence, 'Avis n° 18-A-03 du 6 mars 2018 portant sur l'exploitation des données dans le secteur de la publicité sur internet. The FCA examined various practices including bundling/tying, "low prices" and exclusivity, leveraging effects, discriminatory treatment, restrictions on interoperability, and restrictions on the ability to collect and access data' (2018) <http://www.autoritedelaconcurrence.fr/pdf/avis/18a03.pdf> accessed 18 October 2018

Barbero M 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

Borzacchiello M T, Boguslawski R, Pignatelli F, '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

British Academy and the Royal Society, 'Data Management and Use: Governance in the 21st Century' (2017) <https://royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf> accessed 24 October 2018

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

Bundeskartellamt and Autorité de la concurrence, 'Competition Law and Data' (2016) <https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2> accessed 18 October 2018

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

Bundesministerium für Verkehr und digitale Infrastruktur, 'Einsetzung der Kommission Wettbewerbsrecht 4.0' (BMVI) <https://www.bmwi.de/Redaktion/DE/Downloads/E/einsetzung-der-kommission-wettbewerbsrecht-4-0.pdf?__blob=publicationFile&v=6Its> accessed 18 October 2018

Carrara W, San Chan W, Fischer S and van Steenbergen E, 'Creating Value Through Open Data. Study on the Impact of Re-use of Public Data Resources' (European Commission 2015) <https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf> accessed 18 October 2018

Carrara W, Radu C and Vollers H, 'Open Data Maturity in Europe 2017' (European Data Portal 2017) <https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n3_2017.pdf> accessed 18 October 2018

Castro D, 'The False Promise of Data Nationalism' (ITIF 2013) <http://www2.itif.org/2013-false-promise-data-nationalism.pdf> accessed 17 October 2018

Cavoukian A, 'Privacy by Design: The 7 Foundational Principles' (PbD 2011) <https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf> accessed 16 October 2018

Commission de la protection de la vie privée, 'Big Data Rapport' (CPVP 2017) <https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf> accessed 16 October 2018

Competition & Markets Authority, 'Energy Market Investigation' (CMA 2016) <https://assets.publishing.service.gov.uk/media/5773de34e5274a0da3000113/final-report-energy-market-investigation.pdf> accessed 23 October 2018

Consten A and others, 'Deliverable 4.3 Report about Sea Traffic Analyses using AIS-data. Version 2017-07-21' (ESSnet Big Data 2017) <https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/5/5c/WP4_Deliverable_4.3_2017_07_21_v1.0.pdf> accessed 18 October 2018

D'Acquisto G and others, '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 16 October 2018

Damiani E and others, 'Big Data Threat Landscape and Good Practice Guide' (ENISA 2016) <https://www.enisa.europa.eu/publications/bigdata-threat-landscape> accessed 17 October 2018

Danezis G and others, 'Privacy and Data Protection by Design – from Policy to Engineering' (ENISA 2014) <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design> accessed 16 October 2018

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.org/wp-content/uploads/2017/11/Deloitte-EC-Digital-Content.pdf> accessed 17 October 2018

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-FB14-08A4-31F51A564C60F2A7_46146.pdf> accessed 26 October 2018

Directorate-General for Internal Policies – Policy Department C: Citizens' rights and constitutional affairs, 'Sale of Goods and Supply of Digital Content - Two Worlds Apart?' (European Commission 2016) 18-19 <http://www.europarl.europa.eu/cmsdata/98774/pe%20556%20928%20EN_final.pdf> accessed 26 October 2018

Directorate-General for Maritime Affairs and Fisheries, 'Legal Aspects of Maritime Monitoring & Surveillance Data: Summary Report' (European Commission 2009) <https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/docs/body/legal_aspects_maritime_monitoring_summary_en.pdf> accessed 18 October 2018

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

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 18 October 2018

European Commission, 'Annex to the Synopsis Report. Detailed Analysis of the Public Online Consultation Results on 'Building a European Data Economy" (European Commission 2017) <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 17 October 2018

European Commission, 'Synopsis Report: Consultation on the "Building a European Data Economy" Initiative' (European Commission 2017) <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

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

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 23 October 2018

European Data Protection Supervisor, 'Preliminary Opinion of the European Data Protection Supervisor. Privacy and Competitiveness in the Age of Big Data: The Interplay between Data

Protection, Competition Law and Consumer Protection in the Digital Economy' (EDPS 2014) <https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf> accessed 16 October 2018

European Data Protection Supervisor, 'Opinion 4/2015. Towards a New Digital Ethics. Data, Dignity and Technology' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf> accessed 16 October 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 16 October 2018

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

European Data Protection Supervisor, 'Opinion 8/2016. EDPS Opinion on Coherent Enforcement of Fundamental Rights in the Age of Big Data' (EDPS 2016) <https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf> accessed 18 September 2018

European Digital Rights, 'Feedback on the Free Flow of Non-personal Data' (EDRi 2017) <https://edri.org/files/freeflowdata_consultation_EDRi_20180122.pdf> accessed 17 October 2018

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 17 October 2018

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

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

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

González Fuster G and Scherrer A, '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)

<http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf> accessed 16October 2018

Hansen M, Hoepman J-H and Jensen M, 'Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies: Methodology, Pilot Assessment, and Continuity Plan' (ENISA 2015) <https://www.enisa.europa.eu/news/enisa-news/enisa-proposes-pets-maturity-assessment-methodology> accessed 16 October 2018

Henke N and others, 'The Age of Analytics: Competing in a Data-driven World', (McKinsey & Company 2016) <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx> accessed 19 October 2018

Herrera Anchustegui I and Nowag J, 'Buyer Power in the Big Data and Algorithm Driven World: the Uber and Lyft Example', (Competition policy international 2017) <https://www.competitionpolicyinternational.com/wp-content/uploads/2017/09/CPI-Anchustegui-Nowag.pdf> accessed 19 October 2018

Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (ICO 2012) <https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 17 October 2018

Information Commissioner's Office, 'Conducting Privacy Impact Assessments Code of Practice' (ICO 2014) <https://www.pdpjournals.com/docs/88317.pdf> accessed 16 October 2018

Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (ICO 2017) <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 16 October 2018

Jentzsch N, 'State-of-the-Art of the Economics of Cyber-Security and Privacy, IPACSCO – Innovation Framework for ICT Security Deliverable, No. 4.1' (Waterford Institute of Technology 2016) <https://www.econstor.eu/bitstream/10419/126223/1/Jentzsch_2016_State-Art-Economics.pdf> accessed 17 October 2018

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

Kommerskollegium, 'No Transfer, No Trade – the Importance of Cross-Border Data Transfers for Companies Based in Sweden' (Swedish National Board of Trade 2014) <https://www.kommers.se/Documents/In_English/Publications/PDF/No_Transfer_No_Trade.pdf> accessed 17 October 2018

Naydenov R and others, 'Big Data Security. Good Practices and Recommendations on the Security of Big Data Systems (ENISA 2016) <https://www.enisa.europa.eu/publications/big-data-security> accessed 17 October 2018

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

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

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(0).pdf> accessed 17 October 2018

OECD, 'Algorithms and Collusion – Note from the United Kingdom' (DAF/COMP/WD(2017)19, OECD 2017) <https://one.oecd.org/document/DAF/COMP/WD(2017)19/en/pdf> accessed 19 October 2018

OECD, 'Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact' (OECD Digital Government Studies, OECD Publishing 2018) <https://doi.org/10.1787/9789264305847-en> accessed 18 October 2018

Surakitbanharn C A 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

The Roadmap to Secure Control Systems in the Transportation Sector Working Group, 'Roadmap to Secure Control Systems in the Transportation Sector' (August 2012) <https://ics-cert.us-cert.gov/sites/default/files/documents/TransportationRoadmap20120831.pdf> accessed 17 October 2018

Tinholt D, 'The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data' (Capgemini Consulting 2013) <https://www.capgemini.com/wp-content/uploads/2017/07/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf> accessed 18 October 2018

TNS Opinion & Social, 'Special Eurobarometer 423 – Cyber Security Report' (European Commission 2015) <http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_423_en.pdf> accessed 17 October 2018

Triaille J-P, 'Study on the Legal Framework of Text and Data Mining (TDM)' (De Wolf & Partners 2014) <https://publications.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en> accessed 17 September 2018

Ubeda J A and others, 'Transforming Transport. Summary of deliverable' (Transforming Transport 2018) <https://transformingtransport.eu/sites/default/files/2018-08/D8.3_PUBLIC.pdf> accessed 16 October 2018

Van Gorp N and Batura O, 'Challenges for Competition Policy in a Digitalised Economy' (Directorate-General for Internal Policies Policy Department A Economic and Scientific Policy 2015) <http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU%282015%29542235> accessed 19 October 2018

van Til H, van Gorp N and Price K, 'Big Data and Competition' (Ecorys 2017) <https://www.rijksoverheid.nl/documenten/rapporten/2017/06/13/big-data-and-competition> accessed 23 October 2018

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

World Intellectual Property Organization, 'Guide to the Copyright and Related Right Treaties Administered by WIPO and Glossary of Copyright and Related Rights Terms' (WIPO 2003) <http://www.wipo.int/edocs/pubdocs/en/copyright/891/wipo_pub_891.pdf> accessed 17 October 2018

Speeches:

Vestager M, 'Making Data Work for us' (Data Ethics Event on Data as Power, Copenhagen, 9 September 2016) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/making-data-work-us_en> accessed on 19 September 2018

Vestager M, 'Big Data and Competition' (EDPS-BEUC Conference on Big Data, Brussels, 29 September 2016) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/big-data-and-competition_en> accessed 19 October 2018

Vestager M, 'Algorithms and Competition' (Bundeskartellamt 18th Conference on Competition, Berlin, 16 March 2017) <https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/bundeskartellamt-18th-conference-competition-berlin-16-march-2017_en> accessed 19 October 2018